



ECOLOGICAL economics

PRINCIPLES
AND APPLICATIONS

Herman E. Daly and Joshua Farley

About Island Press

Island Press is the only nonprofit organization in the United States whose principal purpose is the publication of books on environmental issues and natural resource management. We provide solutions-oriented information to professionals, public officials, business and community leaders, and concerned citizens who are shaping responses to environmental problems.

In 2004, Island Press celebrates its twentieth anniversary as the leading provider of timely and practical books that take a multidisciplinary approach to critical environmental concerns. Our growing list of titles reflects our commitment to bringing the best of an expanding body of literature to the environmental community throughout North America and the world.

Support for Island Press is provided by the Agua Fund, Brainerd Foundation, Geraldine R. Dodge Foundation, Doris Duke Charitable Foundation, Educational Foundation of America, The Ford Foundation, The George Gund Foundation, The William and Flora Hewlett Foundation, Henry Luce Foundation, The John D. and Catherine T. MacArthur Foundation, The Andrew W. Mellon Foundation, The Curtis and Edith Munson Foundation, National Environmental Trust, National Fish and Wildlife Foundation, The New-Land Foundation, Oak Foundation, The Overbrook Foundation, The David and Lucile Packard Foundation, The Pew Charitable Trusts, The Rockefeller Foundation, The Winslow Foundation, and other generous donors.

The opinions expressed in this book are those of the author(s) and do not necessarily reflect the views of these foundations.

Ecological Economics

*To Andrea and Marcia
And to the next generation, especially Liam, Anna, Will, and Isabel*



*“The human mind, so frail, so perishable, so full of inexhaustible dreams and hungers,
burns by the power of a leaf.”*

—Loren Eiseley

Copyright © 2004 Herman E. Daly and Joshua Farley

All rights reserved under International and Pan-American Copyright Conventions. No part of this book may be reproduced in any form or by any means without permission in writing from the publisher: Island Press, Suite 300, 1718 Connecticut Ave., NW, Washington, DC 20009.

Library of Congress Cataloging-in-Publication Data.

Daly, Herman E.

Ecological economics : principles and applications / by Herman E. Daly and Joshua Farley.

p. cm.

Includes bibliographical references and index.

ISBN 1-55963-312-3 (cloth : alk. paper)


1. Environmental economics. I. Farley, Joshua C., 1963— . II. Title.

HD75.6.E348 2003

333.7—dc22

British Cataloguing-in-Publication Data available.

Design by Mary McKeon

Printed on recycled, acid-free paper 

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Ecological Economics

Principles and Applications



Herman E. Daly and Joshua Farley

Island Press

Washington • Covelo • London



Contents

Acknowledgments / xv

A Note to Instructors / xvii

Introduction / xix

PART I AN INTRODUCTION TO ECOLOGICAL ECONOMICS / 1

Chapter 1 Why Study Economics? / 3

What Is Economics? / 3

The Purpose of This Textbook / 5

Coevolutionary Economics / 7

The Era of Ecological Constraints / 10

Big Ideas to Remember / 13

Chapter 2 The Fundamental Vision / 15

The Whole and the Part / 15

Optimal Scale / 16

Diminishing Marginal Returns and Uneconomic Growth / 19

A Paradigm Shift / 23

Say's Law: Supply Creates Its Own Demand / 26

Leakages and Injections / 26

Linear Throughput and Thermodynamics / 29

Big Ideas to Remember / 35

Chapter 3 Ends, Means, and Policy / 37

Ends and Means—A Practical Dualism / 37

The Presuppositions of Policy / 43

Determinism and Relativism / 44

The Ends-Means Spectrum / 48

Three Strategies for Integrating Ecology and Economics / 50

Big Ideas to Remember / 56

Conclusions to Part I / 57

PART II THE CONTAINING AND SUSTAINING ECOSYSTEM: THE WHOLE / 59

Chapter 4 The Nature of Resources and the Resources of Nature / 61

A Finite Planet / 62

The Laws of Thermodynamics / 64

Stock-Flow Resources and Fund-Service Resources / 70

Excludability and Rivalness / 72
Goods and Services Provided by the Sustaining System / 74
Big Ideas to Remember / 76

Chapter 5 Abiotic Resources / 77

Fossil Fuels / 78
Mineral Resources / 83
Water / 87
Ricardian Land / 88
Solar Energy / 89
Summary Points / 91
Big Ideas to Remember / 92

Chapter 6 Biotic Resources / 93

Ecosystem Structure and Function / 93
Renewable Resources / 98
Ecosystem Services / 103
Waste Absorption Capacity / 107
Big Ideas to Remember / 110

Chapter 7 From Empty World to Full World / 111

Fossil Fuels / 113
Mineral Resources / 116
Water / 116
Renewable Resources / 118
Waste Absorption Capacity / 119
Big Ideas to Remember / 121
Conclusions to Part II / 122

PART III MICROECONOMICS / 123

Chapter 8 The Basic Market Equation / 125

Components of the Equation / 126
What Does the Market Equation Mean? / 131
Monopoly and the Basic Market Equation / 133
Non-Price Adjustments / 134
Supply and Demand / 136
Big Ideas to Remember / 140

Chapter 9 Supply and Demand / 141

A Shift in the Curve Versus Movement Along the
Curve / 141
Equilibrium P and Q, Shortage and Surplus / 143
Elasticity of Demand and Supply / 146
The Production Function / 148
The Utility Function / 152
Big Ideas to Remember / 155

Chapter 10 Market Failures / 157

Characteristics of Market Goods / 157

Rivalness / 159

Open Access Regimes / 161

Excludable and Nonrival Goods / 164

Pure Public Goods / 168

Externalities / 175

Missing Markets / 180

Summary Points / 182

Big Ideas to Remember / 183

Chapter 11 Market Failures and Abiotic Resources / 185

Fossil Fuels / 186

Mineral Resources / 192

Freshwater / 196

Ricardian Land / 198

Solar Energy / 200

Big Ideas to Remember / 200

Chapter 12 Market Failures and Biotic Resources / 201

Renewable Resource Stocks and Flows / 201

Renewable Resource Funds and Services / 211

Waste Absorption Capacity / 215

Biotic and Abiotic Resources: The Whole System / 218

Big Ideas to Remember / 219

Conclusions to Part III / 220

PART IV MACROECONOMICS / 221**Chapter 13 Macroeconomic Concepts: GNP and Welfare / 223**

A Troubled Marriage / 224

Gross National Product / 228

Sustainable Income / 232

Alternative Measures of Welfare: MEW and ISEW / 233

Beyond Consumption-Based Indicators of Welfare / 237

Big Ideas to Remember / 244

Chapter 14 Money / 245

Virtual Wealth / 248

Seigniorage / 249

The Fractional Reserve System / 252

Money as a Public Good / 253

Money and Thermodynamics / 255

Big Ideas to Remember / 258

Chapter 15 Distribution / 259

- Pareto Optimality / 259
- The Distribution of Income and Wealth / 262
- The Functional and Personal Distribution of Income / 263
- Measuring Distribution / 264
- Consequences of Distribution for Community and Health / 267
- Intertemporal Distribution of Wealth / 268
- Big Ideas to Remember / 276

Chapter 16 The IS-LM Model / 277

- IS: The Real Sector / 279
- LM: The Monetary Sector / 282
- Combining IS and LM / 285
- Exogenous Changes in IS and LM / 286
- IS-LM and Monetary and Fiscal Policy / 289
- IS-LM in the Real World / 299
- Adapting IS-LM to Ecological Economics / 302
- Big Ideas to Remember / 304
- Conclusions to Part IV / 305

PART V INTERNATIONAL TRADE / 307

Chapter 17 International Trade / 309

- The Classical Theory: Comparative Advantage / 309
- Kinks in the Theory / 312
- Capital Mobility and Comparative Advantage / 314
- Absolute Advantage / 316
- Globalization vs. Internationalization / 317
- The Bretton Woods Institutions / 318
- The World Trade Organization / 319
- Summary Points / 321
- Big Ideas to Remember / 321

Chapter 18 Globalization / 323

- Efficient Allocation / 323
- Sustainable Scale / 331
- Just Distribution / 333
- Summary Points / 340
- Big Ideas to Remember / 341

Chapter 19 International Flows and Macroeconomic Policy / 343

- Balance of Payments / 344
- Exchange Rate Regimes / 344
- Capital Mobility and National Policy Levers / 346

Economic Stability / 348
 Big Ideas to Remember / 355
 Conclusions to Part V / 356

PART VI POLICY / 357

Chapter 20 General Policy Design Principles / 359

The Six Design Principles / 360
 Which Policy Comes First? / 363
 Controlling Throughput / 365
 Price vs. Quantity as the Policy Variable / 366
 Source vs. Sink / 368
 Policy and Property Rights / 369
 Big Ideas to Remember / 372

Chapter 21 Sustainable Scale / 373

Direct Regulation / 373
 Pigouvian Taxes / 376
 Pigouvian Subsidies / 378
 Tradeable Permits / 379
 Policy in Practice / 385
 Big Ideas to Remember / 387

Chapter 22 Just Distribution / 389

Caps on Income and Wealth / 390
 Minimum Income / 393
 Distributing Returns from the Factors of Production / 395
 Additional Policies / 404
 Big Ideas to Remember / 404

Chapter 23 Efficient Allocation / 405

Pricing and Valuing Nonmarket Goods and Services / 405
 Macro-Allocation / 412
 Spatial Aspects of Nonmarket Goods / 415
 Redefining Efficiency / 422
 Big Ideas to Remember / 424

Looking Ahead / 425

Glossary / 429
Suggested Readings / 443
About the Authors / 447
Index / 449

Acknowledgments

We are grateful to our many colleagues in the International Society for Ecological Economics for their intellectual contributions and the community of scholarship and support they provide. We especially wish to acknowledge Robert Costanza, and the faculty at the Gund Institute for Ecological Economics at the University of Vermont, as well as our colleagues at the University of Maryland School of Public Affairs. We are also immensely grateful to the Santa Barbara Family Foundation for financial support, and to Jack Santa Barbara for encouragement and substantive help far beyond the financial. Our editor at Island Press, Todd Baldwin, provided many helpful suggestions, in addition to guiding the whole process from idea to published book.

A Note to Instructors

A textbook is usually a pedagogically efficient presentation of the accepted concepts and propositions of a well-defined academic discipline. Although we try to be pedagogically efficient, this book is not a textbook in the above sense, because ecological economics is not a discipline, nor does it aspire to become one. For lack of a better term, we call it a “transdiscipline.” We think that the disciplinary structure of knowledge is a problem of fragmentation, a difficulty to be overcome rather than a criterion to be met. Real problems do not respect academic boundaries. We certainly believe that thinking should be “disciplined” in the sense of respecting logic and facts, but not “disciplinary” in the sense of limiting itself to traditional methodologies and tools that have become enshrined in the academic departments of neoclassical economics. Furthermore, ecological economics is still “under construction,” and therefore no fully accepted methodologies and tools exist. Instead, its practitioners draw on methodologies and tools from various disciplines to address a specific problem.

Much of what we present is more contentious and less cut-and-dried than what you would find in a standard economic principles text. While we are especially critical of standard economics’ excessive commitment to GNP growth and its neglect of the biophysical system in which the economy is embedded, we also recognize that much environmental destruction and other forms of misery are caused by insufficient attention to standard economics. For example, subsidized prices for natural resources, neglect of external costs and benefits, and political unwillingness to respect the basic notions of scarcity and opportunity cost are problems we join standard economists in decrying.

As will be clear to any economist, the sections presenting basic micro- and macroeconomics, as well as other parts discussing distribution and trade, are based on standard economics. We want to be very clear: We are not claiming that ecological economists invented supply and demand, or national income accounting, or comparative advantage. Although it may be unnecessary to state this, experience has taught us to be very explicit in recognizing the origins of certain economic concepts, even if they are now so accepted as to be in the public domain. There are enough real points of contention between standard and ecological economics that we don’t need to add any fictitious ones! On contentious issues we do not shy away from controversy, but we do try to avoid the temptation to fan the flames unnecessarily, and to remember that the conflict is primarily between ideas, rather

than the people holding them. It would be dishonest, however, not to admit that we *all* hold certain of our ideas with passion. If that were not the case, then no one would stand for anything, and studying economics would be very boring indeed. But among our passions should be a commitment to fairness—first in considering the views of others, and second in demanding equal treatment for our own views.

We the authors are both economists trained in the standard neoclassical Ph.D. programs that one finds in nearly all American universities. Between us we have taught and practiced economics for over 50 years in universities and development institutions in various countries. We are not “non-economists,” nor do we consider that the epithet describes an irredeemably fallen state. After all, most policy makers are “non-economists,” a fact for which we are sometimes grateful. We accept more of traditional economics than we reject, although we certainly do reject some of the things we were taught. We have little patience with anti-economists who want to abolish money, who consider all scarcity to be an artificial social construct, or who think that all of nature’s services should be free. On the other hand, we do not share the view of many of our economist colleagues that growth will solve the economic problem, that narrow self-interest is the only dependable human motive, that technology will always find a substitute for any depleted resource, that the market can efficiently allocate all types of goods, or that the laws of thermodynamics are irrelevant to economics. Precisely because ecological economists have some basic disagreements with standard economics, it is necessary to emphasize that these divergences are branchings from a common historical trunk, not the felling of that common trunk.

We have provided a workbook to supplement this textbook. We emphasize that the textbook is self-contained and in no way depends on the workbook. Nevertheless, some instructors and students will find the workbook helpful, especially regarding case studies, applications, and design of class projects.

Introduction

Probably the best introduction to our book is the conclusion of another book. The other book is *Something New Under the Sun* by historian J. R. McNeill.¹ McNeill argues that the Preacher in Ecclesiastes remains mostly but not completely right—there is indeed “nothing new under the sun” in the realm of vanity and wickedness. But the place of humankind within the natural world is not what it was. The enormity and devastating impact of the human scale on the rest of creation really is a new thing under the sun. And it greatly amplifies the consequences of vanity and wickedness. McNeill’s findings help place ecological economics in historical context, and to explain why it is important.² His conclusions are worth quoting at length:

Communism aspired to become the universal creed of the twentieth century, but a more flexible and seductive religion succeeded where communism failed: the quest for economic growth. Capitalists, nationalists—indeed almost everyone, communists included—worshipped at this same altar because economic growth disguised a multitude of sins. Indonesians and Japanese tolerated endless corruption as long as economic growth lasted. Russians and eastern Europeans put up with clumsy surveillance states. Americans and Brazilians accepted vast social inequalities. Social, moral, and ecological ills were sustained in the interest of economic growth; indeed, adherents to the faith proposed that only more growth could resolve such ills. Economic growth became the indispensable ideology of the state nearly everywhere. How?

This state religion had deep roots in earlier centuries, at least in imperial China and mercantilist Europe. But it succeeded fully only after the Great Depression of the 1930s. . . . After the Depression, economic rationality trumped all other concerns except security. Those who promised to deliver the holy grail became high priests.

These were economists, mostly Anglo-American economists. They helped win World War II by reflatting and managing the American and British economies. The international dominance of the United States after 1945 assured wide acceptance of American ideas, especially in economics, where American success was most conspicuous. Meanwhile the USSR proselytized within its geopolitical sphere, offering a version of the growth fetish administered by engineers more than by economists.

¹New York: Norton, 2000, pp. 334–336.

²For another discussion of the place of ecological economics in recent intellectual and historical context, see Peter Hay, *Main Currents in Western Environmental Thought*, Bloomington: Indiana University Press, 2002, Chapter 8.

American economists cheerfully accepted credit for ending the Depression and managing the war economies. Between 1935 and 1970 they acquired enormous prestige and power because, or so it seemed, they could manipulate demand through minor adjustments in fiscal and monetary policy so as to minimize unemployment, avoid slumps, and assure perpetual economic growth. They infiltrated the corridors of power and the groves of academe, provided expert advice at home and abroad, trained legions of acolytes from around the world, wrote columns for popular magazines—they seized every chance to spread the gospel. Their priesthood tolerated many sects, but agreed on fundamentals. Their ideas fitted so well with social and political conditions of the time that in many societies they locked in as orthodoxy. All this mattered because economists thought, wrote, and prescribed as if nature did not.

This was peculiar. Earlier economists, most notably the Reverend Thomas Malthus (1766–1834) and W.S. Jevons (1835–1882), tried hard to take nature into account. But with industrialization, urbanization, and the rise of the service sector, economic theory by 1935 to 1960 crystallized as a bloodless abstraction in which nature figured, if at all, as a storehouse of resources waiting to be used. Nature did not evolve, nor did it twitch and adjust when tweaked. Economics, once the dismal science, became the jolly science. One American economist in 1984 cheerfully forecast 7 billion years of economic growth—only the extinction of the sun could cloud the horizon. Nobel Prize winners could claim, without risk to their reputations, that “the world can, in effect, get along without natural resources.” These were extreme statements, but essentially canonical views. If Judeo-Christian monotheism took nature out of religion, Anglo-American economists (after about 1880) took nature out of economics.

The growth fetish, while on balance quite useful in a world with empty land, shoals, of undisturbed fish, vast forests, and a robust ozone shield, helped create a more crowded and stressed one. Despite the disappearance of ecological buffers and mounting real costs, ideological lock-in reigned in both capitalist and communist circles. No reputable sect among economists could account for depreciating natural assets. The true heretics, economists who challenged the fundamental goal of growth and sought to recognize value in ecosystem services, remained outside the pale to the end of the century. Economic thought did not adjust to the changed conditions it helped to create; thereby it continued to legitimate, and indeed indirectly to cause, massive and rapid ecological change. The overarching priority of economic growth was easily the most important idea of the twentieth century.

From about 1880 to 1970 the intellectual world was aligned so as to deny the massive environmental changes afoot. While economists ignored nature, ecologists pretended humankind did not exist. Rather than sully their science with the uncertainties of human affairs, they sought out pristine patches in which to monitor energy flows and population dynamics. Consequently they had no political, economic—or ecological—impact.

Is McNeill correct in his assessment that “the overarching priority of economic growth was easily the most important idea of the twentieth century”? It’s hard to imagine a more important one. There are still very few who question the priority of economic growth.³ Yet many students are turned off by economics for the reasons also given by McNeill, namely the economists’ total abstraction from nature and their extreme devotion to economic growth as the *summum bonum*. While this aversion is understandable, it would be very sad if the only students who studied economics were those who don’t realize the fundamental limits of the discipline, or those who, realizing that something was wrong, didn’t have the energy or courage to try to reform it.

Professor McNeill specifically meant ecological economics in his reference to the “true heretics” who remained outside the pale to the end of the century. The purpose of our textbook is to try to change that deplorable situation—to help the next generation of economists take proper account of nature and nature’s limits. Achieving this objective will require a fusion of insights and methodologies from numerous disciplines to create a transdisciplinary approach to economics. Such an approach is necessary if we hope to understand nature’s limits and create policies that allow our economy to develop within those limits. However, to achieve anything more than random outcomes, we must direct available means toward specific ends. McNeill convincingly argues that ever-greater material consumption provided by never-ending economic growth is the agreed-upon end for the majority of modern society. This emphasis on an impossible and probably undesirable end is arguably a more serious shortcoming to traditional economics than a limited understanding of means.

■ THE CALL FOR CHANGE

As this is written, there are news reports of a group of economics students in French and British universities who are rebelling against what they are being taught. They have formed a Society for Post-Autistic Economics. Their implicit diagnosis is apt, since autism, like conventional economics, is characterized by “abnormal subjectivity; an acceptance of fantasy rather than reality.” Ecological economics seeks to ground economic thinking in the dual realities and constraints of our biophysical and moral environments. Current “canonical assumptions”⁴ of insatiable wants and infinite resources, leading to growth forever, are simply not founded in reality.

³For an interesting political history of how growth came to dominate U.S. politics in the post-war era, see R. M. Collins, *More: The Politics of Growth in Postwar America*, New York: Oxford University Press, 2000.

⁴*Canonical* literally means “according to religious law” and commonly means according to accepted usage.

Their dire consequences are evident. And that truly is something new under the sun.

In the early days of ecological economics, it was hoped that the ecologist would take over the economist's territory and redeem the failures of an economics that neglected nature. While ecologists made many important contributions, one is forced to accept McNeill's assessment that their influence has basically been disappointing—most have been unwilling “to sully their science with the uncertainties of human affairs.” The reasons that many ecologists appear to have had difficulty in dealing with policy will be the subject of speculation in Chapter 2. While ecology may not share the same inadequacies as economics, studying ecosystems as if they were isolated from human affairs on a planet of six billion humans also suggests an inclination to accept fantasy over reality. Ecological economics, therefore, is not simply bringing the light of ecology into the darkness of economics. Both disciplines need fundamental reform if their marriage is to work.

Nor is autistic tunnel vision limited to economics and ecology. Most universities these days educate students within the narrow confines of traditional disciplines. Rather than training students to examine a problem and apply whatever tools are necessary to address it, they are trained in a set of discipline-specific tools that they are then expected to apply to all problems. The difficulty is that the most pressing problems we face today arise from the interaction between two highly complex systems—the human system and the ecological system that sustains it. Such problems are far too complex to be addressed from the perspective of a single discipline, and efforts to do so must either ignore those aspects of the problem outside the discipline, or apply inappropriate tools to address them. “Abnormal subjectivity” (autism) is the inevitable result of education in a single subject. Applying insights from one discipline to another can serve to dispel the fantasies to which each alone is prone. For example, how could an economist conversant with ecology or physics espouse infinite growth on a finite planet? Effective problem-solving research must produce a mutually intelligible language for communication across disciplines. Otherwise, each discipline shall remain isolated in its own autistic world, unable to understand the world around it, much less to resolve the problems that afflict it.

■ A TRANSDISCIPLINARY SCIENCE

Ecological economists must go well beyond the fusion of ecology and economics alone. The complex problems of today require a correspondingly complex synthesis of insights and tools from the social sciences, natural sciences, and humanities. We frequently see research in which teams of researchers trained in different disciplines separately tackle a single problem and then strive to combine their results. This is known as multidisci-

plinary research, but the result is much like the blind men who examine an elephant, each describing the elephant according to the single body part they touch. The difference is that the blind men can readily pool their information, while different academic disciplines lack even a common language with which their practitioners can communicate. Interdisciplinary research, in which researchers from different disciplines work together from the start to jointly tackle a problem and can reduce the language barrier as they go, is a step in the right direction. But while universities have disciplines, the real world has problems. Ecological economics seeks to promote truly transdisciplinary research in which practitioners accept that disciplinary boundaries are academic constructs irrelevant outside of the university, and allow the problem being studied to determine the appropriate set of tools, rather than vice versa.

Just as effective problem solving requires the insights and tool sets of a variety of disciplines, defining the goals toward which we should strive would benefit from open discussion of the value sets of different ideologies. Unfortunately, the two dominant ideologies of the twentieth century seem to lack sufficient diversity within their value sets to stimulate this discussion. Specifically, the former USSR and the West, though differing in important ways, shared a fundamental commitment to economic growth as the first priority. The Marxist's deterministic ideology of dialectical materialism refused any appeals to morality or justice. The "new socialist man" would emerge only under objective conditions of overwhelming material abundance, which in turn required maximum economic growth. Bourgeois selfishness would disappear only with the disappearance of scarcity itself. In the U.S. and the West generally, the bulk of society did not reject appeals to justice and morality, but we did come to believe that our moral resources were very scarce relative to our natural resources and technological powers. Our strategy was to grow first, in the hope that a bigger pie would be easier to divide than a smaller one.⁵

⁵Though for a large and powerful group within the West, the primacy of growth and the belief in the powers of the market has turned common conceptions of morality on their head. Ayn Rand, a highly influential philosopher and author, who counts Alan Greenspan (chairman of the Federal Reserve, and arguably one of the most powerful men in the world) among her fervent admirers argues that altruism is evil and selfishness is a virtue. Milton Friedman, Nobel Prize-winning economist of the Chicago school, argues in the same vein: "Few trends could so thoroughly undermine the very foundation of our free society as the acceptance by corporate officials of a social responsibility other than to make as much money for their stockholders as possible" (*Capitalism and Freedom*, Chicago: University of Chicago Press, 1962, p. 135. *Selfishness: A New Concept of Egoism*, New York: Signet, 1964. Milton Friedman, Nobel Prize-winning economist of the Chicago school, argues in the same vein: "Few trends could so thoroughly undermine the very foundation of our free society as the acceptance by corporate officials of a social responsibility other than to make as much money for their stockholders as possible" (*Capitalism and Freedom*, Chicago: University of Chicago Press, 1962, p. 135).

But in practice, McNeill's "growth fetish" dominated both systems, and both were unmindful of the costs of growth.

Infinite growth in a finite system is an impossible goal and will eventually lead to failure. The USSR failed first because its system of central planning, along with its neglect of human rights, was more inefficient than the decentralized markets and greater respect for human rights in the West. The USSR was less mindful of the social and environmental costs of growth than the West, so it collapsed sooner. Because of its greater efficiency, the West can keep going for a bit longer in its impossible quest. But it, too, will collapse under the accumulating cost of growth. However, thus far the collapse of the USSR has been recognized by the West only as a validation of our superior efficiency. The possibility that efficiency only buys time and that unlimited growth must eventually fail in the West as well is something we have not yet considered. Alternatives to our misguided goal of infinite growth and limitless material consumption will be discussed throughout this text.

THINK ABOUT IT!

Think about a problem society currently faces, one that you know something about. Does the information required to resolve this problem fit snugly within the boundaries of one academic discipline, or are insights from several different disciplines required to solve it? What disciplines might be involved? In your university, how much interaction exists between the departments (professors and students) of those different disciplines?

■ AN OVERVIEW OF THE TEXT

Part I of this book is an introduction to the subject of ecological economics. Ecological economics seeks not only to explain how the world works, but also to propose mechanisms and institutions for making it work better. Chapter 1 explains the basic subject matter of neoclassical and ecological economics, in order to show the full scope of the new transdiscipline of ecological economics. Having defined the territory, we first establish basic agreement on the fundamental nature of the system we propose to analyze. Chapter 2 begins by describing the core (preanalytic) vision of ecological economics, that the economic system is a part or subsystem of a larger global ecosystem that sustains it. This view is contrasted with the fundamental vision of neoclassical economics, that the economic system is a self-sufficient whole entity unto itself. If we seek to make a system work better, we need to know the resources available to us—the means—and the desired outcomes—the ends. Chapter 3 focuses on the ends-and-means spectrum, an essential step for understanding a science

that defines itself as a mechanism for connecting scarce means to alternative ends.

Part II focuses on the containing and sustaining Whole, the Earth and its atmosphere. In these chapters, we delve deeper into the nature of the Whole—the global ecosystem that sustains us by providing the resources that feed the economic process and the sinks where we dispose of our wastes. Chapter 4 establishes the fundamental importance of low entropy (useful, ordered matter-energy) in economic production, and the inevitability of its conversion via the economic process to high-entropy, disordered, useless waste. Chapter 5 addresses the tangible forms in which low entropy manifests itself, the abiotic goods and services provided by nature, and examines their specific market-relevant characteristics. Chapter 6 does the same for biotic resources. Chapter 7 shows that many of the goods and services provided by nature were formerly superabundant, and it made little difference if an economic system dedicated to allocating scarce resources ignored them. Now, however, as we have discussed in this Introduction, these resources have become scarce, and their allocation has become critically important.

Part III begins our examination of the part of the whole in which we are most interested, the economic subsystem. We draw out the useful elements of neoclassical economic theory and integrate them into ecological economics. Microeconomics, macroeconomics, or international trade each provides sufficient material for years of study, and this text conveys no more than the essentials.

Chapters 8 and 9 introduce microeconomics, the study of mechanisms for efficiently allocating specific scarce resources among specific alternative ends. They explain the self-organizing properties of a competitive market economy, through which millions of independent decision makers, freely acting in their own self-interest, can generate the remarkable outcomes alluded to at the start of Chapter 1. These chapters also explain how neoclassical production and utility functions must be modified to address the concerns of ecological economics.

In Chapter 10, we take a step back from the traditional microeconomic analysis of allocation to examine the specific characteristics resources must have if they are to be efficiently allocated by the market mechanism. We find that few of the goods and services provided by nature exhibit all of them. Attempts to allocate resources that do not have the appropriate characteristics via the unregulated market result in inefficient, unfair, and unsustainable outcomes. Rather than individual self-interest creating an invisible hand that maximizes social well-being, market allocation of such “nonmarket” goods creates an invisible foot that can kick the common good in the pants. Careful analysis of these market-relevant characteristics of scarce resources is an essential prelude to policy formulation. Thus,

Chapter 11 applies the concepts of market failures to abiotic resources, and Chapter 12 applies them to biotic ones.

In Part IV, we turn to macroeconomics. As we stated earlier, ecological economics views the economy as a part of a larger finite system. This means that the traditional goal of macroeconomic policy—unlimited economic growth in the physical dimension—is impossible. Thus, in ecological economics, optimal scale replaces growth as a goal, followed by fair distribution and efficient allocation, in that order. Scale and distribution are basically macroeconomic issues. Hence, in addition to the fiscal and monetary policy tools that dominate the discussion in traditional texts, we will introduce policies that can help the economy reach an optimal scale.⁶ Chapter 13 focuses on the basic macroeconomic concepts of GNP, and welfare. It starts by examining economic accounting, or the measurement of desirable ends ranging from gross national product to human needs assessment. Chapter 14 discusses the role of money in our economy. Chapter 15 focuses on the issue of distribution within and between generations, and Chapter 16 briefly develops the basic macroeconomic model of how saving and investing behavior combines with the supply and demand of money to determine the interest rate and level of national income. We then relate the macroeconomic model to policy levers designed to achieve the ecological economics goals of sustainable scale and just distribution.

Part V addresses international trade. In Chapters 17 and 18, we discuss how different economies interact and the troublesome issue of global economic integration. We consider especially the consequences of global integration for policy making. Chapter 19 looks at financial issues such as balance of payments and exchange rates, and examines the implications of globalization for macroeconomic policy.

Part VI focuses on policy. Chapter 20 presents the general design principles of policy. Chapter 21 reviews a number of specific policy options that primarily affect scale, Chapter 22 reviews policies that primarily affect distribution, and Chapter 23 reviews policies that primarily affect allocation.

Our concluding chapter, *Looking Ahead*, once again reflects on the ethical assumptions of ecological economics. We call for a return to the beginnings of economics as a moral philosophy explicitly directed toward raising the quality of life of this and future generations.

In summary, neoclassical economic theory arose primarily as an effort to explain the market economy. Ecological economics takes a more inclu-

⁶Optimal scale is the point where the marginal benefits of additional growth are just equal to the marginal costs of the reduction in ecosystem function that this growth imposes. As we will show in the text, numerous factors can affect optimal scale.

sive, and activist, position. We describe the nature of scarce resources and the ends for which they should be used and proactively prescribe appropriate institutions for their efficient allocation in a social context of just distribution and sustainable scale. We have the basic allocative institution of the market—it needs improvement, but at least it exists. We have no institution for limiting scale, and our institutions for governing distribution (antitrust, progressive taxation) have been allowed to atrophy. We know that building institutions is a political task and that “politics is the art of the possible.” That is a wise conservative counsel. Yet that dictum also prohibits attempting true physical impossibilities in a vain effort to avoid apparent political “impossibilities.” When faced with the unhappy dilemma of choosing between a physical or a political impossibility, it is better to attempt the politically “impossible.”

PART I

An Introduction to Ecological Economics





Why Study Economics?

■ WHAT IS ECONOMICS?

Economics is the study of the **allocation** of limited, or scarce, resources among alternative, competing ends. We can choose, for example, to allocate steel to plowshares or SUVs. These products in turn are apportioned to different individuals—Somalian farmers or Hollywood stars, for example. Of course, as a society we don't consciously choose to allocate steel to a particular number of plows or SUVs. But we do have collective desires, the sum of the individual choices that each of us makes to buy one thing or another. Really, economics is about what we desire and what we're willing to give up to get it.

In fact, three critical questions guide economic inquiry, and there is a clear order in which they should be asked:

1. What ends do we desire?
2. What limited, or scarce, resources do we need to attain these ends?
3. What ends get priority, and to what extent should we allocate resources to them?

This last question cannot be answered without deep reflection on the answers to the first two questions.

Traditionally, economists have said that the answer to the first question is “utility” or human welfare.¹ Welfare depends on what people want, which they reveal through market transactions—by what goods and services they buy and sell. Naturally, this only reveals preferences for market goods and implicitly assumes that nonmarket goods contribute little to

Allocation is the process of apportioning resources to the production of different goods and services. *Neoclassical economics* focuses on the market as the mechanism of allocation. *Ecological economics* recognizes that the market is only one possible mechanism for allocation.

¹Many neoclassical economists actually argue that economics is a positive science (i.e., based on value neutral propositions and analysis). Since desired ends are normative (based on values), they would therefore lie outside the domain of economic analysis.

welfare. Humans are assumed to be insatiable,² so welfare is increased through the ever-greater provision of goods and services, as measured by their market value. Thus, unending economic growth is typically considered an adequate, measurable proxy for the desirable end.

This view is fundamental to the main school of economics today, known as **neoclassical economics (NCE)**. Since neoclassical economists assume that markets reveal most desired ends and that most scarce resources are market goods, they devote most of their attention to the mechanism for allocating resources to alternative ends, which is, of course, the market. The reason the market is considered the appropriate mechanism is that under certain restrictive assumptions it is efficient, and efficiency is considered a value-free, objective criterion of “the good.” **Efficient allocation** is shorthand for **Pareto efficient allocation**, a situation in which no other allocation of resources would make at least one person better off without making someone else worse off. (The name Pareto is for the economist Vilfredo Pareto.) Efficiency is so important in neoclassical economics that it is sometimes taken to be an end in itself.³

Efficient allocation is shorthand for Pareto efficient allocation, a situation in which no other allocation of resources would make at least one person better off without making someone else worse off.

But we should bear in mind that if our ends were evil, then efficiency would just make things worse. After all, Hitler was rather efficient in killing Jews. Efficiency is only worthwhile if our ends are in fact good and well-ordered—a job not worth doing is not worth doing well. We will return to this in our discussion of an ends-means spectrum in Chapter 3.

Ecological economics takes a different approach than its neoclassical counterpart. In **ecological economics**, efficient allocation is important, but far from being an end in itself. Take the example of a ship. To load a ship efficiently is to make sure that the weight on both sides of the keel is the same, and the load is distributed from front to back so that the ship floats evenly in the water. While it is extremely important to load the cargo efficiently, it is even more important to make sure that not too much cargo is placed on the ship. It is of little comfort if an overloaded ship founders efficiently! Who is entitled to place their cargo on the ship is also important; we wouldn't want the passengers in first class to hog all the cargo space so that those in steerage lack adequate food and clothing for their voyage.

Ecological economists look at the Earth as a ship and gross material production of the economy as the cargo. The seaworthiness of the ship is determined by its ecological health, the abundance of its provisions, and its design. Ecological economists recognize that we are navigating unknown seas and no one can predict the weather for the voyage, so we don't know exactly how heavy a load is safe. But too heavy a load will cause the ship to sink.

²Insatiability means that we can never have enough of all goods, even if we can get enough of any one good at a given time.

³D. Bromley, *The Ideology of Efficiency: Searching for a Theory of Policy Analysis*, *Journal of Environmental Economics and Management* 19: 86–107 (1990).

Neoclassical economists focus solely on allocating the cargo efficiently. **Environmental economics**, a subset of neoclassical economics, recognizes that welfare also depends to a large extent on ecosystem services and suffers from pollution, but is still devoted to efficiency. As markets rarely exist in either ecosystem services or pollution, environmental economists use a variety of techniques to assign market values to them so that they, too, may be incorporated into the market model. Ecological economists insist on remaining within the weight limits (or in nautical terms, respecting the Plimsoll line⁴) determined by the ship design and the worst conditions it is likely to encounter, and making sure that all passengers have sufficient resources for a comfortable voyage. Once those two issues have been safely resolved, the hold is efficiently loaded.

Substantial evidence exists that the cargo hold is already too full for a safe voyage, or at least nearing capacity, and many passengers have not been allowed to load the basic necessities for the voyage. Certainly we seem to have too many greenhouse gases in the hold, too many toxic compounds. To make room for an ever-growing cargo, we have ripped out components of the ship we deem unimportant. But we live on a very complicated ship, and we know very little about its design and the impact of our choices on its structural integrity. How many forests and wetlands are required to keep it afloat? What species are crucial rivets, whose loss will compromise the ship's seaworthiness? Ecological economics addresses these issues. It also assumes that our goal is not simply to load the ship to the limit, but to maintain areas of the ship for our comfort and enjoyment, to revel in the exquisite beauty of its craftsmanship, and to maintain it in excellent condition for future generations.

So why study economics? If we do not, we will likely end up serving less important ends first and running out of resources while more important ends remain unmet. We are also likely to overload and swamp the ship unless we have studied the seas in which it will be sailing, as well as the ship's own design and functioning.

■ THE PURPOSE OF THIS TEXTBOOK

This textbook is designed to introduce ecological economics as a necessary evolution of conventional economic thought (neoclassical economics) that has dominated academia for over a century. Our text will

⁴In 1875, Samuel Plimsoll supported Britain's Merchant Shipping Act, requiring that a load-limit line be painted on the hull of every cargo ship using British ports. If the waterline exceeded the Plimsoll line, the ship was overloaded and prohibited from entering or exiting the port. Because of England's seafaring dominance, the practice was adopted worldwide. Yet shipowners who profited from overloading their ships fiercely resisted the measure. They could buy insurance at rates that made it profitable to occasionally risk losing an overloaded ship. The Plimsoll line has saved the lives of many sailors.

critique not only neoclassical economic theory but also the pro-growth market economy that in many people's minds has come to be virtually synonymous with American democracy. Ecological economists do not call for an end to markets. Markets are necessary. What must be questioned is the prevailing belief that markets reveal all our desires; that they are the ideal system not only for allocating all resources efficiently, but also for distributing resources justly among people; and that markets automatically limit the overall macroeconomy⁵ to a physical scale that is sustainable within the biosphere.

Part of our goal is to explain markets and show what they do well. Another part of our goal is to show why the unregulated market system is inadequate for allocating most of the goods and services provided by nature. This portion of the text should not be controversial—most of the basic arguments actually come from neoclassical economics, and it is only by drawing attention to their full implications that we depart from orthodoxy.

More contentious (and more important) is the call by ecological economics for an end to growth. We define **growth** as an increase in **throughput**, which is the flow of natural resources from the environment, through the economy, and back to the environment as waste. It is a quantitative increase in the physical dimensions of the economy and/or of the waste stream produced by the economy. This kind of growth, of course, cannot continue indefinitely, as the Earth and its resources are not infinite. While growth must end, this in no way implies an end to **development**, which we define as qualitative change, realization of potential, evolution toward an improved, but not larger, structure or system—an increase in the quality of goods and services (where quality is measured by the ability to increase human well-being) provided by a given throughput. Most of you have ceased growing physically, yet are probably studying this text in an effort to further develop your potential as humans. We expect human society to continue developing, and indeed argue that only by ending growth will we be able to continue developing for the indefinite future. Fortunately, many desirable ends require few physical resources.

The idea of “sustainable development,” to be discussed later, is development without growth—that is, qualitative improvement in the ability to satisfy wants (needs and desires) without a quantitative increase in throughput beyond environmental carrying capacity. Carrying capacity is the population of humans that can be sustained by a given ecosystem at a given level of consumption, with a given technology. Limits to growth do not necessarily imply limits to development.

Conventional neoclassical economists might define economic growth

Growth is a quantitative increase in size, or an increase in throughput.

Throughput is the flow of raw materials and energy from the global ecosystem, through the economy, and back to the global ecosystem as waste.

Development is the increase in quality of goods and services, as defined by their ability to increase human well-being, provided by a given throughput.

⁵Microeconomics focuses primarily on how resources are allocated toward the production and consumption of different goods and services. Macroeconomics traditionally focuses primarily on economic growth (i.e., the size of the economy), employment, and inflation.

as the increase in an economy's production of goods and services, typically measured by their market value, that is, an increase in gross national product (GNP). However, an economy can develop without growing, grow without developing, or do both at the same time. GNP lumps together quantitative growth with qualitative development—two very different things that follow very different laws—and is thus not a very useful measure.

In spite of the distinction between growth and development, calling for an end to growth requires an almost revolutionary change in social perceptions of the good (our ends and their ranking), a theme that will recur throughout this text. As we are all aware, the transition from adolescence to maturity is a difficult time for individuals, and will be for society as well.

The market economy is an amazing institution. Market forces are justly credited with an unprecedented and astonishingly rapid increase in consumer goods over the past three centuries. Poor people in affluent countries today have many luxuries that kings of Europe could not have dreamed of in centuries past, and we have achieved this through a system that relies on free choice. In the market in its pure form, individuals are free to purchase and produce any market good they choose, and there is no controlling authority apart from the free will of individual humans. Of course, the pure form exists only in textbooks, but competitive markets do show impressive powers of self-regulation. Arguments for modifying such an admittedly impressive system must be persuasive indeed. However, a brief detour into the history of markets and economics suggests that such modifications occur all the time.

■ COEVOLUTIONARY ECONOMICS⁶

As Karl Polanyi showed in his classic *The Great Transformation*,⁷ the economic system is embedded as a component of human culture, and like our culture, it is in a constant state of evolution. In fact, our ability to adapt to changing environmental circumstances through cultural evolution is something that most clearly distinguishes humans from other animals. Economic, social, and political systems, as well as technological advances, are examples of cultural adaptations. All these systems have adapted in response to changes in the environment, and these adaptations in turn provoke subsequent environmental change, to which we must

⁶Many of the basic ideas here come from the work of Richard Norgaard, including R. Norgaard, Coevolutionary Development Potential, *Land Economics* 60: 160–173 (1984) and R. Norgaard, Sustainable Development: A Coevolutionary View, *Futures*: 606–620 (1988).

⁷K. Polanyi. *The Great Transformation: The Political and Economic Origins of Our Time*. Boston: Beacon Press (2001).

again adapt in a coevolutionary process. Examples of some of the major coevolutionary adaptations and their implications for future change will help illustrate this concept.

From Hunter-Gatherer to Industrialist

For more than 90% of human history, humans thrived as small bands of nomadic hunter-gatherers. Anthropology and archaeology together provide us with a reasonable understanding of the hunter-gatherer economy. Rather than the “nasty, brutish and short” life that many imagine, early people met their basic needs by working only a few hours a day, and resources were sufficient to provide for both young and old who contributed little to gathering food. A recent study of the !Kung, who live in a very arid, marginal environment, found that 10% of the population was over 60, which compares favorably with populations in many industrialized countries.⁸

Small bands of hunter-gatherers would deplete local resources and then move on to places where resources were more abundant, allowing the resource base in the previous encampment to recover. Mobility was essential to survival, and accumulating goods reduced mobility. Numerous chronicles by anthropologists attest that hunter-gatherers show very little concern for material goods, readily discarding their possessions, confident in their ability to make new ones as needed.⁹ Property rights to land made no sense in a nomadic society, and prior to domestication some 10,000 years ago, property rights to animal herds were virtually impossible. Food was also shared regardless of who provided it, perhaps partly because of technological limits. Some food simply cannot be harvested in discrete bundles, and if hunters bring home a large game animal, unshared food would simply rot or attract dangerous predators.¹⁰ Studies of the !Kung and other tribes found that both young and old were generally exempt from food gathering, and even many mature men and women simply chose not to participate in this activity very often, yet were given equal shares of the harvest.¹¹

⁸R. Lee, “What Hunters Do for a Living.” In J. Gowdy, ed. *Limited Wants, Unlimited Means: A Reader on Hunter-Gatherer Economics and the Environment*. Washington, DC, Island Press, 1998.

⁹M. Sahlins, “The Original Affluent Society.” In J. Gowdy, op. cit.

¹⁰Recent anecdotal evidence supports this relationship between storage technology and property rights. In an indigenous community in Alaska, the government provided freezers for food storage, and the impact was dramatic. Where previously successful hunters would share their game with the community, freezers (probably contemporaneous with the breakdown of other social structures) enabled hunters to store their game for their own leisurely consumption. Older, younger, or weaker members of the community were left without a source of subsistence.

¹¹Lee, op. cit.

If for most of human existence, private property and wealth accumulation were impractical and absent from human society, it is hard to argue that these are inherent characteristics of human nature, rather than cultural artifacts.

Gradually hunter-gatherer societies developed the technology to store large quantities of food for months on end, an essential precursor to agriculture. Agriculture ended the nomadic lifestyle for many early peoples. People began to settle in towns or small communities, which led to greater population concentrations than had previously been possible.¹² The technologies of storage and agriculture changed the nature of property rights, and were in fact required before property rights could make sense. Certainly agriculture itself made some form of property rights to land essential. Surplus production allowed greater division of labor and specialization, which in turn led to ever-greater production, fostering extensive trade and eventually the development of money. Greater populations, the need to protect increasing riches against other groups, and the need to defend property rights within the community meant more need for government, and ruling classes developed.¹³ Ruling classes and the needs of the state clearly had to be supported through the productive capacity of others, which inevitably led to some sort of tax system and concentrations of wealth in the upper echelons of the hierarchy.

The chain of evolutionary events did not end there, of course. Higher populations and agriculture would have disrupted local ecosystems, eventually decreasing their capacity to produce food and materials independently of agriculture. This only increased the demands society would place on agriculture. These demands, accompanied by a more rapid exchange of ideas in denser communities, stimulated new technologies, such as large-scale irrigation.¹⁴ Irrigation over time led to increased soil salinity, eventually reducing the capacity of the ecosystem to sustain such high population levels without further agricultural innovations or migration.

The Industrial Revolution

Ever-greater surplus production, accompanied by better ships, allowed trade on an expanding scale. Traders exchanged not only goods but also ideas, further speeding up the rate of technological progress. Among the

¹²J. Diamond, *Guns, Germs, and Steel: The Fates of Human Societies*, New York: Random House, 1997.

¹³Many political philosophers argue that the primary purpose of government is to protect private property. In the words of John Locke: "Government has no other end but the preservation of property," from "An Essay Concerning the True Original, Extent, and End of Civil Government."

¹⁴Diamond, *op. cit.*

crucial technological leaps was the ability to use nonrenewable mineral resources. Trade also allowed specialization to take place across regions, not only across individuals within a society. Technological advance and global markets laid the groundwork for the Industrial Revolution.

The Industrial Revolution had profound impacts on the economy, society, and the global ecosystem. For the first time, human society became largely dependent on fossil fuels and other nonrenewable resources (partially in response to the depletion of forests as fuel). Fossil fuels freed us from dependence on the fixed flow of energy from the sun, but it also allowed the replacement of both human and animal labor by chemical energy. This increased energy allowed us ever-greater access to other raw materials as well, both biological and mineral. New technologies and vast amounts of fossil energy allowed unprecedented production of consumer goods. The need for new markets for these mass-produced consumer goods and new sources of raw material played a role in colonialism and the pursuit of empire. The market economy evolved as an extremely efficient way of allocating such goods, and stimulating the production of even more.

International trade exploded, linking countries together as never before. A greater ability to meet basic needs, and advances in hygiene and medical science, resulted in dramatic increases in population, whose needs were met through greater energy use and more rapid depletion of resources. Growing populations quickly settled the last remaining frontiers, removing the overflow valve that had allowed populations to relocate as local resources ran out. Per-capita consumption soared, and with it the waste output that now threatens to degrade our ecosystems.

■ THE ERA OF ECOLOGICAL CONSTRAINTS

As we stated earlier, economics is the science of the allocation of scarce resources among alternative ends. The success of the Industrial Revolution dramatically reduced the scarcity of consumer goods for much of the world's population. The accompanying economic growth, however, now threatens the former abundance of the goods and services produced by nature upon which we ultimately depend. These have become the newly scarce resources,¹⁵ and we must redesign our economic system to address that reality. Unfortunately, our ability to increase consumption while depleting our resource base has led people to believe that humans and the economy that sustains us have transcended nature. In the current system, the greatest claims to wealth have seemingly nothing to do with natural

¹⁵See R. Huefing, *The New Scarcity and Economic Growth: More Welfare Through Less Production?*, Amsterdam: North Holland, 1980.

resources, but rather are acquired through financial transactions on computers that physically do nothing more than move electrons. While knowledge and information are important, ultimately wealth requires physical resources. A recipe is no substitute for a meal, even though a good recipe may improve the meal.

Though the current economic system has been around for a remarkably short time in relation to past systems, it has wrought far greater environmental changes. These changes have redefined the notion of scarce resources, and they demand correspondingly dramatic changes in economic theory and in our economic system. Change in our economic system is inevitable. The only question is whether it will occur as a chaotic response to unforeseen disruptions in the global life support system, or as a carefully planned transition toward a system that operates within the physical limits imposed by a finite planet and the spiritual limits expressed in our moral and ethical values. The answer depends largely on how fast we act, and the burning question is: How much time do we have?

The Rate of Change

For the vast majority of human history, technological, social, and environmental changes occurred at a glacial pace. The agricultural revolution was really not a revolution but a case of evolution. For example, it probably took several thousand years to create corn from the ancestral stock of teosinthe.¹⁶ People generally saw no evidence of change from one generation to the next, and human culture could evolve at a correspondingly slow pace to adapt to the changes that did occur. Only with the Industrial Revolution did change really begin to accelerate to the extent that we could notice it from one generation to the next. And much of what the Industrial Revolution did was to increase the extraction of nonrenewable resources, thereby increasing human material consumption. As a result of this subsidy from nature, the general perception was that the future would always get better, and all that was needed was more of the same. Our response has been to use up this finite subsidy at ever greater rates, so that now, for the first time in human history, we can dramatically change the Earth's systems on a human time scale (a truly new thing under the sun). In fact, it threatens to alter the ability of the Earth to support life. While cultures have continually and slowly evolved in adapting to new technologies and new constraints, the unprecedented rate of change in technology and ecological degradation means we no longer have the luxury of biding our time. Most likely we will have to change our cultural institutions and values in response, particularly the economic institutions and

¹⁶Diamond, *op. cit.*

values that have led to this state of affairs. Since there is certainly some limit to how fast we can adapt culturally, we need also to consider seriously how to slow down the rate of change that is forcing the adaptations. It is worth remembering that not all change is desirable, and that even desirable change can be too fast.

The Difficulty of Achieving Desirable Change

It would be foolish to underestimate the difficulty of finding the right balance between limiting and adapting to change. Currently, our economic system is focused primarily on the microeconomic issue of efficient allocation. Applied economics also focuses on the macroeconomic issue of maximizing growth. Ecological economics, however, focuses primarily on the larger macroeconomic issue of how big is too big. This is the question of **scale**. How large, in its physical dimensions, should the economic system be relative to the ecosystem that sustains it? As soon as we ask this question, we imply that there is an optimal scale (and many believe we have already surpassed it) and hence a need to end growth. If we accept a need to end growth, we must also accept a need to address the distribution issue much more seriously.

Scale is the physical size of the economic subsystem relative to the ecosystem that contains and sustains it.

The Link Between Sustainable Scale and Just Distribution

Distribution is the apportionment of resources among different individuals. Why does ending growth require us to focus on distribution?

First, it seems pretty likely that the negative impacts of our excessive resource use will be worse for future generations than for our own. Thus, concern with scale involves a concern for future generations, or intergenerational distribution. Yet some 1.2 billion people alive today live in abject poverty, while many others have so much wealth they scarcely know what to do with it. It would be a peculiar set of ethical beliefs that would have us care about generations not yet born while ignoring the plight of the miserable today.

THINK ABOUT IT!

Why might excessive resource use have greater impacts on future generations than on the current one? Look back at the definition of Pareto efficient allocation. If the current generation is the de facto owner of all resources, could it be Pareto efficient for the current generation to consume fewer resources so that future generations are better off?

Second, as long as the economy is growing, we can always offer to the poor the future prospect of a slice of a larger pie. We do not need to redistribute now, some argue, because concentrated capital feeds the capitalist system, and if the poor remain patient, their misery will soon be

relieved. This is certainly a much more politically palatable option than redistribution, but as soon as we call for an end to growth, this option is gone. We certainly can't ask today's poor to sacrifice their hopes for a better future so that unborn generations will enjoy necessities of which they can only dream—especially when a reluctance to redistribute wealth today would suggest that the future generations for whom the poor are asked to sacrifice are likely to be someone else's children. Thus, distribution is of central importance to ecological economics.

Neoclassical economics is concerned almost solely with efficient allocation. Ecological economics also considers efficient allocation important, but it is secondary to the issues of scale and distribution. As we will see, in fact an efficient allocation cannot even be theoretically determined without a prior resolution to the distribution and scale questions. Typically that resolution is to take the existing distribution and scale as “given.”

Fortunately, as McNeill reminds us, it is only since the Depression that the growth fetish has taken control of economics. And as readers of this book will learn, if they don't know it already, there is a lot in economics that is true and useful—that is independent of the growth ideology, and that we could hardly do without. Indeed, as we shall show, the basic economic tools of optimization themselves provide the best means for arguing against the preoccupation with growth.

Why study economics, and in particular ecological economics? As we noted at the beginning of this chapter, economics is about what we want, and what we have to give up to get it. Growth is one more thing we may want, and like anything else, we have to give up something to get it. Ecological economists always ask if the extra growth is worth the extra sacrifice it entails. Neoclassical economists tend to forget this question, or to believe that the answer is always affirmative.

BIG IDEAS to remember

-
- | | |
|-----------------------------------|-----------------------------|
| ■ Ends and means | ■ Growth versus development |
| ■ Pareto efficient allocation | ■ Throughput |
| ■ Allocation, distribution, scale | ■ Coevolutionary economics |
-



The Fundamental Vision

■ THE WHOLE AND THE PART

Opportunity cost is the best alternative given up when a choice is made. For example, if a farmer cuts down a forest to expand his cropland, and if the consequent loss of timber, firewood, and water purification is the next best use of the land, then the value of timber, firewood, and water purification is the opportunity cost of the expanded cropland.

Ecological economics shares many concepts with conventional neoclassical economics. For example, both take as basic the concept of **opportunity cost**, defined as the best alternative that has to be sacrificed when you choose to do something. But ecological economics has a fundamentally different starting point—a different vision at its core of the way the world really is. To put it starkly, conventional economics sees the economy, the entire macroeconomy, as the whole. To the extent that nature or the environment are considered at all, they are thought of as parts or sectors of the macroeconomy—forests, fisheries, grasslands, mines, wells, ecotourist sites, and so on. Ecological economics, by contrast, envisions the macroeconomy as part of a larger enveloping and sustaining Whole—namely, the Earth, its atmosphere, and its ecosystems. The economy is seen as an open subsystem of that larger “Earthsystem.” That larger system is finite, nongrowing, and materially closed, although open to solar energy.

It is important to understand the distinctions among open, closed, and isolated systems. An **open system** takes in and gives out both matter and energy. The economy is such a system. A **closed system** imports and exports energy only; matter circulates within the system but does not flow through it. The Earth closely approximates a closed system. An **isolated system** is one in which neither matter nor energy enters or exits. It is hard to think of an example of an isolated system, except perhaps the universe as a whole. We say the Earth is approximately a closed system because it does not exchange significant amounts of matter with outer space—an occasional meteor comes in, an occasional rocket never returns, and we have a moon rock in a stained glass window in the National Cathedral. Maybe material exchanges will be greater someday, but so far they are negligible.

However, we do have a significant flow-through or throughput of energy in the form of incoming sunlight and exiting radiant heat. That throughput, like the ecosystem, is also finite and nongrowing. For the Earth, the basic rule is: Energy flows through, material cycles within.

Back to the problem of the whole and the part. Why is it so important? Because if the economy is the whole, then it can expand without limit. It does not displace anything and therefore incurs no opportunity cost—nothing is given up as a result of physical expansion of the macroeconomy into unoccupied space. But if the macroeconomy is a part, then its physical growth encroaches on other parts of the finite and nongrowing whole, exacting a sacrifice of something—an opportunity cost, as economists would call it. In this case, if we choose to expand the economy, the most important natural space or function sacrificed as a result of that expansion is the opportunity cost. The point is that growth has a cost. It is not free, as it would be if we were expanding into a void. The Earth-ecosystem is not a void, it is our sustaining, life-supporting envelope. It is therefore quite conceivable that at some point the further growth of the macroeconomy could cost us more than it is worth. Such growth is known as **uneconomic growth**. This leads to another insight that is fundamental to ecological economics and distinguishes it from conventional economics: Growth can be uneconomic as well as economic. There is an optimal scale of the macroeconomy relative to the ecosystem.¹ How do we know we have not already reached or passed it?

■ OPTIMAL SCALE

The idea of **optimal scale** is not strange to standard economists. It is the very basis of microeconomics. As we increase any activity, be it producing shoes or eating ice cream, we also increase both the costs and the benefits of the activity. For reasons we will investigate later, it is generally the case that after some point, costs rise faster than benefits. Therefore, at some point the extra benefits of growth in the activity will not be worth the extra costs. In economist's jargon, when the **marginal costs** (extra costs) equal the marginal benefits, then the activity has reached its optimal scale.² If we grow beyond the optimum, then costs will go up by more

¹Beyond optimal scale physical expansion becomes uneconomic growth, even if we misleadingly still call it "economic" growth. We use the word "economic" in two senses: (1) of or pertaining to the economy, and (2) yielding net benefits above costs. If the entity we call "the economy" physically grows, then we call that economic growth in sense 1. But growth in sense 1 may be economic or uneconomic in sense 2. Our linguistic habit of using sense 1 often leads us to prejudge the issue in terms of sense 2.

²"Marginal" means the last unit; in this case, the last unit of something obtained, produced, or consumed. Marginal cost (benefit) is the cost (benefit) of a very small increase in some activity.

than benefits. Subsequently, growth will make us poorer rather than richer. The basic rule of microeconomics, that optimal scale is reached when marginal cost equals marginal benefit ($MC = MB$), has aptly been called the “when to stop rule”—that is, when to stop growing. In macroeconomics, curiously, there is no “when to stop rule,” nor any concept of the optimal scale of the macroeconomy. The default rule is “grow forever.” Indeed, why not grow forever if there is no opportunity cost of growth? And how can there be an opportunity cost to growth of the macroeconomy if it is the whole?

Even if one adopts the basic vision of ecological economics and considers the economy as a subsystem of the ecosystem, there still would be no need to stop growing as long as the subsystem is very small relative to the larger ecosystem. In this “empty-world vision,” the environment is not scarce and the opportunity cost to expansion of the economy is insignificant. But continued growth of the physical economy into a finite and non-growing ecosystem will eventually lead to the “full-world economy” in which the opportunity cost of growth is significant. We are already in such a full-world economy, according to ecological economists.

This basic ecological economics vision is depicted in Figure 2.1. As growth moves us from the empty world to the full world, the welfare from economic services increases while the welfare from ecological services diminishes. For example, as we cut trees to make tables, we add the economic service of the table (holding our plates so we won't have to eat off the floor) and lose the ecological service of the tree in the forest (photosynthesis, securing soil against erosion, providing wildlife habitat, etc.). Traditionally, economists have defined capital as produced means of production, where produced implies “produced by humans.” Ecological economists have broadened the definition of capital to include the means of production provided by nature. We define capital as a stock that yields a flow of goods and services into the future. Stocks of manmade capital include our bodies and minds, the artifacts we create, and our social structures. **Natural capital** is a stock that yields a flow of natural services and tangible natural resources. This includes solar energy, land, minerals and fossil fuels, water, living organisms, and the services provided by the interactions of all of these elements in ecological systems.

We have two general sources of welfare: services of manmade capital (dark gray stuff) and services of natural capital (light gray stuff), as represented by the thick arrows pointing to “Welfare” in Figure 2.1. Welfare is placed outside the circle because it is a psychic, not a physical, magnitude (an experience, not a thing). Within the circle, magnitudes are physical. If we object to having a nonphysical magnitude in our basic picture of the economy on the grounds that it is metaphysical and unscientific, then we will have to content ourselves with the view that the economic system is

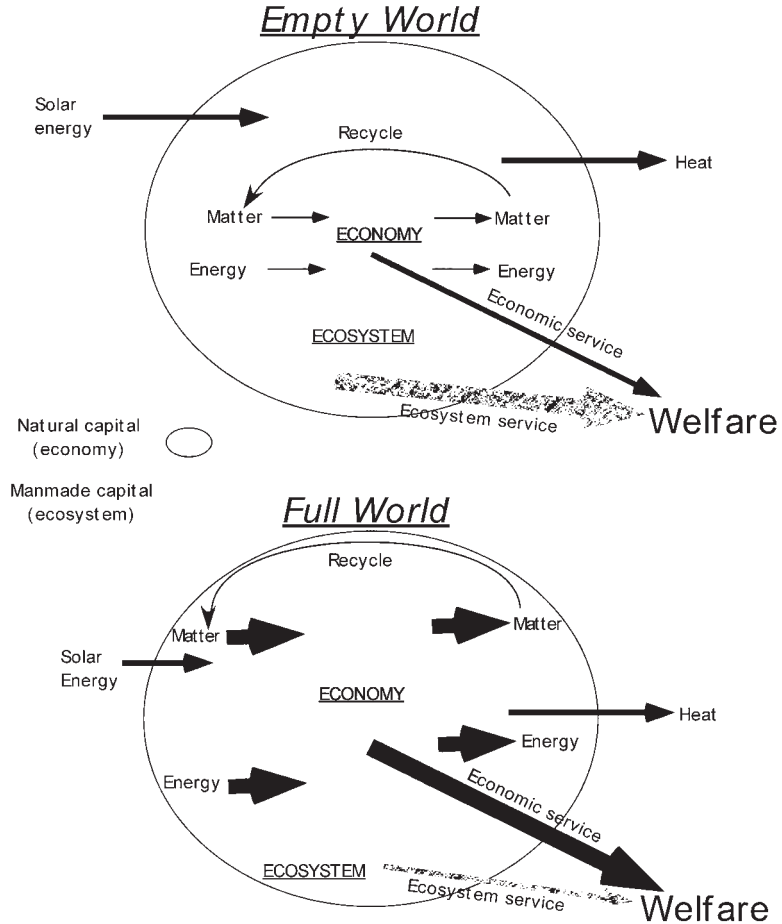


Figure 2.1 • From empty world to full world.

just an idiotic machine for turning resources into waste for no reason. The ultimate *physical* output of the economic process is degraded matter and energy—waste. Neglecting the biophysical basis of economics gives a false picture. But neglecting the psychic basis gives a meaningless picture. Without the concept of welfare or enjoyment of life, the conversion of material resources first into goods (production) and then into waste (consumption) must be seen as an end in itself—a pointless one. Both conventional and ecological economics accept the psychic basis of welfare, but they differ on the extent to which manmade and natural capital contribute to it.

■ DIMINISHING MARGINAL RETURNS AND UNECONOMIC GROWTH

As the economy grows, natural capital is physically transformed into man-made capital. More manmade capital results in a greater flow of services from that source. Reduced natural capital results in a smaller flow of services from that source. Moreover, as growth of the economy continues, the services from the economy grow at a decreasing rate. As rational beings, we satisfy our most pressing wants first, hence the law of diminishing marginal utility (to which we will return). As the economy encroaches more and more on the ecosystem, we must give up some ecosystem services. As rational beings, we presumably will sequence our encroachments so that we sacrifice the least important ecosystem services first. This is the best case, the goal. In actuality we fall short of it because we do not understand very well how the ecosystem works, and have only recently begun to think of ecosystem services as scarce. But the consequence of such rational sequencing is a version of the law of increasing marginal cost (to which we will return)—for each further unit of economic expansion, beyond some threshold, we must give up a more important ecosystem service. Marginal costs increase while marginal benefits decrease. At some point increasing marginal costs will equal declining marginal benefits.

Box 2-1 MARGINAL UTILITY VS. MARGINAL COST

- **Marginal utility:** The marginal utility of something is the additional benefit or satisfaction you derive from obtaining an additional unit of that thing. The **law of diminishing marginal utility** states that the more one has of something, the less satisfaction an additional unit provides. For example, the first slice of pizza on an empty stomach offers considerable satisfaction, but each additional slice provides less satisfaction than the previous one.
- **Marginal cost:** Marginal cost is the additional cost of producing one more unit. The **law of increasing marginal cost** is similar to that of diminishing marginal utility. For each additional ton of wheat harvested, you have to make use of inferior land and workers (you used the best first). Also, once you've used all the land for wheat, adding more labor, fertilizer, and so on is the only way to increase the wheat harvest. But with fixed land, we will have diminishing returns to the variable factors (labor, fertilizer)—more and more laborers and fertilizer will be required for each additional ton of harvest. Diminishing returns is a further reason for increasing marginal costs. Neoclassical economics is constantly comparing increasing marginal costs with declining marginal benefits, looking for their point of intersection that

defines the optimal scale of each microeconomic activity. It does not apply this logic to the macroeconomy, or recognize that it has an **optimal scale**. Ecological economics insists that the logic of optimal scale is relevant to the entire macroeconomy, as well as to its parts.

This first step in analyzing the core or preanalytic vision of ecological economics can be expressed graphically (Figure 2.2). The basic logic goes back to William Stanley Jevons (1871) and his analysis of labor supply in terms of balancing the marginal utility of wages with the marginal disutility of labor to the worker. Put another way, Jevons asked: When does the effort of working begin to exceed the value of the wage to the worker? Ecological economists ask: When does the cost to all of us of displacing the Earth's ecosystems begin to exceed the value of the extra wealth produced? In Figure 2.2, the marginal utility (MU) curve reflects the dimin-

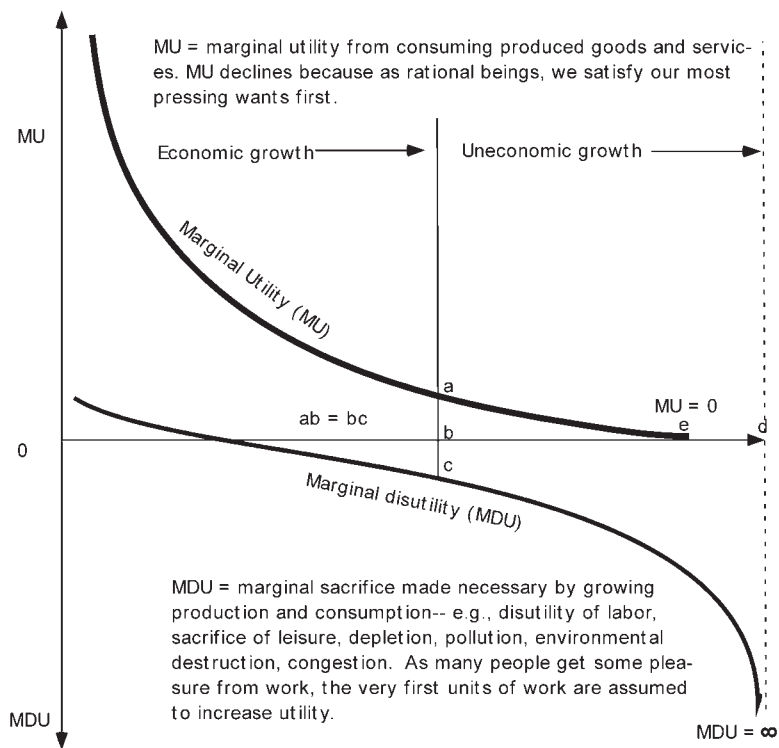


Figure 2.2 • Limits to growth of the macroeconomy. Point b = economic limit or optimal scale, where marginal utility (MU) = marginal disutility (MDU) (maximum net positive utility); e = futility limit, where MU = 0 (consumer satiation); d = catastrophe limit, where MDU = infinity. At point d, we have gone beyond sustainable scale.

ishing marginal utility of additions to the stock of manmade capital. The marginal disutility (MDU) curve reflects the increasing marginal cost of growth (sacrificed natural capital services and disutility of labor), as more natural capital is transformed into manmade capital. The optimal scale of the macroeconomy (economic limit to growth) is at point d, where $MU = MDU$, or where $ab = bc$, and net positive utility is a maximum (the area under the MU curve minus the area above the MDU curve).

Two further limits are noted: point e where $MU = 0$ and further growth is futile even with zero cost; and point d, where an ecological catastrophe is provoked, driving MDU to infinity. For example, some relatively unnecessary chemical pollutant in sufficient amounts might unexpectedly prohibit photosynthesis, rendering plant life incapable of capturing the solar energy upon which all life depends. These “outer limits” need not occur in the order depicted. We could have an ecological disaster before reaching satiation. The diagram shows that growth out to point b is literally economic growth (benefiting us more than it costs), while growth beyond point b is literally uneconomic growth (costing us more than it benefits us). Beyond point b, GNP, “that which seems to be wealth,” does indeed become “a gilded index of far-reaching ruin,” as John Ruskin predicted over a century ago.³ The nice thing about point b, the economic limit, is that it occurs first, allowing us to maximize net benefits while stopping us from destroying the capacity of the Earth to support life.

The concepts of optimal scale and uneconomic growth have a universal logic—they apply to the macroeconomy just as much as to microeconomic units.⁴ How did we come to forget this in macroeconomics? How did we come to ignore the existence of the MDU curve and the issue of optimal scale of the macroeconomy? We suggest two possibilities. One is the “empty-world vision” that recognizes the concept of uneconomic growth, but claims that we are not yet at that point; neoclassical economists tend to think that MU is still very large, and MDU is still negligible. In this case we can look at the factual evidence to resolve the difference, as will be done later.

The other possibility for explaining the total neglect of the costs of growth is a paradigm difference: The economy is simply not seen as a subsystem of the ecosystem, but rather the reverse—the ecosystem is a subsystem of the economy (Figure 2.3). Here we are discussing different

³J. Ruskin, *Unto This Last*, (1862) in Lloyd J. Hubenka, ed., *Four Essays on the First Principles of Political Economy*, Lincoln: University of Nebraska Press, 1967.

⁴It is a mistake to think that microeconomics is about little things and macroeconomics is about big things. Microeconomics means the economics of the part, macroeconomics means the economics of the whole or aggregate. Parts can be big, aggregates can be small. Although MB = MC is a rule of microeconomic analysis, we can apply it to something big, the economic subsystem, as long as the big thing is a part, not the whole.

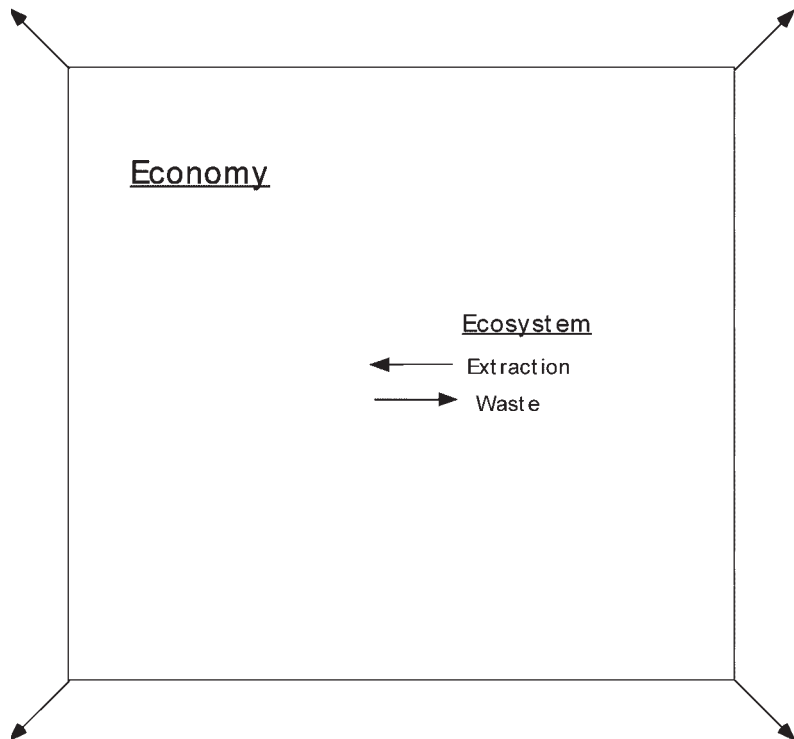


Figure 2.3 • The ecosystem as a subsystem of the economy.

conceptual worlds, and no empirical measurements will resolve the difference.

In the vision of Figure 2.3, the ecosystem is merely the extractive and waste disposal sector of the economy. Even if these services become scarce, growth can still continue forever because technology allows us to “grow around” the natural sector by the substitution of manmade for natural capital, following the dictates of market prices. Nature is, in this view, nothing but a supplier of various indestructible building blocks, which are substitutable and superabundant. The only limit to growth, in this view, is technology, and since we can always develop new technologies, there is no limit to economic growth. The very notion of “uneconomic growth” makes no sense in that paradigm. Since the economy is the whole, the growth of the economy is not at the expense of anything else; there is no opportunity cost to growth. On the contrary, growth enlarges the total to be shared by the different sectors.⁵ Growth does not increase

⁵A note of caution: the dark gray stuff in Figure 2.1 is in physical dimensions. The dark gray stuff in Figure 2.3 is probably thought of by neoclassical economists as GNP; it is in units of value and therefore not strictly physical. But value is price times quantity, and the latter has an irreducible physical component. Indeed it is mainly changes in that physical component that economists seek to measure in calculating *real* GNP—i.e., changes in GNP not due to changes in prices.

the scarcity of anything; rather, it diminishes the scarcity of everything. How can one possibly oppose growth?

■ A PARADIGM SHIFT

Where conventional economics espouses growth forever, ecological economics envisions a steady-state economy at optimal scale. Each is logical within its own preanalytic vision, and each is absurd from the viewpoint of the other. The difference could not be more basic, more elementary, or more irreconcilable.

In other words, ecological economics calls for a “paradigm shift” in the sense of philosopher Thomas Kuhn,⁶ or what we have been calling, following economist Joseph Schumpeter,⁷ a change in preanalytic vision. We need to pause to consider more precisely just what these concepts mean. Schumpeter observes that “analytic effort is of necessity preceded by a preanalytic cognitive act that supplies the raw material for the analytic effort” (p. 41). Schumpeter calls this preanalytic cognitive act “Vision.” One might say that vision is the pattern or shape of the reality in question that the right hemisphere of the brain abstracts from experience, and then sends to the left hemisphere for analysis. Whatever is omitted from the preanalytic vision cannot be recaptured by subsequent analysis. Correcting the vision requires a new preanalytic cognitive act, not further analysis of the old vision. Schumpeter notes that changes in vision “may reenter the history of every established science each time somebody teaches us to see things in a light of which the source is not to be found in the facts, methods, and results of the preexisting state of the science.” (p. 41). It is this last point that is most emphasized by Kuhn (who was apparently unaware of Schumpeter’s discussion).

Kuhn distinguished between “normal science,” the day-to-day solving of puzzles within the established rules of the existing preanalytic vision, or “paradigm” as he called it, and “revolutionary science,” the overthrow of the old paradigm by a new one. It is the common acceptance by scientists of the reigning paradigm that makes their work cumulative, and that separates the community of serious scientists from quacks and charlatans. Scientists are right to resist scientific revolutions. Most puzzles or anomalies, after all, do eventually get solved, one way or another, within the existing paradigm. And it is unfortunate when people who are too lazy to master the existing scientific paradigm seek a shortcut to fame by summarily declaring a “paradigm shift” of which they are the leader. Nevertheless, as Kuhn demonstrates, paradigm shifts, both large and small, are

⁶T. Kuhn, *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 1962.

⁷J. Schumpeter, *History of Economic Analysis*, New York: Oxford University Press, 1954.

undeniable episodes in the history of science—the shift from the Ptolemaic (Earth-centered) to the Copernican (sun-centered) view in astronomy, and Newton’s notions of absolute space and time versus Einstein’s relativity of space and time—are only the most famous. As Kuhn demonstrates, there does come a time when sensible loyalty to the existing paradigm becomes stubborn adherence to intellectual vested interests.

Paradigm shifts are obscured by textbooks whose pedagogical organization is, for good reason, logical rather than historical.⁸ Physics students would certainly be unhappy if, after learning in the first three chapters all about the ether and its finely grained particles, they were suddenly told in Chapter 4 to forget all that stuff about the ether because we just had a Newtonian paradigm shift and now accept action at a distance unmediated by fine particles (gravity)!

Thirty years ago, a course in the history of economic thought was required in all graduate economics curricula. Today such a course is usually not even available as an elective. This is perhaps a measure of the (over)confidence economists have in the existing paradigm. Why study the errors of the past when we now know the truth? Consequently, the several changes in preanalytic vision in the history of economic thought are unknown to students and to many of their professors.

A change in vision from seeing the economy as the whole to seeing it as a part of the relevant Whole—the ecosystem—constitutes a major paradigm shift in economics. In subsequent chapters, we will consider more specific consequences of this shift.

The Circular Flow and the Linear Throughput

Differing preanalytic visions lead to a few basic analytical differences as well, although many tools of analysis remain the same between standard and ecological economics, as we’ll discuss later.

Given that standard economics has a preanalytic vision of the economy as the whole, what is its first analytic step in studying this whole? It is depicted in Figure 2.4, the familiar circular flow diagram with which all basic economics texts begin. In this view, the economy has two parts: the production unit (firms) and the consuming unit (households). Firms produce and supply goods and services to households; households demand goods and services from firms. Firm supply and household demand meet in the goods market (lower loop), and prices are determined there by the interaction of supply and demand.

⁸Textbooks are designed to initiate the student into the reigning paradigm as efficiently as possible. Chapter 2 builds on Chapter 1, Chapter 3 builds on Chapters 1 and 2, etc. Efficient pedagogy is logical and cumulative. But the history of science is not so tidy. In history there are times when we have to throw out earlier chapters and start over. This textbook is not immune to this danger, although we have tried to be sensitive to it.

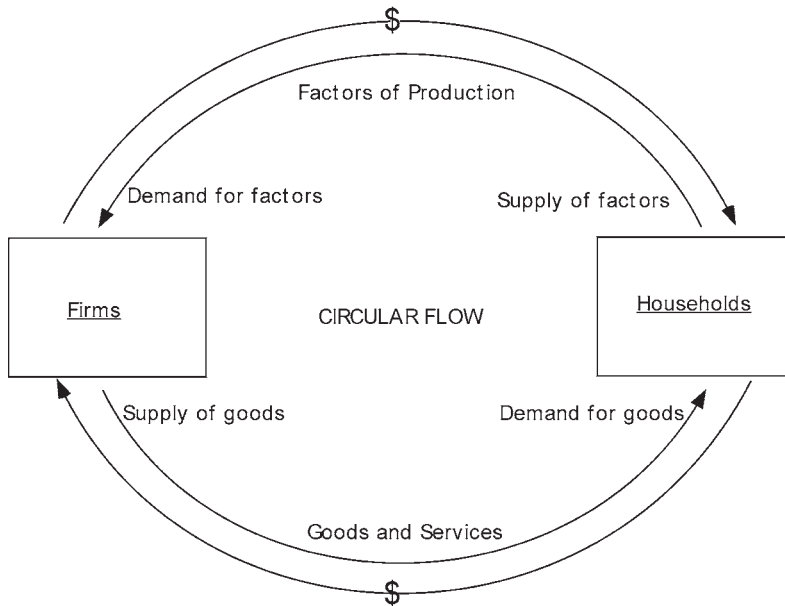


Figure 2.4 • The circular flow of the economy.

At the same time, firms demand factors of production from the households, and households supply factors to the firms (upper loop). Prices of factors (land, labor, capital) are determined by supply and demand in the factors market. These factor prices, multiplied by the amount of each factor owned by a household, determine the income of the household. The sum of all these factor incomes of all the households is National Income. Likewise, the sum of all goods and services produced by firms for households, multiplied by the price at which each is sold in the goods market, is equal to National Product. By accounting convention, National Product must equal National Income. This is so because profit, the value of total production minus the value of total factor costs, is counted as part of National Income.

THINK ABOUT IT!

Would the equality still hold if profits were negative? Explain.

The upper and lower loops are thus equal, and in combination they form the circular flow of exchange value. This is a very important vision. It unifies most of economics. It shows the fundamental relationship between production and consumption. It is the basis of microeconomics, which studies how the supply-and-demand plans of firms and households emerge from their goals of maximizing profits (firms) and maximizing utility (households). It shows how supply and demand interact under different

market structures to determine prices, and how price changes lead to changes in the allocation of factors to produce a different mix of goods and services. In addition, the circular flow diagram also provides the basis for macroeconomics—it shows how the aggregate behavior of firms and households determines both National Income and National Product.

■ SAY'S LAW: SUPPLY CREATES ITS OWN DEMAND

The equality of National Income and National Product, as mentioned, guarantees that there is always enough purchasing power in the hands of households in the aggregate to purchase the aggregate production of firms. Of course, if some firms produce things households do not want, the prices of those things will fall, and if they fall below what it cost to produce them, those firms will make losses and go out of business. The circular flow does not guarantee that all firms will sell whatever they produce at a profit. But it does guarantee that such a result is not impossible because of an overall glut of production in excess of overall income. This comforting feature of the economy is known as **Say's Law**—supply creates its own demand. For a long time, economists believed Say's Law ruled out any possibility of long-term and substantial unemployment, such as occurred during the Great Depression. However, the experience of the Depression led John Maynard Keynes to reconsider Say's Law and the comforting conclusion of the circular flow vision.

There may indeed always be enough income generated by production to purchase what is produced. But there is no guarantee that all the income will be spent, or spent in the current time period, or spent on goods and services, or spent in the national market. In other words, there are leakages out of the circular flow. There are also corresponding injections into the circular flow. But there is no guarantee that the leakages and injections will balance each other.

■ LEAKAGES AND INJECTIONS

What are these leakages and injections? One leakage from the expenditure stream is savings. People refrain from spending now in order to be able to spend later. The corresponding injection is investment. Investment results in expenditure now, but increased production only in the future. Thus, the circular flow can be restored if Saving equals Investment. This recycling of savings into investment is accomplished through financial markets and interest rates. In Figure 2.5, the shaded rectangles represent the financial institutions that collect savings and lend to investors.

A second leakage from the circular flow is payment of taxes. The corresponding injection is government expenditure. The rectangle represents

the institutions of public finance. Public finance policies can balance the taxes and government spending, or intentionally unbalance them to compensate for imbalances in Savings and Investment. For example, if saving exceeds investment, the government might avoid a recession by allowing government expenditures to exceed taxes by the same amount.

The third leakage from the national circular flow is expenditure on imports. The corresponding injection is expenditure by foreigners for our exports. International finance and foreign exchange rates are mechanisms for balancing exports and imports. Again the corresponding institutions are represented by the rectangle. The circular flow is restored if the sum of leakages equals the sum of injections, that is, if savings plus taxes plus imports equals investment plus government spending plus exports. If the sum of leakages is greater than the sum of injections, unemployment or deflation tends to result. If the sum of injections is greater than the sum of leakages, we tend to have either expansion or inflation.

Leakages and injections are shown in Figure 2.5, an expanded circular flow

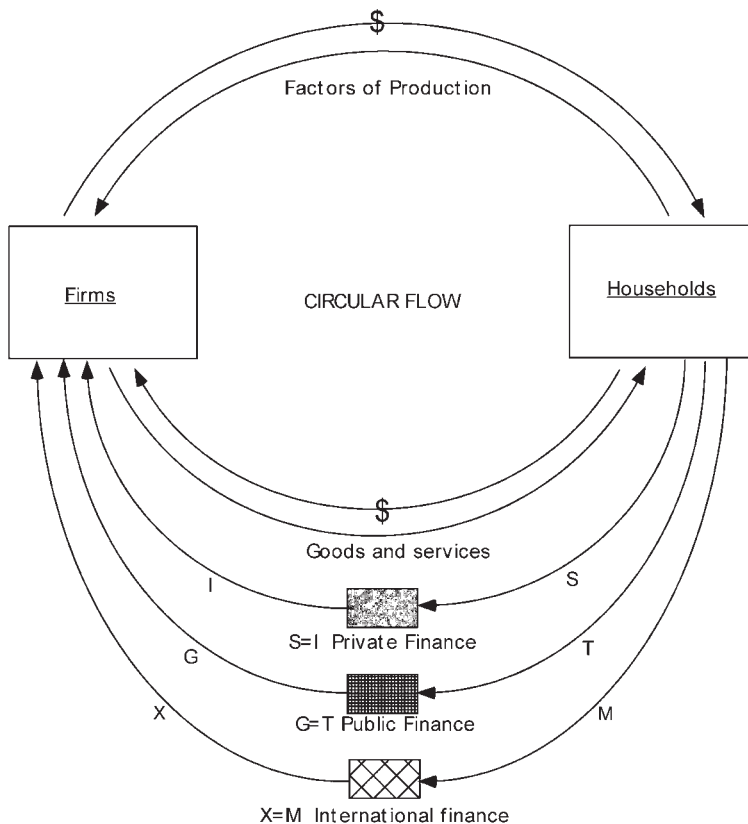


Figure 2.5 • The circular flow with leakages and injections. S = savings, I = investment, G = government expenditure, T = taxes, X = exports, M = imports.

flow diagram. For simplicity we have assumed that households are net savers, net taxpayers, and net importers, while firms are net investors, net recipients of government expenditure, and net exporters.

The circular flow diagram unites not only micro- and macroeconomics, but also shows the basis for monetary, fiscal, and exchange rate policy in the service of maintaining the circular flow so as to avoid unemployment and inflation. With so much to its credit, how could one possibly find fault with the circular flow vision?

There is no denying the usefulness of the circular flow model for analyzing the flow of exchange value. However, it has glaring difficulty as a description of a real economy. Notice that the economy is viewed as an isolated system. Nothing enters from outside the system; nothing exits the system to the outside. But what about all the leakages and injections just discussed? They are just expansions of the isolated system that admittedly make the concept more useful, but they do not change the fact that nothing enters from outside and nothing exits to the outside. The whole idea of analyzing leakages and injections is to be able to reconnect them and close the system again. Why is the isolated system a problem? Because an isolated system has no outside, no environment. This is certainly consistent with the view that the economy is the whole. But a consequence is that there is no place from which anything can come, or to which it might go. If our preanalytic vision is that the economy is the whole, then we cannot possibly analyze any relation of the economy to its environment. The whole has no environment.

What is it that is really flowing around and around in a circle in the circular flow vision? Is it really physical goods and services, and physical laborers and land and resources? No. It is only abstract exchange value, the purchasing power represented by these physical things.⁹ The “soul” embodied in goods by the firms is abstract exchange value. When goods arrive to the households, the “soul” of exchange value jumps out of its embodiment in goods and takes on the body of factors for its return trip to the firms, whereupon it jumps out of the body of factors and reincorporates itself once again into goods, and so on. But what happens to all the discarded bodies of goods and factors as the soul of exchange value transmigrates from firms to households and back *ad infinitum*? Does the

⁹We are careful to say “abstract exchange value” rather than “money” because not even money in the sense of currency can circulate as an isolated system. Money wears out and has to be replaced by new money. The physical wear and tear of hand-to-hand circulation means that even money has to have a throughput to maintain its circulation. Because fractional money circulates more rapidly than notes of higher denomination, we usually adopt metal coins rather than paper to withstand the higher velocity of circulation of small denominations. For this reason, as inflation has eroded its value, the U.S. Treasury has periodically attempted to issue the dollar in coin form, though without much success.

system generate wastes? Does the system require *new* inputs of matter and energy? If not, then the system is a perpetual motion machine, a contradiction to the Second Law of Thermodynamics (about which more later). If it is not to be a perpetual motion machine (a perfect recycler of matter and energy), then wastes must go somewhere and new resources must come from somewhere outside the system. Since there is no such thing as perpetual motion, the economic system cannot be the whole. It must be a subsystem of a larger system, the Earth-ecosystem.

The circular flow model is in many ways enlightening, but like all abstractions, it illuminates only what it has abstracted out of reality and leaves in darkness all that has been abstracted from. What has been abstracted from, left behind, in the circular flow model is the linear throughput of matter-energy by which the economy lives off its environment. Linear throughput is the flow of raw materials and energy from the global ecosystem's sources of low entropy (mines, wells, fisheries, croplands), through the economy, and back to the global ecosystem's sinks for high entropy wastes (atmosphere, oceans, dumps). The circular flow vision is analogous to a biologist describing an animal only in terms of its circulatory system, without ever mentioning its digestive tract. Surely the circulatory system is important, but unless the animal also has a digestive tract that connects it to its environment at both ends, it will soon die either of starvation or constipation. Animals live from a metabolic flow—an entropic throughput from and back to their environment. The **law of entropy** states that energy and matter in the universe move inexorably toward a less ordered (less useful) state. An entropic flow is simply a flow in which matter and energy become less useful; for example, an animal eats food and secretes waste, and cannot ingest its own waste products. The same is true for economies. Biologists, in studying the circulatory system, have not forgotten the digestive tract. Economists, in focusing on the circular flow of exchange value, have entirely ignored the metabolic throughput. This is because economists have assumed that the economy is the whole, while biologists have never imagined that an animal was the whole, or was a perpetual motion machine.

■ LINEAR THROUGHPUT AND THERMODYNAMICS

The linear throughput is in physical units and is strictly subject to the laws of conservation of mass and energy, and the law of entropy. The circular flow is in units of abstract exchange value and is not subject to any obvious physical limits. The circular flow can nominally grow forever by virtue of inflation, but we set this case aside to ask if the real economic value in the sense of satisfying wants, of qualitative development, can grow forever.

The Fallacy of Misplaced Concreteness

Obviously, a model that abstracts from the environment and considers the economy in isolation from it cannot shed any light on the relation of the economy to the environment. This kind of mistake was given a name by the philosopher and mathematician Alfred North Whitehead. He called it the **fallacy of misplaced concreteness**. By that he meant the error of mistaking the map for the territory, the error of treating an abstract model, made with the purpose of understanding one aspect of reality, as if it were adequate for understanding everything, or entirely different things, things that had been abstracted from in making the model. Whitehead was no enemy of abstract thought. He emphasized that we cannot think without abstraction. All the more important, therefore, to be aware of the limits of our abstractions. The power of abstract thought comes at a cost. The fallacy of misplaced concreteness is to forget that cost.

Let's take a closer look at what standard economists have abstracted from in the circular flow model—namely, the throughput, the metabolic flow from raw material inputs to waste outputs. The throughput is in physical units. Consequently, the laws of physics apply strictly to it.

By the **First Law of Thermodynamics**, the conservation of matter and energy, we know that throughput is subject to a balance equation: Input

The First Law of Thermodynamics states that neither matter nor energy can be either created or destroyed.

Box 2-2 THE LAWS OF THERMODYNAMICS

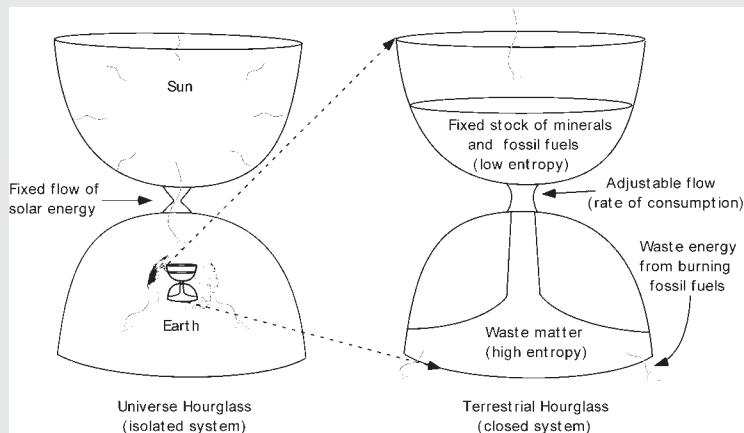


Figure 2.6 • The entropy hourglass (based on Georgescu-Roegen).

The hourglass on the left is an isolated system; no sand enters, no sand exits. Also, within the hourglass there is neither creation nor destruction of sand; the amount of sand in the hourglass is constant. This, of course,

is the analog of the First Law of Thermodynamics, the conservation of matter and energy. Finally, there is a continual running down of sand in the top chamber and an accumulation of sand in the bottom chamber. Sand in the bottom chamber has used up its potential to fall and thereby do work. It is high-entropy or unavailable matter-energy. Sand in the top chamber still has potential to fall; it is low-entropy or available matter-energy (still useful). This is the analogy of the Second Law of Thermodynamics: Entropy, or “used-up-ness,” increases in an isolated system. The hourglass analogy is particularly apt because entropy is “time’s arrow” in the physical world—that is, we can tell earlier from later by whether or not entropy has increased. However, unlike a real hourglass, the entropy hourglass cannot be turned upside down!

With a bit of license, we can extend the basic analogy by considering the sand in the upper chamber to be the stock of low-entropy fossil fuel on Earth, depicted in the right-hand figure. Fossil energy is used at a rate determined by the constricted middle of the hourglass, but unlike a normal hourglass, humans alter the width (i.e., they change the rate of consumption of fossil fuels). Once consumed, the sand falls to the bottom of the chamber, where it accumulates as waste and interferes with terrestrial life processes.

To represent solar energy, the top of the hourglass on the left would be vast (from the human perspective) as would the bottom; solar energy, too, ends as waste heat, but it is not confined to the Earth. It does not disappear, but it radiates into outer space, and unlike waste matter does not accumulate on Earth. The constricted middle, however, would be quite small, and humans would be unable to adjust it. The solar source of low entropy is stock-abundant but flow-limited. In other words, there is a lot of it, but we get only a little at a time. The terrestrial source is stock-limited but flow-abundant, until the stock runs out. The asymmetry is important. With industrialization we have come to depend more and more on the least abundant source of low entropy. However convenient in the short run, this will be uneconomic in the long run.

equals output plus accumulation. If there is accumulation, the economic subsystem is growing. In steady-state equilibrium, growth and accumulation would be zero, and input flow would equal output flow. In other words, all raw material inputs eventually become waste outputs. The throughput has two ends: depletion of environmental sources, and pollution of environmental sinks. Ignoring throughput is the same as ignoring depletion and pollution. Unlike exchange value, the flow of throughput is not circular; it is a one-way flow from low-entropy sources to high-entropy sinks. This is a consequence of the **Second Law of Thermodynamics**, the entropy law. We can recycle materials, but never 100%; recycling is a circular eddy in the overall one-way flow of the river. Energy, by the entropy law, is not recyclable at all. More precisely, it is recyclable, but

The *Second Law of Thermodynamics* states that entropy never decreases in an isolated system. Although matter and energy are constant in quantity (First Law), they change in quality. The measure of quality is entropy, and basically it is a physical measure of the degree of “used-up-ness” or randomization of the structure or capacity of matter or energy to be useful to us. Entropy increases in an isolated system. We assume the universe is an isolated system, so the Second Law says that the natural, default tendency of the universe is “shuffling” rather than “sorting.” In everyday terms, left to themselves, things tend to get mixed up and scattered. Sorting does not occur by itself.

it always takes more energy to do the recycling than the amount that can be recycled. Thus, recycling energy is not physically impossible but always economically a loser—regardless of the price of energy. No animal can directly recycle its own waste products as its own food. If it could, it would be a perpetual motion machine. In strict analogy, no economy can function by directly reusing only its own waste products as raw materials.

The circular flow diagram gives the false impression that the economy is capable of direct reuse. Some very good textbook writers have explicitly affirmed this false impression. For example, Heilbroner and Thurow,¹⁰ in a standard economics text, tell us that “the flow of output is circular, self-renewing, self-feeding.” In other words the economy is a perpetual motion machine. To drive the point home, the first study question at the end of the chapter is, “Explain how the circularity of the economic process means that the outputs of the system are returned as fresh inputs.” It would have been reasonable to ask how dollars spent reappear as dollars earned in the circular flow of exchange value, and how purchasing power is regenerated in the act of production. But explaining how outputs are returned as inputs, indeed *fresh* inputs, requires the student to discover the secret of perpetual motion! Of course, the authors do not really believe in perpetual motion; they were trying to get across to the student the importance of replenishment—how the economic process reproduces itself and keeps going for another round. Certainly this is an important idea to stress, but the key to understanding it is precisely that replenishment must come from *outside* the economic system. This is a point conventional economists tend to neglect, and it leads to the mistaking of the part for the whole. If the economy is the whole, it has no outside; it is an isolated system.

The error in the text cited is fundamental, but not unique. It is representative of most standard texts. Heilbroner and Thurow have the virtue of clear expression—a virtue that makes it easier to spot errors. Other texts leave the student with the same erroneous impression, but without forthrightly stating the implication in words that cause us to think again. Nor is the error confined to standard economists. Karl Marx’s models of simple and expanded reproduction are also isolated circular flow models. Marx, with his theory that labor was the source of all value, was even more eager than standard economists to deny any important role to nature in the functioning of the economy and creation of value. For Marx, the idea that nature embodied scarcity was an abomination. All poverty was the result of unjust social relations, or class exploitation, not the “niggardliness of nature.” Thomas Malthus had argued that overpopulation relative to

¹⁰R. Heilbroner and L. Thurow, *The Economic Problem*, New York: Prentice-Hall, 1981, pp. 127, 135.

natural capacities was also an independent cause of poverty, and that social revolution could not eliminate poverty. Marx felt that Malthus' ideas were a threat to his, and treated him with contempt and vituperation. Political debates between neo-Marxists and neo-Malthusians continue to this day.¹¹

The Importance of Throughput

Let's turn from the theoretical reasons for the importance of throughput to an empirical look at its size and composition in modern economies. The following paragraph is from a book about the dependence of the economy on the environment:

Researchers have calculated that industry moves, mines, extracts, shovels, burns, wastes, pumps, and disposes of 4 million pounds of material in order to provide one average middle-class American family's needs for a year. In 1990, the average American's economic and personal activities mobilized a flow of roughly 123 dry-weight pounds of material per day—equivalent to a quarter of a billion semitrailer loads per year. This amounts to 47 pounds of fuel, 46 of construction materials, 15 of farmland, 6 of forest products, 6 of industrial minerals, and 3 of metals of which 90% is iron and steel. Net of 6 pounds of recycled materials, that Average American's daily activities emitted 130 pounds of gaseous material into the air, created 45 pounds of material artifacts, generated 13 pounds of concentrated wastes, and dissipated 3.5 pounds of nongaseous wastes into the environment in such scattered forms as pesticides, fertilizers, and crumbs of material rubbed off tires. In addition, the person's daily activities required the consumption of about 2,000 pounds of water that after use is sufficiently contaminated that it cannot be reintroduced into marine or riparian systems, and produced 370 pounds of rock, tailings, overburden, and toxic water as a result of extracting oil, gas, coal, and minerals. . . . In sum, Americans waste or cause to be wasted nearly 1 million pounds of materials per person per year.¹²

That's a lot of throughput to abstract from—to leave out of our model! It all ends up as waste, but necessary waste to support our population at our standard of consumption, with our present technology. Better technologies, as well as a better ordering of our priorities, can reduce the throughput without lowering the quality of life. However, by how much, and by what policies, are big issues in ecological economics.

In 1997, a coalition including the World Resources Institute (WRI, U.S.), the Wuppertal Institute (Germany), the Netherlands Ministry for

¹¹See H. Daly, A Marxian-Malthusian View of Poverty and Exploitation, *Population Studies*, May 1971.

¹²P. Hawken, A. Lovins, and H. Lovins, *Natural Capitalism*, Boston: Little, Brown, 1999, pp. 51–52.

Housing Spatial Planning and Environment, and the National Institute for Environmental Studies (Japan) attempted to measure throughput in each of their industrial countries for the period 1975–1993.¹³ Their basic finding was that total material requirements (per-capita annual flows) for each of the four countries did not change much between 1975 and 1993. The range was 45–85 metric tons of natural resources per person per year, with the U.S. at the high end, Japan at the low end, and Germany and the Netherlands in between. Over the period, the U.S. flows declined slightly and those of the other countries rose slightly. Most of the decline in U.S. requirements was accounted for by better soil erosion control, not better industrial efficiency. The roughly constant total resource requirements over time are the product of a declining resource requirement per dollar of gross domestic product (GDP) with a growing number of dollars of GDP, in all four countries. We have become more *efficient* but not more *frugal*. It is as if we developed cars that got twice as many miles per gallon and then drove twice as many miles, thereby burning the same number of gallons.

THINK ABOUT IT!

Which goal do you think should come first, efficiency or frugality? We will come back to this, but maybe you can answer it already.

While it is important to have empirical information on the physical size, composition, and change over time in the throughput, we also must have some basis for judging the environmental costs of these flows. How large are they relative to the capacity of the ecosystem to absorb and regenerate them? Exactly what opportunity costs do these flows inflict on us? On other species? Partial answers are given by the World Wildlife Fund (WWF):¹⁴

While the state of the Earth's natural ecosystems has declined by about 33 per cent over the last 30 years ("Living Planet Index"), the ecological pressure of humanity on the Earth has increased by about 50 per cent over the same period ("World Ecological Footprint"), and exceeds the biosphere's regeneration rate.

In terms of our "economy as subsystem" diagram (Figure 2.1), this means that the capacity of natural capital (light gray stuff) to supply life-support services has declined by about 33%, and that the demand generated by manmade capital (dark gray stuff) for life-support services, provided by the light gray stuff, has increased by about 50%—and this has occurred over the past 30 years. There are two blades to this scissors:

¹³Resource Flows: The Material Basis of Industrial Economies, Washington, DC: WRI, 1997.

¹⁴World Wildlife Fund, UNEP, *Living Planet Report 2000*, Gland, Switzerland: WWF International, 2000, p. 1.

increasing demand for carrying capacity (ecological footprint), and decreasing supply of carrying capacity (living planet index). Both blades of the scissors are being squeezed by the same hand—namely, growth. The “ecological footprint” is the number of hectares of productive land or sea required to support one average person at the world average consumption level. The study estimates that as of 1997 the ecological footprint of the Earth’s total population was at least 30% higher than the Earth’s biological reproductive capacity. This deficit is made up by consuming or drawing down natural capital, thus “borrowing from” or perhaps “robbing” the future. Scholars may have statistical arguments over the best measures of carrying capacity demanded and supplied, but the basic qualitative conclusion of unsustainable trends is hard to deny.

BIG IDEAS to remember

-
- | | |
|--------------------------------------|-------------------------------------|
| ■ Whole and part | ■ Linear throughput |
| ■ Open, closed, and isolated systems | ■ Say’s Law |
| ■ Optimal scale | ■ Leakages and injections |
| ■ Full world versus empty world | ■ Laws of thermodynamics |
| ■ Diminishing marginal utility | ■ Fallacy of misplaced concreteness |
| ■ Increasing marginal costs | ■ Entropy hourglass |
| ■ Paradigm and preanalytic vision | ■ Measures of throughput volume |
| ■ Circular flow | |
-



Ends, Means, and Policy

■ ENDS AND MEANS—A PRACTICAL DUALISM

Ecological economics has at least as much in common with standard economics as it has differences. One important common feature is the basic definition of economics as the study of the allocation of scarce means among competing ends. There are disagreements about what is scarce and what is not, what are appropriate mechanisms for allocating different resources (means), and about how we rank competing ends in order of importance—but there is no dispute that using means efficiently in the service of ends is the subject matter of economics. Using means in the service of ends implies policy. Alternatively, policy implies knowledge of ends and means. Economics, especially ecological economics, is inescapably about policy, although the rarefied levels of abstraction sometimes reached by economists may lead us to think otherwise.

If economics is the study of the allocation of scarce means in the service of competing ends, we have to think rather deeply about the nature of ends and means. Also, policy presupposes knowledge of two kinds: of possibility and purpose; of means and ends. Possibility reflects how the world works. In addition to keeping us from wasting time and money on impossibilities, this kind of knowledge gives us information about trade-offs among real alternatives. Purpose reflects desirability, our ranking of ends, our criteria for distinguishing better from worse states of the world. It does not help much to know how the world works if we cannot distinguish better from worse states of the world. Nor is it useful to pursue a better state of the world that happens to be impossible. Without both kinds of knowledge, policy discussion is meaningless.

To relate this to economic policy, we need to consider two questions. First, in the realm of possibility, the question is: What are the means at our

disposal? Of what does our ultimate means consist? By “ultimate means” we mean a common denominator of possibility or usefulness that we can only use up and not produce, for which we are totally dependent on the natural environment. Second, what ultimately is the end or highest purpose in whose service we should employ these means? These are very large questions, and we cannot answer them completely, especially the latter. But it is essential to raise the questions. There are some things, however, that we say by way of partial answers, and it is important to say them.

Means

Ultimate means, the common denominator of all usefulness, consist of low-entropy matter-energy.¹ Low-entropy matter-energy is the physical coordinate of usefulness; the basic necessity that humans must use up but cannot create, and for which the human economy is totally dependent on nature to supply. Entropy is the qualitative physical difference that distinguishes useful resources from an equal quantity of useless waste. We do not use up matter and energy *per se* (First Law of Thermodynamics), but we do irrevocably use up the quality of usefulness as we transform matter and energy to achieve our purposes (Second Law of Thermodynamics). All technological transformations require a before and after, a gradient or metabolic flow from concentrated source to dispersed sink, from high to low temperature.² The capacity for entropic transformations of matter-energy to be useful is therefore reduced both by the emptying of finite sources and by the filling up of finite sinks. If there were no entropic gradient between source and sink, the environment would be incapable of serving our purposes or even sustaining our lives. Technical knowledge helps us use low entropy more efficiently; it does not enable us to eliminate or reverse the direction of the metabolic flow.

Matter can of course be recycled from sink back to source by using more energy (and more material implements) to carry out the recycling. Energy can only be recycled by expending more energy to carry out the recycling than the amount recycled, so it is never economic to recycle energy—regardless of price. Recycling also requires material implements for collection, concentration, and transportation. The machines used to collect, concentrate, and transport will themselves wear out through a process of **entropic dissipation**—the gradual erosion and dispersion of their material components into the environment in a one-way flow of low entropy usefulness to high entropy waste. Any recycling process must be

¹By matter-energy we mean just matter and energy, but with the recognition that they are convertible according to Einstein's famous formula, $E = mc^2$.

²For a scholarly development of this theme, see N. Georgescu-Roegen, *The Entropy Law and the Economic Process*, Cambridge, MA: Harvard University Press, 1971.

efficient enough to replace the material lost to this process. Nature's biogeochemical cycles powered by the sun can recycle matter to a high degree—some think 100%. But this only underlines our dependence on nature's services, since in the human economy we have no source equivalent to the sun, and our finite sinks fill up because we are incapable of anything near 100% materials recycling.

Information: The Ultimate Resource? There is a strong tendency to deny our dependence on nature to achieve our purposes. Among the more explicit denials is that from George Gilder:³

Gone is the view of a thermodynamic world economy, dominated by "natural resources" being turned to entropy and waste by human extraction and use. . . . The key fact of knowledge is that it is anti-entropic: it accumulates and compounds as it is used. . . . Conquering the microcosm, the mind transcends every entropic trap and overthrows matter itself.

According to *The Economist*, George Gilder is "America's foremost technology prophet" whose recommendation can cause the share price of a company to increase by 50 percent the next day.⁴ If Gilder is really that influential, it simply proves that stock prices are often based on erroneous information and irrational expectations. To cast further doubt on Gilder's Gnostic⁵ prophecy, one need only recall the aphorisms of Nobel chemist Frederick Soddy: "No phosphorus, no thought"⁶ and of Loren Eiseley: "The human mind . . . burns by the power of a leaf." As Kenneth Boulding—one of the pioneers of ecological economics—pointed out, knowledge has to be imprinted on physical structures in the form of improbable arrangements of matter before it is effective in the economy. And low entropy is the quality of matter-energy that increases its capacity to receive and retain the improbable imprint of human knowledge. For example, to receive the imprint, a typical computer microelectronics plant producing 5000 wafers per day generates some 5 million liters of organic and aqueous solvent waste (i.e., high entropy) per year,⁷ in addition to raw materials and

³G. Gilder, *Microcosm: The Quantum Revolution in Economics and Technology*, New York: Simon & Schuster, 1989, p. 378. Similar views are expressed by the late Julian Simon in *The Ultimate Resource*, Princeton, NJ: Princeton University Press, 1981. Recently Peter Huber has continued the tradition in *Hard Green: Saving the Environment from the Environmentalists*, New York: Basic Books, 2000.

⁴March 25, 2000, p. 73.

⁵Gnosticism was an early Christian heresy teaching that salvation was available only to those who through esoteric spiritual knowledge could transcend matter, and who believed that Christ was noncorporeal; hence, any view that denigrates the material world and sees knowledge as an escape therefrom.

⁶Phosphorous is an essential component of chlorophyll, which is required for photosynthesis, which in turn is required for life, which is required for thought.

⁷J.M. Desimone, Practical Approaches to Green Solvents, *Science* 297 (5582) (2002).

energy used. With regard to retaining the imprint, recent estimates suggest that the information technology (IT) economy in the U.S. currently consumes 13% of the electricity we use as a nation, and this level is increasing rapidly.⁸

Furthermore, as important as knowledge is, it is misleading to say it grows by compounding accumulation. New dollars from compound interest paid into a bank account are not offset by any decline in old dollars, that is, the principal. Yet new knowledge often renders old knowledge obsolete, as we saw in our discussion of scientific revolutions and paradigm shifts. Do scientific theories of phlogiston⁹ and the ether¹⁰ still count as knowledge? And when knowledge becomes obsolete, the artifacts that embody that knowledge become obsolete as well. Again the IT economy is the best example. Experts estimate that 300 million personal computers, each containing 7–10 pounds of toxic metal, will end up in U.S. landfills over the next 5 years alone.¹¹ For every three computers that enter the market, two become obsolete. This is hardly antientropic. Physicists will not be surprised, because they have never found anything that is antientropic.

As E. J. Mishan noted, technological knowledge often unrolls the carpet of increased choice before us by the foot, while simultaneously rolling it up behind us by the yard.¹² Yes, knowledge develops and improves, but it does not grow exponentially like money compounding in the bank. Furthermore, new knowledge need not always reveal new possibilities for growth; it can also bring serious harm and reveal new limitations. The new knowledge of the fire-resisting properties of asbestos increased its usefulness; subsequent new knowledge of its carcinogenic properties reduced its usefulness. New knowledge can cut both ways. Finally, and most obviously, knowledge has to be actively learned and taught every generation—it cannot be passively bequeathed like an accumulating stock portfolio. When society invests little in the transfer of knowledge to the next generation, some of it is lost and its distribution often becomes more

⁸M.P. Mills, "Kyoto and the Internet: The Energy Implications of the Digital Economy," Testimony of Mark P. Mills, Science Advisor, The Greening Earth Society, Senior Fellow, The Competitive Enterprise Institute, President, Mills-McCarthy & Associates, before the Subcommittee on National Economic Growth, Natural Resources, and Regulatory Affairs, U.S. House of Representatives, 2000.

⁹Phlogiston is a hypothetical substance formerly thought to be a volatile constituent of all combustible material, released as flame in combustion.

¹⁰Ether is an all-pervading, infinitely elastic, massless medium once thought to fill the upper regions of space, as air fills the lower regions. Its hypothetical existence avoided confronting the mystery of "action at a distance," later recognized in the concept of gravity.

¹¹"Electronic Product Recovery and Recycling Baseline Report: Recycling of Selected Electronic Products in the United States," National Safety Council, May 1999.

¹²E. J. Mishan, *The Costs of Economic Growth*, New York: Praeger, 1967.

concentrated, contributing to the growing inequality in the distribution of income, as well as to a general dumbing-down of the future.

It is a gross prejudice to think that the future will always know more than the past. Every new generation is born totally ignorant, and just as we are always only one failed harvest away from starvation, we are also always only one failed generational transfer of knowledge away from darkest ignorance. Although it is true that today many people know many things that no one knew in the past, it is also true that large segments of the present generation are more ignorant than were large segments of past generations. The level of policy in a democracy cannot rise above the average level of understanding of the population. In a democracy, the distribution of knowledge is as important as the distribution of wealth.

Waste as a Resource? The common view among economists and many others is that waste is just a resource we have not yet learned to use, that nature supplies only the indestructible building blocks of elemental atoms, and that all the rest either is or can be done by humans. What counts to economists is value added by human labor and capital—that to which value is added is thought to be totally passive stuff, not even worthy of the name *natural resources*, as evidenced by Gilder's putting the term in quotation marks. Natural processes, in this view, do not add value to the elemental building blocks—and even if they did, manmade capital is thought to substitute for such natural services.

The brute facts remain, however, that we can only get so much energy from a lump of coal, we cannot burn the same lump twice, and the resulting ashes and heat scattered into nature's sinks really are polluting wastes and not just matter-energy of equally useful potential, if only we knew how to use it. Eroded topsoil washed to the sea and chlorofluorocarbons in the ozone layer are also polluting wastes on a human time scale, not just "resources out of place." No one denies the enormous importance of knowledge.¹³ But this denigration of the importance of the physical world, and exclusive emphasis on knowledge as our ultimate resource, seems to be a modern version of Gnosticism. It appears to be religiously motivated by a denial of our creaturehood as part of the material world, by the belief that we have, or soon will have, transcended the world of material creation and entered an unlimited realm of esoteric knowledge, albeit technical now rather than spiritual. Thus, even in the discussion of means we are pushed out of the purely biophysical realm to

¹³For interesting discussions of the limitations of knowledge, see P. R. Ehrlich et al., Knowledge and the Environment, *Ecological Economics* 30:267–284 (1999) and M. H. Huesemann, Can Pollution Problems Be Effectively Solved by Environmental Science and Technology? An Analysis of Critical Limitations, *Ecological Economics* 37:271–287 (2001).

consider alternative religious philosophies, including most prominently the revival of the ancient Christian heresy of Gnosticism.

Ends

We argued earlier that there is such a thing as ultimate means, and that it is low-entropy matter-energy. Is there such a thing as an **ultimate end**, and if so, what is it? Following Aristotle, we think there are good reasons to believe that there must be an ultimate end, but it is far more difficult to say just what it is. In fact we will argue that, while we must be dogmatic about the existence of the ultimate end, we must be very humble and tolerant about our hazy and differing perceptions of what it looks like.

In an age of “pluralism,” the first objection to the idea of ultimate end is that it is singular. Do we not have many “ultimate ends”? Clearly we have many ends, but just as clearly they conflict and we must choose among them. We rank ends. We prioritize. In setting priorities, in ranking things, something—only one thing—has to go in first place. That is our practical approximation to the ultimate end. What goes in second place is determined by how close it came to first place, and so on. Ethics is the problem of ranking plural ends or values. The ranking criterion, the holder of first place, is the ultimate end (or its operational approximation), which grounds our understanding of objective value—better and worse as real states of the world, not just subjective opinions.

We do not claim that the ethical ranking of plural ends is necessarily done abstractly, *a priori*. Often the struggle with concrete problems and policy dilemmas forces decisions, and the discipline of the concrete decision helps us implicitly rank ends whose ordering would have been too obscure in the abstract. Sometimes we have regrets and discover that our ranking really was not in accordance with a subsequently improved understanding of the ultimate end.

Neoclassical economists reduce value to the level of individual tastes or preferences, about which it is senseless to argue. But this apparent tolerance has some nasty consequences. Our point is that we must have a dogmatic belief in objective value, an objective hierarchy of ends ordered with reference to some concept of the ultimate end, however dimly we may perceive the latter. This sounds rather absolutist and intolerant to modern devotees of pluralism, but a little reflection will show that it is the very basis for tolerance. If A and B disagree regarding the hierarchy of values, and they believe that objective value does not exist, then there is nothing for either of them to appeal to in an effort to persuade the other. It is simply A's subjective values versus B's. B can vigorously assert her preferences and try to intimidate A into going along, but A will soon get wise to that. They are left to resort to physical combat or deception or manipulation, with no possibility of truly reasoning together in search of a clearer shared

vision of objective value, because, by assumption, the latter does not exist. Each knows his own subjective preferences better than the other, so no “values clarification” is needed. If the source of value is in one’s own subjective preferences, then one does not really care about the other’s preferences, except as they may serve as means to satisfying one’s own. Any talk of tolerance becomes a sham, a mere strategy of manipulation, with no real openness to persuasion.¹⁴

Of course, we must also be wary of dogmatic belief in a too explicitly defined ultimate end, such as those offered by many fundamentalist religions.¹⁵ In this case, again, there is no possibility of truly reasoning together to clarify a shared perception, because any questioning of revealed truth is heresy.

■ THE PRESUPPOSITIONS OF POLICY

Ecological economics is committed to policy relevance. It is not just a logical game for autistic academicians. Because of our commitment to policy, we must ask: What are the necessary presuppositions for policy to make sense, to be worth discussing? We see two.

First, we must believe that there are real alternatives among which to choose. If there are no alternatives, if everything is determined, then it hardly makes sense to discuss policy—what will be, will be. If there are no options, then there is no responsibility, no need to think.

Second, even if there are real alternatives, policy dialogue would still make no sense unless there were a real criterion of value to use for choosing among the alternatives. Unless we can distinguish better from worse states of the world, it makes no sense to try to achieve one state of the world rather than another. If there is no value criterion, then there is no responsibility, no need to think.

In sum, serious policy must presuppose: (1) nondeterminism—that the world is not totally determined, that there is an element of freedom that offers us real alternatives; and (2) nonnihilism—that there is a real criterion of value to guide our choices, however vaguely we may perceive it.

The fact that many people engaged in discussing and making policy reject one or both of these presuppositions is, in A. N. Whitehead’s term, “the lurking inconsistency,” a contradiction at the basis of the modern worldview that enfeebles thought and renders action halfhearted. If we even halfway believe that purpose is an illusion foisted on us by our genes

¹⁴For a fuller exposition of this argument, see C. S. Lewis, *The Abolition of Man*, New York: Macmillan, 1947.

¹⁵Many economists seem to view “efficiency” and/or growth as the ultimate end and border on religiosity in their convictions. See Robert Nelson, *Economics as Religion*, University Park: Pennsylvania State University Press, 2002.

to somehow make us more efficient at procreation,¹⁶ or that one state of the world is as good as another, then it is hard to get serious about real issues. And ecological economics must be serious about real issues. As Whitehead noted, “Scientists animated by the purpose of proving that they are purposeless constitute an interesting subject for study.”¹⁷

■ DETERMINISM AND RELATIVISM

The preceding section may seem pretty obvious and consistent with common sense. What is the point of stating the obvious? The point is that many members of the intelligentsia deny either nondeterminism, nonnihilism, or both, yet they want to engage in a policy dialogue. It is not just that we disagree on exactly what our alternatives are in a particular instance, or about what our value criterion implies for a concrete case—that’s part of the reasonable policy dialogue. The point is that determinists who deny the effective existence of alternatives, and nihilists or relativists who deny the existence of a value criterion beyond the level of subjective personal tastes, have no logical basis for engaging in policy dialogue—and yet they do! We cordially and respectfully invite them to remember and reflect deeply upon their option of remaining silent—at least about policy.¹⁸

One may well agree with the logic of our position—that policy rules out determinism and nihilism—but argue that there are so few real determinists and nihilists around that in effect we are kicking at an open door, or attacking a straw man. We hope this is true. However, one leading biologist, Paul R. Ehrlich, who has contributed much to ecological economics, recently wrote a book with this stated purpose:¹⁹ “to give an evolutionist’s antidote to the extreme hereditary determinism that infests much of the current discussion of human behavior—the idea that we are somehow simply captives of tiny, self-copying entities called genes” (p. x). In other words, Ehrlich felt that the influence of the hard-line determinists is sufficiently toxic to require a 500-page antidote, even if a rather mild and general one.

A stronger and more specific antidote was thought necessary by Wen-

¹⁶As asserted in E. O. Wilson’s *Consilience* (New York: Knopf, 1998) and R. Dawkins’ *The Blind Watchmaker* (New York: Norton, 1996).

¹⁷A. Whitehead, *The Function of Reason*, Princeton, NJ: Princeton University Press, 1929, p. 12.

¹⁸In the sciences, these are hard-line neo-Darwinists and sociobiologists; in the humanities, postmodern deconstructionists; in theology, hard-line Calvinist believers in predestination; and in the social sciences, some evolutionary psychologists, and so-called “value-free” economists who reduce value to subjective individual tastes, any one of which is as good as another.

¹⁹P. R. Ehrlich, *Human Natures: Genes, Cultures, and the Human Prospect*, Washington, DC: Island Press, 2000.

dell Berry, who took particular aim at the influential writings of Edward O. Wilson, especially his recent book *Consilience*. Berry deserves to be quoted at length:²⁰

A theoretical materialism as strictly principled as Mr. Wilson's is inescapably deterministic. We and our works and acts, he holds, are determined by our genes, which are determined by the laws of biology, which are determined ultimately by the laws of physics. He sees that this directly contradicts the idea of free will, which even as a scientist he seems unwilling to give up, and which as a conservationist he cannot afford to give up. He deals with this dilemma oddly and inconsistently.

First, he says that we have, and need, "the illusion of free will," which, he says further, is "biologically adaptive." I have read his sentences several times, hoping to find that I have misunderstood them, but I am afraid that I understand them. He is saying that there is an evolutionary advantage in illusion. The proposition that our ancestors survived because they were foolish enough to believe an illusion is certainly optimistic, but it does not seem very probable. And what are we to think of a materialism that can be used to validate an illusion? Mr. Wilson nevertheless insists upon his point; in another place he speaks of "self-deception" as granting to our species the "adaptive edge." Later, in discussing the need for conservation, Mr. Wilson affirms the Enlightenment belief that we can "choose wisely." How a wise choice can be made on the basis of an illusory freedom of the will is impossible to conceive, and Mr. Wilson wisely chooses not to try to conceive it. (p. 26)

We have learned from personal conversation with Wilson that he considers the question of how one squares scientific determinism with purposeful policy to be "the mother of all questions." Mutual humility in the face of mystery and paradox is more easily expressed, and understood, in friendly conversation over wine and dinner than in dry academic print. No one can, in practice, live by the creeds of determinism or nihilism. In this sense, no one takes these creeds seriously, not even the advocates themselves. So we tend to discount any effect on policy of these doctrines. However, many open-minded citizens halfway suspect that the learned scholars who publicly proclaim these views might know something that they do not. Maybe I really am just a robot controlled by my selfish genes; maybe purpose really is just an epiphenomenal illusion; maybe better and worse really are just meaningless terms for lending undue authority to subjective personal preferences, or to class-based, gender-based, or race-based interests. The fact that determinist or nihilist views cannot consistently be lived out in practice by individuals does not mean that their

²⁰W. Berry, *Life Is a Miracle: An Essay Against Modern Superstition*, Washington, DC: Counterpoint Press, 2000.

existence, lurking in the back of the collective mind, is not capable of disabling policy.

In the Introduction, we referred briefly to the difficulty some ecologists have in dealing with policy, the messy world of human affairs. To the extent that the ecologist, like some biologists, is a determinist, then policy of any kind would be silly. Such an ecologist would necessarily be more laissez-faire than the most extreme free market economist. Hence our view that ecological economics is not simply a matter of bringing the light of ecology to dispel the darkness of economics. There is that to be sure, but there is also some darkness within ecology that economists do not need to import.

Perhaps we should take some cues from modern physics, just as traditional economics takes cues from nineteenth-century mechanical physics. Quantum indeterminacy and chaos theory have upset the “scientific” foundations of determinacy. And many of our greatest modern physicists, those who have best come to understand the physical matter underlying the scientific materialism paradigm, increasingly question its ability to provide any ultimate truths. For example, Einstein points out that scientific knowledge “of what is does not open the door directly to what should be.” He goes on to ask “what should be the goal of our human aspirations? The ultimate goal itself and the longing to reach it must come from another source.”²¹ In Schrodinger’s words, “the scientific picture of the real world around me is very deficient. It gives a lot of factual information, puts all our experience in a magnificently consistent order, but it is ghastly silent about all and sundry that is really near to our heart, that really matters to us—we do not belong to the material world the science constructs for us.”²²

Policy students, including economists, implicitly assume that the world offers more than one possibility to choose from, and that some choices really are better than others. This is also true, of course, for ecological economists who, while continuing to take biology and ecology very seriously, must not fall into the metaphysical traps of determinism or nihilism that seem to have ensnared some in those disciplines.

To be sure, not every conceivable alternative is a real alternative. Many things really are impossible. But the number of viable possibilities permitted by physical law and past history is seldom reduced to only one. Through our choices, value and purpose lure the physical world in one direction rather than the other. Purpose is independently causative in the world.

²¹A. Einstein, *Ideas and Opinions*. New York: Crown, 1954. Quoted in T. Maxwell “Integral Spirituality, Deep Science, and Ecological Awareness.” *Zygon: Journal of Religion and Science* 38(2): 257–276 (2003).

²²E. Schrodinger, *My View of the World*, Cambridge: Cambridge University Press, 1964. Quoted in Maxwell, op. cit.

Box 3-1**DETERMINISM IN THE HISTORY OF PHILOSOPHY**

Materialism, determinism, and mechanism are closely related metaphysical doctrines about the basic nature of reality. If you study the history of philosophy, you will see that they go back to Epicurus, Democritus, and Lucretius, over 2000 years ago, and these doctrines are still very much with us today. It would be arrogant for two economists to think that they can resolve this ancient puzzle, but also naïve to think that we can sidestep it, since economics is unavoidably about choice. If choice is an illusion, what does that say about economics?

Because humans are part of reality, it follows that, if matter in motion is all there is to reality, then that is all there is to humans as well. Since the motions of matter are determined by mechanical law, it follows that the same laws ultimately determine human action. This **determinism** rules out free will—it means that our purposes are not independently causative in the world. Only mechanical motion of matter is causative. Purposes, intentions, values, choices are all more or less dreams or subjective hallucinations. They are effects, not causes.

Nihilism, the rejection of all moral values, is the ethical consequence of the materialist, determinist cosmology. Things are what they are, and you can do nothing about it because your will and purpose have no power to change things. You can have no responsibility for what cannot be otherwise. For Epicurus this was a great relief—much better than worrying about the gods' anger and retribution, about responsibility and guilt and punishment. Relax, don't worry, do your best to enjoy life. Nothing can really hurt you, because when you are dead, that's the end and you can no longer suffer. This view is still very much alive in the modern secular world, although it has a long history of conflict with Christianity, Judaism, and Islam, as well as other philosophies that reject materialism as an adequate view of reality. They insist that good and evil are as real in our experience as matter, and that humans have at least some capacity for choice between them. To ignore our direct experience of good, evil, and freedom is considered antiempirical and against the deeper spirit of science.

It is not our intent to convert you either to or from Epicurianism, Christianity, or any other position. Maybe you do not yet have any position on this question. But logic does have its demands, and no doctrine is exempt from them. Even the early materialists recognized the contradictions involved in a doctrine that ruled out freedom, novelty, and choice. Epicurus tried to restore a modicum of freedom in an *ad hoc* manner by introducing the notion of the “clinamen”—the idea that atoms swerved from their determined motions for unexplained reasons, and that this was the source of novelty, and perhaps some degree of freedom. Our advice is to be skeptical of any easy answer to a problem that has been around for 2500 years, and also to be humble in the face of any logical contradictions that you cannot resolve.

■ THE ENDS-MEANS SPECTRUM

Ultimate means and the ultimate end are two extremes of an **ends-means spectrum** in the middle of which economic value is determined. In everyday life, it is our mid-range ends and means that interact, not their ultimate origins in the realms of the spirit or the electron. We will discuss this intermediate, mid-spectrum interaction in our consideration of the function of markets and relative prices (see Chapter 8). But for now it is useful to think of the entire ends-means spectrum depicted in Figure 3.1. The economic choices that exist in the mid-range of the spectrum are not illusory. They are not totally determined by material causes from below, nor are they rendered meaningless by an absence of final cause from above, or the presence of a predestining final cause. As we will discuss later, prices, relative values, are determined by supply and demand. But supply reflects alternative conditions of relative possibility, of the reality of ultimate means, while demand reflects independent conditions of relative desirability, rooted in notions of better and worse, of ethical choices based on some perception of the ultimate end.

In its largest sense, humanity's ultimate economic problem is to use ultimate means efficiently and wisely in the service of the ultimate end.

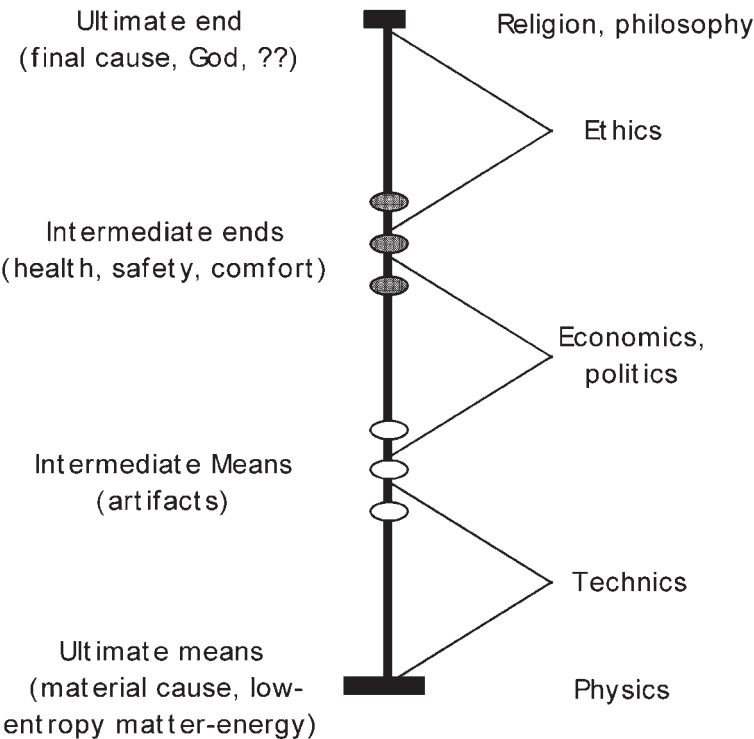


Figure 3.1 • The ends-means spectrum.

Stated in this way, the problem is overwhelming in its inclusiveness. Therefore, it's not hard to understand why in practice it has been broken up into a series of sub-problems, each dealt with by a different discipline, as indicated on the right side of the ends-means spectrum.

At the top of the spectrum, we have the ultimate end, studied by religion and philosophy. It is that which is intrinsically good and does not derive its goodness from any instrumental relation to some other or higher good. It is the highest good, to which all other good is instrumental and derivative. Needless to say, it is not well defined. As noted earlier, there are unacceptable consequences from denying its existence, but the dimness of our vision of the ultimate end is part of the human condition and requires a great deal of mutual tolerance. The error of treating as ultimate that which is not is, in theological terms, idolatry.

At the bottom of the spectrum is ultimate means, the useful stuff of the world—low-entropy matter-energy, which we can only use up and cannot create or replenish, and whose net production cannot possibly be the end result of any human activity. The ultimate end is much harder to define than ultimate means. Our current approximation to the ultimate end, unfortunately, seems to be economic growth, and part of the critique of economic growth is that our devotion to it has become idolatrous, worshipping a false god, so to speak, because it is not really ultimate. But it is not easy to formulate a central organizing principle of society that does not border on idolatry.

To reiterate, since we are forced by scarcity to choose which of our many intermediate ends will be satisfied and which will be sacrificed, we must rank our intermediate ends. Ranking means establishing priority. Priority means that something goes in first place. That holder of first place is our operational estimate of the ultimate end. It provides the ordering criterion for ranking other intermediate ends. Second place goes to whatever is nearest to or best serves first place, and so on. This ranking of intermediate ends relative to our vision of the ultimate end is the problem of ethics. Economists traditionally take the solution to the ethical problem as given and start their analysis with a given ranking of intermediate ends, or with the assumption that one person's ranking is as good as another's, so that ethics is indistinguishable from personal subjective tastes.

At the bottom of the spectrum, physics studies ultimate means, and technics studies the problem of turning ultimate means into artifacts specifically designed to satisfy each of our intermediate ends. Economists also habitually assume the technical problem to have been solved; that is, technology is taken as given. Thus, the remaining segment of the spectrum is the middle one of allocating *given* intermediate means to the service of a *given* hierarchy of intermediate ends. This is the significant and

important economic problem, or rather political economic problem, quite distinct from the ethical and technical problems.

The middle-range nature of the problem of political economy is significant. It means that, from the perspective of the entire spectrum, economics is, in a sense, both too materialistic and not materialistic enough. In abstracting from the ethical and religious problem it is “too materialistic,” and in abstracting from the technical and biophysical problem it is not materialistic enough. Economic value has both physical and moral roots. Neither can be ignored. Yet many thinkers are attracted to a monistic philosophy that focuses only on the biophysical or only on the psychic root of value. Ecological economics adopts a kind of practical dualism. Dualism is not as simple as monism, and it entails the mysterious problem of how the material and the spiritual interact. That is indeed a large and enduring mystery. But on the positive side, dualism is more radically empirical than either monism, refusing to deny or ride roughshod over inconvenient facts, just to avoid confronting a mystery.²³

■ THREE STRATEGIES FOR INTEGRATING ECOLOGY AND ECONOMICS

Previous attempts to integrate economics and ecology have been based on one of three strategies: (1) economic imperialism; (2) ecological reductionism; (3) steady-state subsystem. Each strategy may be thought of as beginning with the picture of the economy as a subsystem of the ecosystem. The differences concern the way they each treat the boundary between the economy and the rest of the ecosystem (Figure 3.2).

Economic Imperialism

Economic imperialism seeks to expand the boundary of the economic subsystem until it encompasses the entire ecosystem. The goal is one system, the macroeconomy as the whole. This is to be accomplished by complete internalization of all external costs and benefits into prices. Price, of course, is the ratio (e.g., dollars per gallon) at which something is exchanged for money (or for some other commodity) by individuals in the market. Those aspects of the environment not customarily traded in markets can be treated as if they were by imputation of “shadow prices”—the economist’s best estimate of what the price of the function or thing would be if it were traded in a competitive market. Everything in the ecosystem is theoretically rendered comparable in terms of its ability to help or hin-

²³Honesty requires facing up to mystery. Although we suspect that mystery is an enduring part of the human condition and not just another word for “future knowledge not yet discovered,” we nevertheless respect the scientific and philosophical quest to solve mysteries, including the mystery inherent in the dualism we advocate as a practical working philosophy.

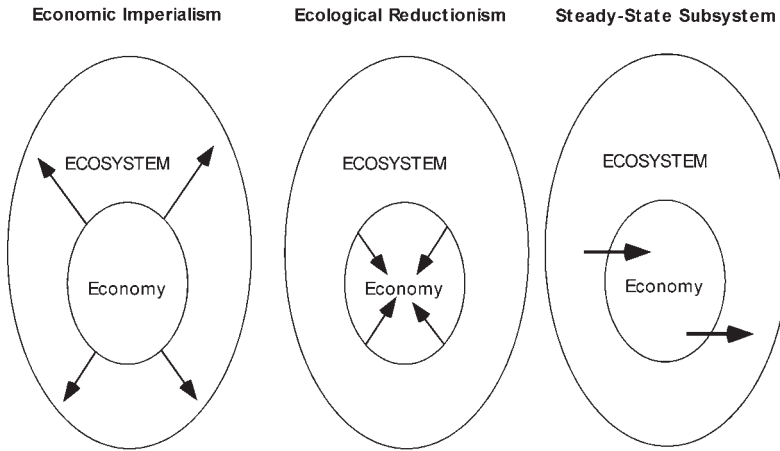


Figure 3.2 • Three strategies for integrating ecology and economics.

der individuals in satisfying their wants. Implicitly, the single end pursued is ever-greater levels of consumption, and the only intermediate means to effectively achieve this end is growth in market goods. Economic imperialism is basically the neoclassical approach.

Subjective individual preferences, however whimsical, uninstructed, or ill-considered, are taken as the source of all value. Since subjective wants are thought to be infinite in the aggregate, as well as sovereign, there is a tendency for the scale of activities devoted to satisfying them to expand. The expansion is considered legitimate as long as “all costs are internalized.” But most of the costs of growth we have experienced have come as surprises. We could not have internalized them if we could not first imagine and foresee them. Furthermore, even after some external costs have become visible to all (e.g., greenhouse warming), internalization has been very slow and partial. As long as the evolutionary fitness of the environment to support life is unperceived by economists, it is likely to be destroyed in the imperialistic quest to make every molecule in creation, including every strand of DNA, pay its way according to the pecuniary rules of present value maximization.

Furthermore, this imperialism sacrifices the main virtue of free market economists, namely their antipathy to the arrogance of central planners. Putting a price tag on everything in the ecosystem requires information and calculating abilities far beyond anything attempted by Gosplan in the old Soviet Union.²⁴ As an example, let’s take a look at what calculations

²⁴Gosplan is the Russian acronym for the State Planning Committee, which centrally developed 5-year and annual plans for the Soviet economy, at all levels from individual enterprises to the national level.

would be required to accurately quantify and internalize the costs associated with global warming. Currently we are incapable of accounting for even carbon dioxide flows, the most basic piece of the puzzle. How much carbon is being absorbed by oceans or terrestrial ecosystems? How will it affect these ecosystems? Will global warming lead to positive-feedback loops, such as a release of methane from a thawing arctic tundra and increased atmospheric water vapor from more rapid evaporation from the oceans (both potent greenhouse gases), or negative-feedback loops via increased sequestration of carbon by forests? How will temperature changes affect global weather patterns over the next century? (And how certain can we be about such estimates, when we cannot even accurately predict the weather next week?) What changes would have occurred even in the absence of global warming? What technologies will evolve to cope with these problems, and how much will changing our rate of greenhouse gas emissions affect the rate of technological advance? Finally, how will these factors affect the economy? Bear in mind that while a meteorologist cannot accurately predict the weather in a week, she can at least stick her head out the window and say “it’s raining.” Economists, on the other hand, at the time of this writing are in the midst of a heated debate over whether or not the economy is in a recession *right now*.

These calculations, a small fraction of those that would be required to estimate the costs of global warming, are clearly beyond the capabilities of modern science, and quite probably beyond the capacity of the human mind. And calculating all the costs at the time they occur is the straightforward part. How do we determine the present value of costs to future generations? The currently favored approach, intertemporal discounting (to which we return in Chapter 10) gives less value to future costs and benefits than those that occur today, and the discount rate we choose in this calculation is likely to be as important as any other of the variables mentioned above. But discounting in this case implies that future generations have no inalienable right to a stable climate, economic growth will continue throughout the discount period, and economic growth is a satisfactory substitute. Yet the discount rate we choose for internalizing costs will itself affect the rate of growth.

The global warming example brings up another serious problem with economic imperialism—the assumption that the most efficient mechanism for allocating virtually any means among any ends is the market. In fact, markets are incapable of allocating goods that cannot be owned, and inefficient at allocating goods for which use does not lead to depletion (either or both of which are properties of the bulk of ecosystem services). Even if we could put an appropriate charge on greenhouse gas emissions to internalize their costs, who would receive the charge? It would seem only fair that it would go to those who bear the costs. Would it even be a

market transaction if when we purchased something, we did not pay the person who bore the costs of production? However, global warming is likely to affect virtually the entire population of the planet for countless generations into the future. This would imply that not only would we need to calculate all the costs, we would need to do so for *each individual*. Strangely enough, as we will discuss later, some neoclassical economists argue that those who bear the costs of externalities should not receive the payments,²⁵ but in this case, how could we say that the result resembles a market solution? A major goal of this text will be to explain exactly why many goods and services are not amenable to market solutions, independently of whether or not we are able to internalize all costs.

Let's play the role of the stereotypical economist and assume away all these problems. There is, then, no doubt that once the scale of the economy has grown to the point that formerly free goods become scarce, it is better that these goods should have a positive price reflecting their scarcity than to continue to be priced at zero. But there remains the prior question: Are we better off at the new scale with formerly free goods correctly priced, or at the old scale with free goods also correctly priced at zero? In both cases, the prices are right. This is the suppressed question of optimal scale, and it is not answered by market prices.

Ecological Reductionism

Ecological reductionism begins with the true insight that humans are not exempt from the laws of nature. It then proceeds to the false inference that human action is totally explainable by, reducible to, the laws of nature. It seeks to explain whatever happens within the economic subsystem by exactly the same naturalistic principles that it applies to the rest of the ecosystem. It shrinks the economic subsystem to nothing, erasing its boundary. Taken to the extreme, in this view energy flows, embodied energy costs, and relative prices in markets are all explained by a mechanistic system that has no room for purpose or will. This may be a sensible vision from which to study some natural systems. But if one adopts it for studying the human economy, one is stuck from the beginning with the important policy implication that policy makes no difference. We encounter again all the problems of determinism and nihilism already discussed.

Economic imperialism and ecological reductionism have in common that they are monistic visions, albeit rather opposite monisms. It is the monistic quest for a single substance or principle by which to explain all value that leads to excessive reductionism on both sides. Certainly one

²⁵E. T. Verhoef, "Externalities." In J. C. J. M. van den Bergh, ed. *Handbook of Environmental and Resource Economics*, Northampton, MA: Edward Elgar, 1999.

should strive for the most reduced or parsimonious explanation possible without ignoring the facts. But respect for the basic empirical facts of chance and necessity on the one hand, and self-conscious purpose and will on the other hand, should lead us to a kind of practical dualism or polarity reflected in the ends-means spectrum. After all, the fact that our being should consist of two fundamental elements offers no greater inherent improbability than that it should rest on one only. How these two fundamental elements of our being interact is a mystery—precisely the mystery that the monists of both kinds are seeking to avoid. But economists are too much in the middle of the spectrum to adopt either monistic “solution.” Economists are better off denying the tidy-mindedness of either monism than denying the untidy and mysterious facts.

The Steady-State Subsystem

The remaining strategy, is the **steady-state subsystem**, the one adopted here. It does not attempt to eliminate the subsystem boundary, either by expanding it to coincide with the whole system or by reducing it to nothing. Rather, it affirms the fundamental necessity of the boundary and the importance of drawing it in the right place. It says that the scale of the human subsystem defined by the boundary has an optimum, and that the throughput by which the ecosystem maintains and replenishes the economic subsystem must be ecologically sustainable. Once we have drawn this boundary in the appropriate place, we must further subdivide the economic subsystem into regions where the market is the most effective means of allocating resources, and regions where it is inappropriate. These regions are determined by inherent characteristics of different goods and services, to be discussed at length in this text.

Box 3-2

THE STEADY-STATE ECONOMY

The idea of a steady-state economy comes from classical economics, and was most developed by John Stuart Mill (1857), who referred to it as the “stationary state.” The main idea was that population and the capital stock were not growing. The constancy of these two physical stocks defined the scale of the economic subsystem. Birth rates would be equal to death rates and production rates equal to depreciation rates, so that both the stock of people (population) and the stock of artifacts (physical capital) would be constant—not static, but in a state of dynamic equilibrium. Most classical economists dreaded the stationary state as the end of progress, but not Mill:^a

It is scarcely necessary to remark that a stationary condition of capital and population implies no stationary state of human improvement. There would be as much scope as ever for all kinds of mental culture, and moral and social progress; as much room for improving the Art of Living and much more likelihood of its being improved, when minds cease to be engrossed by the art of getting on.

Mill thought we would pay more attention to getting better, once we ceased to be so preoccupied with getting bigger. He also recognized that growth could become uneconomic:

If the earth must lose that great portion of its pleasantness which it owes to things that the unlimited increase of wealth and population would extirpate from it, for the mere purpose of enabling it to support a larger, but not a happier or better population, I sincerely hope, for the sake of posterity, that they will be content to be stationary, long before necessity compels them to it.

In physical terms, populations of both human bodies and things are what physicists call “dissipative structures”—things that fall apart, die, and decay if left to themselves. People die, goods wear out. To keep a population of dissipative structures constant requires births equal to deaths, and production equal to depreciation—in other words, input equal to output equal to throughput, a concept with which you are now familiar. But births can equal deaths at low rates or at high rates. Either one will keep the population constant. Which do we want? If we want a long life expectancy for individuals, we must choose low birth rate equal to low death rate. For an equilibrium population with birth equal to death rates at 40 per thousand per year, the average age at death must be 25 years. If we want people to live to be 67 rather than 25, we will have to lower birth and death rates to 15 per thousand per year. Can you explain why? Can you apply the same logic to lifetime or durability of the stock of goods?

To summarize: The main idea of a steady-state economy is to maintain constant stocks of wealth and people at levels that are sufficient for a long and good life. The throughput by which these stocks are maintained should be low rather than high, and always within the regenerative and absorptive capacities of the ecosystem. The system is therefore sustainable—it can continue for a long time. The path of progress in the steady state is no longer to get bigger, but to get better. This concept was a part of classical economics, but unfortunately was more or less abandoned by NCE. More precisely, the terms *stationary* and *steady state* were redefined to refer not to constant population and capital stock, but to their proportional growth—a constant ratio between ever-growing stocks of people and things!

⁹J. S. Mill, *Principles of Political Economy*, Book IV, Chapter VI (1848). Online: <http://www.econlib.org/library/Mill/mIPbl.html>.

BIG IDEAS to remember

- Practical dualism versus monisms
 - Presuppositions of policy
 - Ultimate means
 - Ultimate end
 - Information and knowledge
 - Determinism and materialism
 - Ends-means spectrum
 - Economic imperialism
 - Ecological reductionism
 - Steady-state subsystem
-



Conclusions to Part I

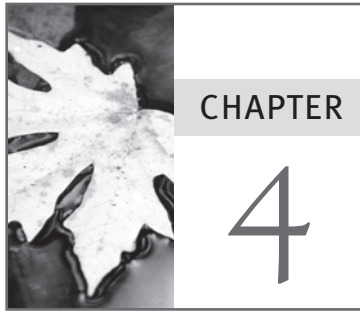
In Part I, we have defined economics as the science of the allocation of scarce resources among alternative desirable ends. Ecological economics distinguishes itself from mainstream economics in its preanalytic vision of the economic system as a subsystem of the sustaining and containing global ecosystem. Economic growth is not an end in itself, continuous physical economic growth is not possible, and eventually the costs imposed by growth on the sustaining system become greater than the benefits of that growth. Economic systems change the environment, then adapt in responses to those changes in a coevolutionary process. Economic growth has increased the scarcity of ecological goods and services relative to manmade goods and services, and the economic system must adapt to that fact. The way in which the economy evolves is not predetermined; it can be influenced by policy in ways that are better or worse.

In the ecological economy, there is an ultimate end, although hard to perceive and not universally agreed upon; the ultimate means is low-entropy matter-energy, and the market is a very useful but by no means sufficient institution for allocating means to the service of ends. Although we shrink from trying to define the ultimate end, as a basis for future discussion in this text, we suggest a working definition of the penultimate end for the ecological economy: the maintenance of ecological life-support systems far from the edge of collapse (which requires an end to material growth of the economy) and healthy, satisfied human populations free to work together in the pursuit and clarification of a still vague ultimate end—for a long, long time.

PART II

The Containing and Sustaining Ecosystem: The Whole





The Nature of Resources and the Resources of Nature

The economic system is a subsystem of the global ecosystem, and one of the major goals of ecological economics is to determine when the benefits of continued growth in the economic subsystem are outweighed by the increasing opportunity costs of encroaching on the sustaining ecosystem. Achieving this goal demands a clear understanding of how the global ecosystem sustains the economy and how economic growth affects the sustaining ecosystem. In addition to determining when economic growth becomes uneconomic, ecological economists must provide the policies necessary to keep the economy within its “optimal” size range. Currently, the dominant tool for determining economic optimality is the market. However, markets only function effectively with goods and services that have certain specific attributes, and they really do not function at all with goods that cannot be exclusively owned. Developing effective policies requires a clear understanding of the specific attributes of goods and services that the economic system must allocate among alternative ends. In this chapter, we introduce you to several concepts that will be useful for understanding scarce resources. These include the difference between stock-flow and fund-service resources, and the concepts of rivalness and excludability. We will also consider further the laws of the thermodynamics.

Chapters 5 and 6 will apply these concepts to the abiotic and biotic scarce resources upon which our economy depends. Chapters 9 and 10 will explain why these concepts are so important to policy analysis.

■ A FINITE PLANET

With the exception of inconsequential bits of material arriving from space, there is only so much water, so much land, and so much atmosphere to our planet. We have finite supplies of soils, minerals, and fossil fuels. Even if we argue that natural processes make more soil and fossil fuels, the rate at which they do so is not only finite; it is exceedingly slow from a human perspective. Fortunately, we are blessed with a steady influx of solar energy that will undoubtedly continue long past the extinction of the human race,¹ but the rate at which this energy arrives is also fixed and finite. Of course, for this energy to be useful, it must be captured, and at present virtually all of that capture is performed by a finite stock of photosynthesizing organisms. In other words, it appears that we live on a finite planet. Why waste words on such an obvious fact?

Continued economic growth is the explicit goal of most economists and policy makers. Many economists even argue that economic growth is not only compatible with a clean environment, it is a prerequisite for achieving one. A clean environment is a luxury good, the story goes. People who are struggling simply to feed themselves cannot be concerned with pollution. The fact that throughout the Third World, poverty forces people to actually live and work in garbage dumps, finding food to eat, clothes to wear, and goods and materials to recycle speaks for itself—survival takes precedence over environment. And work in a factory, no matter how much it pollutes, must be better than life in a dump. Only in rich nations can we afford the luxury of clean water and clean air. This would explain the fact that water quality and air quality in the United States has improved since the 1970s (we will return to this apparent paradox later), and even forest cover is expanding in many areas. The best way to clean up the planet and preserve its remaining ecosystems, it is often argued, is through economic growth.

THINK ABOUT IT!

Do you think the environment in wealthy countries has improved over the past 20 years? Have the global environmental impacts of wealthy countries diminished over the past 20 years? Where do most of the things you buy come from? Do you think their production has negative impacts on the environment?

In contrast to this scenario, the laws of physics tell us that we cannot create something from nothing. Economic production therefore requires

¹The average life span of a mammalian species is only one million years, while the sun is expected to last for several billion years. See R. Foley, "Pattern and Process in Hominid Evolution." In J. Bintliff, ed. *Structure and Contingency*. Leicester, England: Leicester University Press, 1999.

raw material inputs, and the finite supply of those inputs limits the size of the economy. The economic system cannot grow indefinitely, no matter how much we can substitute a new resource for an exhausted one. For example, human populations cannot continue growing forever. A simple calculation shows that even at a continuous 1% rate of growth, the human population would have a mass greater than the entire planet in just over 3000 years.² Similarly, we could not conceivably continue to increase the physical mass of artifacts we own and consume over the next 1000 years at the same rate as we have during the past 50. But population growth rates are already slowing. The most recent U.N. report estimates that the worldwide population will stabilize at about 11 billion people by the year 2200,³ though many ecologists believe planetary ecosystems could not sustain even half that number.⁴ Some also argue that we can produce more using less, so the physical mass of artifacts need not increase. It is true that we can now produce 12 aluminum cans from the same material it once took to produce one, but we still use more aluminum than ever before, and aluminum can only be rolled so thin. Still others assert that economic value is not a measure of a physical quantity, and therefore it is not at all obvious that the production of economic value has physical limits.

It is true that economic value is not a physical quantity. Economic production is really all about creating welfare, quality of life, utility, or whatever else we choose to call this psychic flux of satisfaction. Does it really matter, then, that we live on a finite planet? Certainly it matters in terms of economic production. Economic production, as it is typically understood, is the transformation of raw materials supplied by the ecosystem into something of value to humans. Transformation requires energy, and it inevitably generates waste. Even the service sector requires physical inputs to sustain those who provide the service. We have finite supplies of energy, finite supplies of raw materials, finite absorption capacities for our

²Some would say that this type of calculation is really just a straw man argument, and that no one argues that human population growth will continue indefinitely. However, University of Maryland Professor Julian Simon once claimed that human populations could continue growing at the same rate for the next 7 million years with existing technologies (J. Simon, *The Ultimate Resource*, 2nd ed. Princeton, NJ: Princeton University Press, 1996). A steady 1% growth rate of the current population for that long would leave us with more people than the estimated number of atoms in the universe. Simon is widely and favorably cited in a recent influential book by Bjorn Lomborg, *The Skeptical Environmentalist*, Cambridge: Cambridge University Press, 2001.

³United Nations Population Division, Department of Economic and Social Affairs 1998. World Population Projections to 2150: Executive Summary. Online: <http://www.undp.org/popin/wdtrends/execsum.htm>.

⁴For example, G. Daily, A. Ehrlich, and P. Ehrlich, Optimum Human Population Size. *Population and Environment* 15(6) (1994) argue that 5.5 billion people clearly exceeds the planet's carrying capacity and suggest optimal populations of 1.5 to 2 billion.

wastes, and poorly understood but finite capacities for ecosystems to provide a host of goods and services essential for our survival. And evidence suggests that we are reaching the limits with respect to these resources, as we will describe in greater detail below. With continued growth in production, the economic subsystem must eventually overwhelm the capacity of the global ecosystem to sustain it.

All of this does not mean economic value cannot continue to grow indefinitely. Indeed, we believe that perhaps it can if we define economic value in terms of the psychic flux of human satisfaction, and we learn to attain this satisfaction through nonmaterial means. Ecological economics does not call for an end to economic development, merely to physical growth, while mainstream economists' definitions of economic progress confusingly conflate the two. The problem is that the existing market economy is ill-suited to providing nonmaterial satisfaction. Even if one accepts some variant of the NCEs' assertion that infinite economic "growth" is possible by redefining growth as the ever-greater provision of psychic satisfaction (what we call economic development), the conventional economic paradigm is probably an inadequate guide for achieving this goal—but we'll come back to that later. Our point for now is that constant growth in physical throughput is impossible. Once we understand this, the question becomes how to decide when economic production becomes uneconomic, particularly if this has already happened. Before we address this last question, however, we need to look more closely at the assertion stated above—that infinite growth is impossible in a closed system. The branch of science most relevant to this issue, and indeed most relevant to the economic problem, is thermodynamics.

THINK ABOUT IT!

Of all the activities and objects that give you satisfaction, which ones consume the fewest resources and produce the least waste? Which consume the most resources and produce the most waste? Which of these are produced by the market economy?

■ THE LAWS OF THERMODYNAMICS

A Brief History of Thermodynamics

With the advent of the Industrial Revolution and the machine age at the end of the eighteenth century, scientists became intrigued by the idea of a perpetual motion machine—a machine fueled by the very same heat it generated while it worked. In 1824, the French scientist Sadi Carnot, while trying to calculate the greatest amount of work that could be done by a given amount of heat, realized that a heat engine (e.g., a steam engine) could only perform work by taking heat from one reservoir and

transferring it to another at a lower temperature. In fact, the performance of work in general required a temperature differential between two reservoirs, and all else being equal, the greater the differential, the more work that could be performed. However, even with a temperature differential, it was impossible to convert heat or any kind of energy directly into work with 100% efficiency. It turned out that this was related to the obvious fact that heat would naturally flow from a hotter item to a colder one, and not vice versa. While heat could be made to flow from a colder object to a hotter one, the amount of work required to make this happen was greater than the amount of energy latent in the increased temperature of the hotter object.⁵ To the dismay of industrialists, physical laws did not allow a perpetual motion machine.

Within the course of the next few decades, some other important facts were established. Robert Mayer and Herman Helmholtz showed that energy cannot be created or destroyed, and James Joule performed experiments demonstrating that energy and work are equivalent. Rudolf Clausius recognized that there were two related principles at work here, which came to be called the First and Second Laws of Thermodynamics. The First Law established that energy could not be created or destroyed, and the Second Law established that energy moved inevitably toward greater homogeneity. Because work requires a temperature differential, homogeneity means that energy becomes increasingly unavailable to perform work. In the words of Georgescu-Roegen, “all kinds of energy are gradually transformed into heat, and heat becomes so dissipated in the end that mankind can no longer use it.”⁶ Clausius coined the term *entropy* for the Second Law, derived from the Greek word for transformation, in recognition of the fact that entropy was a one-way street of irreversible change, a continual increase in disorder in the universe. While the First Law of Thermodynamics relates to quantity, the Second Law relates to quality.

A dictionary definition of **entropy** is a measure of the unavailable energy in a thermodynamic system. “Unavailable” means unavailable to do work. Unavailable energy is also known as bound energy, and available energy as free energy. For example, gasoline carries a form of free energy: It can be burned in an internal combustion engine to generate work. Work can be transformed into free energy in a different form (e.g., it can carry a car to the top of a big hill, where it has the potential energy to coast back down) or into heat, which diffuses into the surrounding environment. The energy in the gasoline transformed into heat has not

⁵L. P. Wheeler, *Josiah Willard Gibbs: The History of a Great Mind*, New Haven, CT: Archon Books, 1999.

⁶N. Georgescu-Roegen, *Energy and Economic Myths: Institutional and Analytic Economic Essays*, New York: Pergamon Press, p. 8.

disappeared but has instead become bound energy, unavailable to perform work. In the well-cited example used by Georgescu-Roegen, the ocean contains enormous amounts of energy, but that energy is not available to run a ship.⁷ It is bound energy, because there is no reservoir of a lower temperature to which the energy within the ocean can be transferred, and Carnot showed that such a temperature differential was essential to perform work.

THINK ABOUT IT!

Would you invest in a revolutionary new automobile designed to capture its own exhaust and burn it again?

Does matter, as well as energy, obey the laws of thermodynamics? Einstein's famous $E = mc^2$ established the equivalence between matter and energy, and thus the fact that the First Law applies to matter as well as energy. Georgescu-Roegen argued that the entropy law also applies to matter, and proposed that this be recognized as the fourth law of thermodynamics.⁸ Although physicists dispute the idea of a formal "fourth law," there is no dispute about matter being subject to entropy in the sense of a natural tendency to disorder. When a cube of sugar is dropped into a cup of water it gradually dissolves, losing its order. Nor will that order spontaneously reappear. This is equally obvious for mixing liquids and gases, or more generally for any substance that is soluble in another. It is less obvious for materials in environments in which they are not soluble. However, friction, erosion, and chemical breakdown inexorably lead to the breakdown and diffusion of even the hardest metals over sufficient time, resulting in increased disorder.

It is important to recognize that the laws of thermodynamics were developed more from experimental evidence than from theory, and the mechanism behind entropy is still not completely understood.⁹ When the laws of thermodynamics were first proposed, mechanical physics was the dominant paradigm in science. In a mechanical system, every action has an equal and opposite reaction, and is thus inherently reversible. One theoretical explanation of entropy comes from efforts to harmonize the irreversibility inherent to entropy with the reversibility that characterizes mechanical physics. This has resulted in the field of statistical mechanics, best explained by referring to the example of the sugar cube used above. When in a cube on a shelf, sugar molecules are not free to disperse—there

⁷Ibid., p. 6.

⁸The third law of thermodynamics states that the entropy of any pure, perfect crystalline element or compound at absolute zero (0 K) is equal to zero. This is not particularly relevant to economics.

⁹R. Beard and G. Lozada, *Economics, Entropy and the Environment: The Extraordinary Economics of Nicholas Georgescu-Roegen*, Cheltenham, England: Edward Elgar, 2000.

is only one state space available to them. When placed into a container of water, in which sugar is soluble, sugar molecules are free to move. Suddenly, there are almost countless possible arrangements the sugar molecules can take within that container. Each arrangement may have an equal probability, but only one of those arrangements is that of the cube. Thus, the probability of the cube remaining intact is almost immeasurably small. According to this statistical version of thermodynamics, or statistical mechanics, a sugar cube dissolved in water could spontaneously reassemble, and a cold pot of water could spontaneously come to a boil; it is simply not very likely. But unlikely events are quasi-certain to happen if we wait long enough, and indeed might happen tomorrow with the same (low) probability as for the day after a billion years from now. So the fact that we have never observed a cold pot of water spontaneously come to a boil, or even less significant instances of spontaneous increases in low entropy, remains an empirical difficulty for statistical mechanics.

Statistical mechanics is a far from universally accepted explanation of entropy, and while it does seem to allow for reversibility, which is compatible with mechanical physics, it also depends entirely on random motion, which is incompatible. If the defenders of statistical mechanics believe that it reconciles entropy theory with mechanical physics, they must also believe that if every atom in the universe happened to be traveling in the opposite direction to which it now moves, then heat would move from colder objects to warmer ones, and order would spontaneously appear.¹⁰ If the statistical view of entropy is correct, the gradual dispersion of material via physical and chemical erosion may not be entropy per se, because the physical and chemical erosion of matter is fundamentally different from the dissipation of heat. Regardless of the explanation, however, the end result and the practical implications are the same: Both matter and energy move irreversibly toward less-ordered states, and lower-entropy states can only be restored by converting low entropy to high entropy elsewhere in the system—and the increase in entropy elsewhere will be greater than the local decrease in entropy that it made possible.

Entropy and Life

If all matter-energy moves toward greater disorder, how, then, do we explain life? Is life not a form of spontaneous order that emerged from the chaotic maelstrom that was our early planet? Has not the continued evolution of life on Earth led to highly complex and ordered life forms? And don't ecosystems exhibit yet another level of complexity and order that

¹⁰N. Georgescu-Roegen, *The Entropy Law and the Economic Process*, Cambridge, MA: Harvard University Press, 1971.

arises from the mutual interactions of the organisms of which they are composed? These facts in no way contradict entropy, but to understand why this is so, we remember the distinction made in Chapter 2 between isolated, closed, and open systems. Isolated systems are those in which neither matter nor energy can enter or leave. The universe is such a system. The Earth, in contrast, is a materially closed system, in which radiant energy can enter and leave, but for all practical purposes, matter does not. The Earth is continually bathed in the low entropy of solar radiation that has allowed the complexity and order of life to emerge and increase. Any living thing on our planet is an open system, capable of absorbing and emitting both matter and energy.¹¹ A biological or ecological system is only capable of maintaining its low entropy by drawing on even greater amounts of low entropy from the system in which it exists, and returning high entropy back into the system. Erwin Schrodinger has described life as a system in steady-state thermodynamic disequilibrium that maintains its constant distance from equilibrium (death) by feeding on low entropy from its environment—that is, by exchanging high-entropy outputs for low-entropy inputs.¹² This exchange results in a net increase in entropy. Hence, life on our planet requires a constant flow of low-entropy inputs from the sun simply to maintain itself.

Entropy and Economics

What, then, are the implications of the entropy law for the science of economics? The goal of the early neoclassical economists was to establish economics as a science, and in the words of William Stanley Jevons, “it is clear that economics, if it is to be a science at all, must be a mathematical science.”¹³ The basic argument was that economics focused on quantities of goods, services, and money and therefore was amenable to quantitative (i.e., mathematical) analysis. Such analysis enabled economists to build logically consistent theories from fundamental axioms. These theories could then be applied to problems in the real world. In the words of Leon Walras, “from real type concepts, [the physico-mathematical] sciences abstract ideal-type concepts which they define, and then on the basis of these definitions they construct *a priori* the whole framework of their the-

¹¹H.E. Daly and J. Cobb, *For the Common Good: Redirecting the Economy Towards Community, the Environment, and a Sustainable Future*. Boston: Beacon Press, 1989, p. 253. Such open systems are often called “dissipative structures.” The nonequilibrium thermodynamics of dissipative structures is a field of thermodynamics under development by Nobel laureate physicist Ilya Prigogine and his collaborators. See his *The End of Certainty*, New York: Free Press, 1996.

¹²E. Schrodinger, *What Is Life?* Cambridge, England: Cambridge University Press, 1944.

¹³W. S. Jevons, quoted in R. Heilbronner, *Teachings from the Worldly Philosophy*, New York: Norton, p. 210.

orems and proofs. After that they go back to experience not to confirm but to apply their conclusions.”¹⁴ Mechanical physics was the best-developed and most successful application of this approach in the sciences at the time the original neoclassicals were writing, and thus was explicitly accepted as a model to emulate.¹⁵

In mechanical physics, all processes were considered reversible. For example, if one struck a billiard ball, an equal and opposite strike would return it exactly to its initial position. In contrast, the Second Law of Thermodynamics established the existence of irreversible processes as a fundamental law of physics. Entropy meant that in any isolated system, energy and matter would move toward a thermodynamic equilibrium in which they were equally diffused throughout the closed space. This implies an absence of temperature differentials and an inability to perform work. Quality, or order, was more important than quantity, and net quality changed in one direction only. The universe as a whole is an isolated system, and thus must be inevitably progressing toward a “heat death” in which all energy is evenly dispersed.

This notion was radical in the early nineteenth century and had profound implications for science as well as philosophy. If the laws of mechanical physics were universal, then the universe was governed by the same principles as a pool table. Not only was there no such thing as irreversible change, but if one could determine the position and velocity of every atom in the universe, one would know the past and could predict the future. Though this implies no free will, no alternatives and no sense in worrying about policy, it was during the nineteenth century the reigning worldview among scientists in the West and still holds considerable sway today. In the world of mechanical physics, the circular flow vision of economics discussed in Chapter 2 makes sense, as one can continually return to the same starting point. In a world where entropy reigns, it cannot.

Indeed, if we accept the laws of thermodynamics,¹⁶ the entire nature of the economic system is entropic. The First Law of Thermodynamics tells us that we cannot make something from nothing, and hence that all human production must ultimately be based on resources provided by nature. These resources are transformed through the production process into something of use to humans, and transformation requires work. Only low entropy or free energy can provide work. The First Law also ensures

¹⁴L. Walras, quoted in Heilbronner, *ibid.*, p. 225.

¹⁵Alfred Marshall, perhaps the most famous of the founding fathers of NCE, argued that in the future, the complex science of biology would provide a better model for economics, but in the meantime he relied extensively on the methodologies of physics. Heilbronner, *ibid.*

¹⁶Though in truth, physical laws, such as gravity, function the same whether we accept them or not!

that any waste generated by the economy cannot simply disappear but must be accounted for as an integral part of the production process. And the entropy law tells us that inevitably whatever resources we transform into something useful must disintegrate, decay, fall apart, or dissipate into something useless, returning in the form of waste to the sustaining system that generated the resource. The economy is thus an ordered system for transforming low-entropy raw materials and energy into high-entropy waste and unavailable energy, providing humans with a “psychic flux” of satisfaction in the process. Most importantly, the order in our economic system, its ability to produce and provide us with satisfaction, can only be maintained by a steady stream of low-entropy matter-energy, and this high-quality, useful matter-energy is only a fraction of the gross mass of matter-energy of which the Earth is composed.

THINK ABOUT IT!

Many people have proposed putting our toxic waste output onto rockets and shooting it into space. Based on your knowledge of thermodynamics, do you think this is a feasible solution for the pollution problem? Why or why not?

While we stress the fundamental importance of entropy to the economic process, we do *not* advocate an “entropy theory of value” similar to the classical economists’ “labor theory of value.” Value has psychic roots in want satisfaction, as well as physical roots in entropy. To propose an “entropy theory of value” would be to focus on the supply side only and neglect demand. And even on the supply side, entropy does not reflect many qualitative differences in materials that are economically important (e.g., hardness, strength, ductility, conductivity, etc.). On the other hand, any theory of value that ignores entropy is dangerously deficient.

■ STOCK-FLOW RESOURCES AND FUND-SERVICE RESOURCES

We now turn our attention to an important distinction between different types of scarce resources too often neglected by conventional economists—that of stock-flow and fund-service. Conventional economics uses the phrase “factors of production.” Factors of production are the inputs into a production process necessary to create any output. For example, when you make a pizza, you need a cook, a kitchen with an oven, and the raw ingredients. If you think about it carefully, however, you will clearly see that the cook and kitchen are different in some fundamental ways from the raw ingredients. The cook and kitchen are approximately the same before making the pizza as after, though just a bit more worn out. The raw ingredi-

ents, however, are used up, transformed first into the pizza itself, then rapidly thereafter into waste. The cook and kitchen are not physically embodied in the pizza, but the raw ingredients are. Thousands of years ago, Aristotle discussed this important distinction and divided causation (factors) into *material cause*, that which is transformed, and *efficient cause*, that which causes the transformation without itself being transformed in the process. Raw ingredients are the material cause, and the cook and kitchen are the efficient cause.

Other differences between these factors of production also exist. If we have enough raw ingredients to make 1000 pizzas, those ingredients could be used to make 1000 pizzas in one night, or one pizza a night for 1000 nights (assuming the ingredients were frozen and wouldn't spoil, and we had enough cooks and kitchens). The economy can use the existing stock of raw materials at virtually any rate, and time is not a factor. The productivity of raw ingredients is simply measured as the physical number of pizzas into which they can be transformed. In addition, as the ingredients for a pizza are produced over time, those ingredients can be used when they are produced, or stockpiled for future use. In contrast, while a cook or a kitchen may be capable of producing many thousands of pizzas over the course of their lifetimes, they can produce no more than a few pizzas in any given evening, even if limitless ingredients are available. The productivity of cooks and kitchens is measured as a number of pizzas per hour. However, this productivity cannot be stockpiled. For example, if we rest a cook for 6 nights, his capacity to produce a week's worth of pizzas cannot be used up all on the seventh night.

Georgescu-Roegen used the terms "stock" and "fund" to distinguish between these fundamentally different types of resources. A **stock-flow resource** is materially transformed into what it produces. A stock can provide a flow of material, and the flow can be of virtually any magnitude; that is, the stock can be used at almost any rate desired. Time does not enter into the equation, so the appropriate unit for measuring the production of a stock-flow resource is the physical amount of goods or services it can produce. Further, a flow can be stockpiled for future use. Finally, stock-flow resources are used up, not worn out. A **fund-service resource**, in contrast, suffers wear and tear from production but does not become a part of (does not become embodied in) the thing produced. Instead, a fund provides a service at a fixed rate, and the appropriate unit for measuring the service is physical output per unit of time. The service from a fund cannot be stockpiled for future use, and fund-service resources are worn out, not used up.¹⁷

¹⁷Georgescu-Roegen, *The Entropy Law*, op. cit.

Box 4-1**STOCK-FLOW AND FUND-SERVICE RESOURCES**

In the academic literature, there are many distinct definitions for stocks, flows, funds, and services. To make it clear, we are discussing the specific definitions given here. Future references will be to stock-flow and fund-service resources.

Stock-flow resources:

- Are materially transformed into what they produce (material cause).
- Can be used at virtually any rate desired (subject to the availability of fund-service resources required for their transformation), and their productivity is measured by the number of physical units of the product into which they are transformed.
- Can be stockpiled.
- Are used up, not worn out.

Fund-service resources:

- Are not materially transformed into what they produce (efficient cause).
- Can only be used at a given rate, and their productivity is measured as output per unit of time.
- Cannot be stockpiled.
- Are worn out, not used up.

The stock-flow and fund-service concepts are important when analyzing human production, and probably more so when focusing on the goods and services provided by nature. Note that “material cause” is always stock-flow in nature, and “efficient cause” is always fund-service.

THINK ABOUT IT!

Think about a specific ecosystem—or better yet, go visit one, and take along a field notebook. Make a list of three stock-flow resources provided by (or found in) that ecosystem, and three fund-service resources. (Note that you will need to be very specific about the use of each resource. For example, drinking water is a stock-flow, while water for swimming is a fund-service). Tick off the attributes of stock-flow and fund-service for each (see Box 4.1).

■ EXCLUDABILITY AND RIVALNESS

Excludability and rivalness are also crucial concepts for economic analysis, and rivalness is in fact related to the stock-flow, fund-service distinction. Though conventional economists first introduced these

concepts, they rarely receive the attention they deserve. We believe they are important enough to be described in some detail both here and in Chapter 10.

Excludability is a legal concept that when enforced allows an owner to prevent others from using his or her asset. An **excludable resource** is one whose ownership allows the owner to use it while simultaneously denying others the privilege. For example, in modern society, when I own a bicycle, I can prohibit you from using it. In the absence of social institutions enforcing ownership, nothing is excludable. However, the characteristics of some goods and services are such that it is impossible or else highly impractical to make them excludable. While someone could conceivably own a streetlight on a public street, when that streetlight is turned on, there is no practical way to deny other people on the street the right to use its light. There is no conceivable way that an individual can own climate stability, or atmospheric gas regulation, or protection from UV radiation, since there is no feasible institution or technology that could allow one person to deny all others access. When no institution or technology exists that makes a good or service excludable, it is known as a **nonexcludable resource**.

Rivalness is an inherent characteristic of certain resources whereby consumption or use by one person reduces the amount available for everyone else. A **rival resource** is one whose use by one person precludes its use by another person. A pizza (a stock-flow resource) is clearly rival, because if I eat it, it is no longer available for you to eat. A bicycle (a fund-service resource that provides the service of transportation) is also rival, because if I am using it, you cannot. While you can use it after I am done, the bicycle has worn out a bit from my use and is not the same as it was. A **nonrival resource** is one whose use by one person does not affect its use by another. If I use the light of a streetlight when riding my bike at night, it does not decrease the amount of light available for you to use. Similarly, if I use the ozone layer to protect me from skin cancer, there is just as much left for you to use for the same purpose. It is possible to deplete the ozone layer (through the emission of chlorofluorocarbons, for example), but depletion does not occur through use. Rivalness is a physical characteristic of a good or service and is not affected by human institutions.

Note that all stock-flow resources are rival, and all nonrival goods are fund-service. However, some fund-service goods are rival. For example, my bicycle is a fund that provides the service of transportation, but it is rival; the ozone layer is a fund that provides the service of screening UV rays, but it is nonrival.

As you will see when we turn to allocative mechanisms in subsequent chapters, the concepts of rivalness and excludability are very important.

THINK ABOUT IT!

For the list of resources you made earlier, answer the following questions:

Is the resource rival or nonrival? In general, can you think of any stock-flow resources that are nonrival? Can you think of any fund-service resources provided by nature that are rival?

Is the resource excludable or nonexcludable? (Note that excludability may differ depending on the specific value in question.) If it is nonexcludable, can you think of an institution or technology that could make it excludable? Do you think it should be made excludable? Why or why not?

Is the resource a market good or a nonmarket good?

In general, can you think of any stock-flow resources that cannot be made excludable? Can you think of fund-service resources provided by nature that can be made excludable?

■ **GOODS AND SERVICES PROVIDED BY THE SUSTAINING SYSTEM**

To make this discussion of entropy, fund-services, stock-flows, excludability, and rivalness more concrete, and to really understand the implications for economic theory and policy, we must see how these concepts apply to the specific scarce resources available to our economy—the goods and services provided by nature. We undertake this task in the next two chapters, and conclude this one by simply introducing the scarce resources.

For our purposes, we will present eight types of goods and services provided by nature, divided for convenience into nonliving and living resources. Clearly this is an enormous abstraction from the number and complexity of resources our Earth actually does supply, but these categories illustrate why the specific characteristics of goods and services we have described are of fundamental importance to economic policy.

1. *Fossil fuels.* For practical purposes, fossil fuels are a nonrenewable source of low-entropy energy. They are also very important as material building blocks.
2. *Minerals.* The Earth provides fixed stocks of the basic elements in varying combinations and degrees of purity, which we will refer to hereafter simply as minerals. This is the raw material on which all economic activity and life itself ultimately depends. Rocks in which specific minerals are found in relatively pure form we refer to as ores. Ores in which minerals are highly concentrated are a nonre-

newable source of low-entropy matter. We will refer to mineral resources and fossil fuels together as **nonrenewable resources**, and the first five goods and services in this list as abiotic resources (see Chapter 5).

3. *Water*. The Earth provides a fixed stock of water, of which fresh water is only a miniscule fraction. All life on Earth depends on water, and human life depends on fresh water.
4. *Land*. The Earth provides a physical structure to support us that is capable of capturing the solar radiation and rain that falls upon it. Land as a physical structure, a substrate, or a *site* has economic properties unrelated to the productivity of its soil, and is thus distinct from land as a *source* of nutrients and minerals. To capture this distinction, we will refer to land as a physical structure and location as Ricardian land.¹⁸ The quantity and quality of soil available on a given piece of Ricardian land will be grouped with minerals, discussed below.
5. *Solar energy*. The sustaining system provides solar energy, the ultimate source of low entropy upon which the entire system depends.
6. *Renewable resources*. Life is able to harness solar energy to organize water and basic elements into more useful structures (from the human perspective) that we can use as raw materials in the economic process. Only photosynthesizing organisms are capable of achieving this directly, and virtually all other organisms, including humans, depend on these primary producers. These biological resources are traditionally referred to as **renewable resources**, but they are only renewable if extracted more slowly than the rate at which they reproduce. Clearly, species can be exploited to extinction, so, as we shall see, biological resources are exhaustible in a way that mineral resources are not.
7. *Ecosystem services*. Living species interact to create complex ecosystems, and these ecosystems generate **ecosystem functions**. When functions are of use to humans, we refer to them as **ecosystem services**. Many of these ecosystem services are essential to our survival.
8. *Waste absorption*. Ecosystems process waste, render it harmless to humans, and, in most cases, again make it available to renewable resource stocks as a raw material input. This is really a specific type of ecosystem service, but one whose economic characteristics make it worth classifying on its own. We refer to these last three goods and services as biotic resources (see Chapter 6).

¹⁸Ibid., p. 232.

We refer to all the structures and systems that provide these goods and services as **natural capital**. In the following chapters, we will examine these resources in the light of entropy, fund-services, stock-flows, excludability, and rivalness.

BIG IDEAS to remember

- | | |
|--------------------------------|--|
| ■ Laws of Thermodynamics | ■ Fund-service resource |
| –Conservation of matter-energy | ■ Excludable and nonexcludable resources |
| –The law of increasing entropy | ■ Rival and nonrival resources |
| ■ Stock-flow resource | ■ Eightfold classification of resources |
-



Abiotic Resources

In this chapter, we closely examine the five **abiotic resources** introduced in Chapter 4: fossil fuels, minerals, water, land, and solar energy. Our goal is to explain how the laws of thermodynamics, the distinction between stock-flow and fund-service resources, and the concepts of excludability and rivalness relate to these resources, in order to better understand the role they play in the ecological-economic system. We will also assess the extent to which substitutes are available, and the degree of uncertainty associated with each resource. As we will see, however, abiotic resources are fundamentally different from each other, and it is their even greater dissimilarity from biotic resources that binds them together, more than their similarity to each other. Perhaps the most important distinction is that biotic resources are simultaneously stock-flow and fund-service resources that are self-renewing, but human activities can affect their capacity to renew. Abiotic resources are either nonrenewable (fossil fuels) or virtually indestructible (everything else).

The differences between abiotic resources probably deserve more emphasis than their similarities, and we'll start with a brief summary. Fossil fuels and mineral resources are frequently grouped together under the classification of nonrenewables. The laws of thermodynamics, however, force us to pay attention to an important difference: The energy in fossil fuels cannot be recycled, while mineral resources can be, at least partially. Water is one of the most difficult resources to categorize, precisely because it has so many different forms and uses. Fossil aquifers (those that are not being recharged) are in some ways similar to mineral resources—once used, the water does not return to the ground, and while it cannot be destroyed, it can become less useful when polluted by chemicals, nutrients, or salt. Rivers, lakes, and streams, in contrast, share similarities with biotic resources: They are renewable through the hydrological cycle driven

by solar energy, and they can exhibit stock-flow and fund-service properties simultaneously. However, human activity cannot affect the total stock of water to any meaningful extent, while we can and do irreversibly destroy biotic resources. Similarly, land as a physical substrate, a location (hereafter referred to as Ricardian land) cannot be produced or destroyed in significant amounts by human activity (with the exception of sea level rise induced by anthropogenic climate change), and solar energy flows are not meaningfully affected by humans at all—though we can affect the amount of solar energy that moves in and out of the atmosphere. We now examine these resources in more detail.

■ FOSSIL FUELS

Perhaps the simplest resource to analyze is **fossil fuels**, or hydrocarbons, upon which our economy so dramatically depends. In 1995, crude oil supplied 38% of energy inputs into the global economy, followed by coal at 25% and natural gas at 22%. In all, 85% of the energy in the economy comes from fossil hydrocarbons.¹ In geological terms and as far as humans are concerned, fossil fuels are a fixed stock. For a variety of reasons, however, it is extremely difficult to say precisely how large that stock is.

For practical purposes, we are only concerned with recoverable supplies. But what does recoverable mean? Clearly, hydrocarbons are found in deposits of varying quality, depth, and accessibility, and there are different costs associated with the extraction of different deposits. In economic terms, we can define recoverable supplies as those for which total extraction costs are less than the sales revenues. However, fossil fuel prices fluctuate wildly, and recoverable supplies defined in this way show similarly chaotic variation through time. We could also define recoverable supplies in entropic terms, in which case a hydrocarbon is recoverable if there is a net energy gain from extraction; that is, it takes less than a barrel of oil to recover a barrel of oil. This measure must include all the energy costs, including those of exploration, machinery, transportation, decommissioning, and so on. While technological change can reduce these, there is a certain irreducible limit to the energy costs of extracting fossil fuels. It takes 9.8 joules of energy to lift 1 kilogram 1 meter, and no amount of technology can change that basic fact.

As we deplete the most accessible hydrocarbon supplies first, over time it will take more and more energy to recover remaining supplies. In other words, the *energy return on investment*, which is “the ratio of gross fuel extracted to economic energy required directly and indirectly to deliver the

¹J. Edwards, *AAPG Bulletin* 81(8), August 1997, American Association of Petroleum Geologists.

fuel to society in a useful form,” declines over time.² In entropic terms, the energy cost of oil and natural gas extraction in the United States increased by 40% from 1970 to the 1990s.³ During the 1950s in the U.S., every barrel of oil invested in exploration led to the discovery of about 50 more. By 1999, the ratio was about one to five. Some experts predict that by the year 2005, the ratio will be one to one, and no matter what the price of oil, it will make no sense to search for more in the U.S.⁴ Still, under either the economic or the entropic definition of *recoverable*, estimates of recoverable reserves change constantly. Largely this is the result of new discoveries, but it also results from dramatically different methods for calculating “proven” supplies between different companies and different countries, with frequent changes often based on political or economic motives.⁵ Petroleum geologists can, however, assign reasonable probabilities to different estimates of total stocks.

Box 5-1 ESTIMATING OIL STOCKS

Every year, the world consumes in the neighborhood of 25 billion barrels of oil (Gbo). Yet at the end of most years, reported reserves of oil are greater than they were at the start, and there is a fairly wide range of estimates as to what those reserves actually are. The increase is possible as long as new oil discoveries are greater than oil consumed, but that is rarely the case anymore. For example, in 1997, the world used about 23 Gbo and discovered 7 Gbo, yet estimated reserves *increased* by 11 Gbo. How do we explain this anomaly?

When geologists estimate the quantity of oil in any given field, they assign a probability to the estimate. For example, in the late 1990s, geologists estimated that the Oseberg field in Norway would supply 700 million barrels of oil with 90% certainty (known as probability 90, or P90) and 2500 with 10% certainty (known as P10). Different corporations and countries generally use some number within the P10–P90 range when

²C. Cleveland, R. Costanza, C. Hall, and R. Kaufmann, Energy and the US Economy: A Biophysical Perspective, *Science* 225: 297 (1984).

³C. Cleveland and D. Stern, “Natural Resource Scarcity Indicators: An Ecological Economic Synthesis.” In C. Cleveland, D. Stern, and R. Costanza, eds., *The Economics of Nature and the Nature of Economics*. Cheltenham, England: Edward Elgar, 2001.

⁴J. Hanson, Energetic Limits to Growth, *Energy* (Spring 1999). Online: <http://www.dieoff.com/page175.htm>. This does not mean that perverse economic incentives cannot lead to oil being recovered beyond the 1:1 ratio. During the 1970s, in an effort to decrease dependence on imported oil, the U.S. government offered preferential prices for domestic oil. This meant that cheaper imported oil could be used to extract more expensive domestic oil, and domestic producers could profit even when it took more than one barrel of oil to extract a barrel.

⁵C. J. Campbell and J. H. Laherrère, The End of Cheap Oil, *Scientific American*, March 1998.

stating their reserves, and they are often purposefully vague about what number they use. Higher reported reserves can increase stock prices, provide greater access to credit, and for OPEC countries, increase their quotas. As oil fields are exploited, geologists can use the information acquired to make better estimates about how much they contain. Based on this information and other factors (e.g., moving from P90 to P50 estimates), countries frequently revise their reserve estimates from existing fields, and often upward. In the absence of major new discoveries or technological breakthroughs in the late 1980s, six OPEC countries alone revised their estimates upward by 287 Gbo, 40% more than all the oil ever discovered in the U.S.!

When calculating global oil reserves, it makes most sense to sum the P50 estimates across countries, but even this is no easy task. In addition, revised estimates from existing reserves are not new discoveries and should not be counted as such.^a

^aC. J. Campbell and J. H. Laherrère, The End of Cheap Oil, *Scientific American* (March 1998).

THINK ABOUT IT!

Economists argue that price reflects scarcity. Do you think the price of oil is a good indicator of how much oil is left in the ground? Why or why not?

Regardless of what the stocks of fossil fuels are, however, they are stocks that can be extracted as flows, and the rate of flow is largely determined by human efforts. If we had adequate infrastructure, we could theoretically extract all entropically recoverable fossil energy stocks in a single year, or we could make them last 1000 generations. How long recoverable stocks will last, therefore, is determined as much by how fast we extract them as by how much there actually is. We almost certainly will never exhaust fossil fuel stocks in physical terms, because there will always remain some stocks that are too energy-intensive or too expensive to recover. From this point of view, fossil fuel stocks are nonrenewable but not exhaustible.

As we extract fossil fuels, we will logically extract them from the most accessible and highest-quality known reserves first, where net energy gains are highest.⁶ These stocks essentially offer the lowest-entropy resource. Therefore, as we continue to extract fossil fuels over time, we can expect not only a *quantitative* decrease but also a *qualitative* decline in stocks. For example, the first oil to be extracted actually pooled on the surface and erupted in geysers from wells with no pumping. But as

⁶Note that the largest and most accessible reserves are also the most likely to be discovered first.

stocks diminish, it takes more and more energy to extract energy; ever-larger fractions of a barrel of oil are required as energy inputs to retrieve a barrel of oil as output, until we have reached entropic exhaustion. (As mentioned earlier, we are probably very near this point already in the U.S.)

Of course, resource exhaustion is only one component of fossil fuel use. Used fuel does not disappear; it must return to the ecosystem as waste. Acid rain, global warming, carbon monoxide, heat pollution, and oil spills are unavoidably associated with the use of fossil fuels. On a small scale, some of these wastes could be readily processed by natural systems, but on the current scale, they pose serious threats. Indeed, the growing accumulation of waste products from fossil fuel use and the negative impacts these have on planetary ecosystems is probably a far more imminent threat to human welfare than depletion; the sink will be full before the source is empty.

We must reiterate here that ecosystems (as shorthand for the primary producers they sustain) themselves capture solar energy, and humans make direct use of much of the energy they capture. If waste products from fossil fuel use diminish the ability of these ecosystems to capture energy, there are more energy costs to fossil fuel extraction than the direct ones discussed above. These costs are, however, several degrees of magnitude more difficult to measure—and therefore that much more likely to be ignored (Figure 5.1).

THINK ABOUT IT!

Many people are concerned by the United States' dependence on oil imports from a number of politically unstable regions and countries (e.g., the Middle East, Nigeria, Venezuela, Colombia). Proposed solutions to this problem have included increased domestic drilling and extraction, greater energy efficiency, and the development of renewable energy sources. What do you think are the pros and cons of each approach?

The basic equation here is:

$$\begin{aligned} \text{net recoverable energy from oil} = & (\text{initial total stock of entropically} \\ & \text{recoverable reserves}) - (\text{energy cost of extraction}) - \\ & (\text{loss of solar energy due to induced loss of capacity to capture}) \end{aligned}$$

To obtain net recoverable energy remaining in fossil fuel stocks, we also would need to subtract oil already consumed. Net energy from fossil fuels must account for the damage fossil fuel use causes to the ability of the sustaining system to capture solar energy, a fund-service resource. This lost capacity is measured as energy-flow/time, and we must account for the

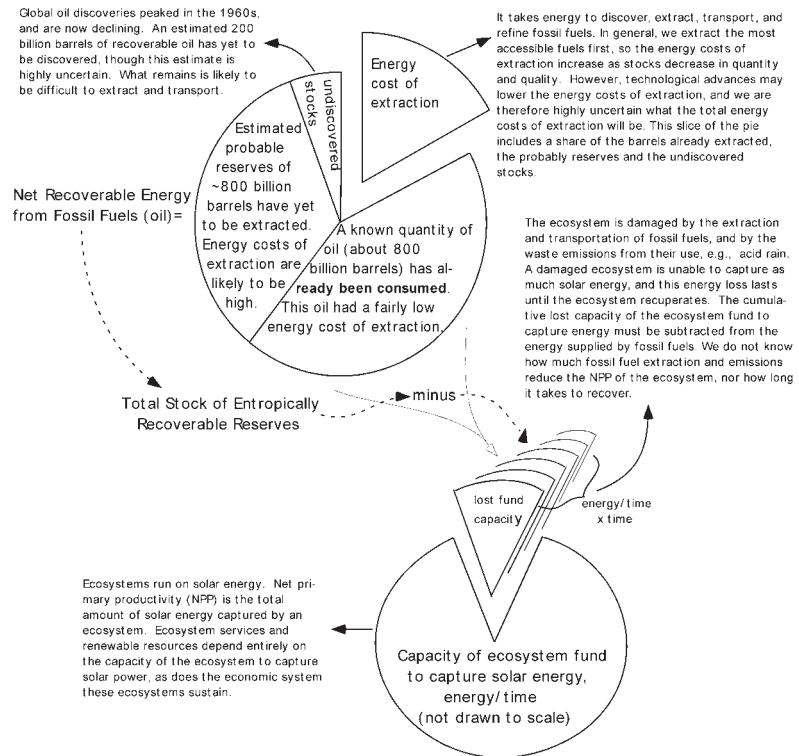


Figure 5.1 • Net recoverable energy from fossil fuels.

total amount of energy not captured from the time the damage occurs to the time the fund-service recovers.⁷

What points can we draw from this discussion of fossil fuels? First, once fossil fuels are used, they are gone forever—they are rival goods. While a seemingly trivial point, this has important implications for economic policy, as we will show in Chapter 11. Second, while fossil fuel stocks are finite, they are a stock-flow resource that can be extracted virtually as quickly as we wish, limited only by existing infrastructure, knowledge of stock locations, and the energy costs of extraction. We have control of the spigot and have been opening it a bit wider every decade. Eventually the reservoir must run dry. This is in stark contrast to flows of solar energy, as we pointed out in Chapter 4.

Third, our current populations and economic systems depend for survival on the use of fossil fuels. Fossil fuels not only supply 85% of our en-

⁷Estimates of oil already consumed, probable reserves, and oil yet to be discovered are from Campbell and Laherrère, *ibid*.

ergy needs, much of which is used to produce food; they also provide the raw materials for a substantial portion of our economic production, including ubiquitous plastics and, even more importantly, the fertilizers, herbicides, and pesticides that help provide food for 6 billion people. At this point, we do not have the technologies available to support 6 billion people in the absence of fossil fuels.

While we may be able to substitute renewable energy for fossil fuels, it is highly uncertain that we can do so before the negative impacts of their waste products force us to stop using them or the fuels themselves are depleted.

THINK ABOUT IT!

The U.S. and Canada have vast deposits of shale oil and tar sands, respectively. Both of these are fossil fuels, but of fairly low quality, requiring more energy to extract and process than conventional fossil fuels, and creating more associated waste. Do you think these resources present possible solutions to our energy problems? You may need to do some quick research on the Internet. Can you find any information on their energy returns to investment and waste outputs?

■ MINERAL RESOURCES

Though typically grouped together with fossil fuels in economics textbooks and labeled nonrenewable resources, minerals differ in important respects from fossil fuels. Like fuels, nonenergy minerals can be analyzed in terms of stocks and flows. We know the total stock is finite, and according to the First Law of Thermodynamics, this imposes a physical limit on their contribution to the material growth of the economy. Again, technology can increase the efficiency with which we extract minerals from ore, but there exists an entropic limit to efficiency. Valuable mineral deposits occur in varying degrees of purity, and, like fuels, the degree of purity can be looked at as a measure of low entropy. Highly concentrated ores are highly ordered low entropy.⁸ It is much easier to extract their mineral content, and they are much more valuable. As our growing economy depletes these most valuable ores first, we must move on to ores of lower and lower purity, incurring higher and higher processing costs.

As in the case of oil, we are not exactly certain of the total stock of any particular mineral, but geologists assign reasonable probabilities to different estimates. Even the most efficient process conceivable will require

⁸Even if we do not accept the notion of entropy in materials, concentrated ores require much less low-entropy energy to process.

some energy to extract minerals from an ore, and the less pure the ore, the more energy that will be required. Currently, mining accounts for about 10% of global energy use.⁹ However, unlike fossil fuels that cannot be burned twice, materials can be recycled (though this, too, requires energy). Therefore, we must think in terms of nonrenewable subterranean stocks as well as aboveground stocks, which accumulate as the subterranean ones are depleted. Still, we cannot avoid the laws of entropy even here, and use leads to dissipation through chemical and physical erosion; therefore, 100% recycling of any material may be impossible.

There is considerable debate over the impossibility of 100% recycling, as well as the implications. Georgescu-Roegen argues that because solar energy can provide a substitute for fossil fuels and nothing can provide a substitute for minerals, mineral depletion is actually more of a concern than fossil fuel depletion, and its inevitability means that a steady-state economy¹⁰ is impossible (see Box 3.2). In contrast, Ayres claims that even if all elements in the Earth's crust were homogeneously distributed (the material equivalent of "heat death" mentioned above), a sufficiently efficient solar-powered extraction machine would enable us to extract these elements,¹¹ presumably at a rate that would provide enough raw materials to maintain the machine and still leave a material surplus. This scenario implicitly assumes that damage caused by extracting all the resources from the Earth's crust in the first place, and their consequent return to the ecosystem as waste, would not irreparably damage the Earth's ability to capture solar energy and sustain life.

Alternatively, we may be able to master the art of creating polymers from atmospheric CO₂, which could provide substitutes for many of the minerals we currently use. If such polymers were biodegradable and simply returned to the atmosphere as CO₂ we would presumably be able to achieve 100% recycling (though in this case we may not want to, at least not before atmospheric CO₂ stabilizes at preindustrial levels). Of course, none of these propositions can currently be proven empirically. Nonetheless, it appears that mineral deposits are sufficiently large, and recycling has the potential to become sufficiently efficient, that with careful use, minimizing waste and appropriate substitution where possible, we could sustain a steady-state economy for a very long time.

Figure 5.2 depicts both the accumulation of extracted minerals into

⁹P. Sampat, From Rio to Johannesburg: Mining Less in a Sustainable World. World Summit Policy Brief #9. Online: <http://www.worldwatch.org/worldsummit/briefs/20020806.html>. (World Watch).

¹⁰N. Georgescu-Roegen, *The Entropy Law and the Economic Process*, Cambridge, MA: Harvard University Press, 1971.

¹¹R. U. Ayres, The Second Law, the Fourth Law, Recycling and Limits to Growth, *Ecological Economics* 29:473–484 (1999).

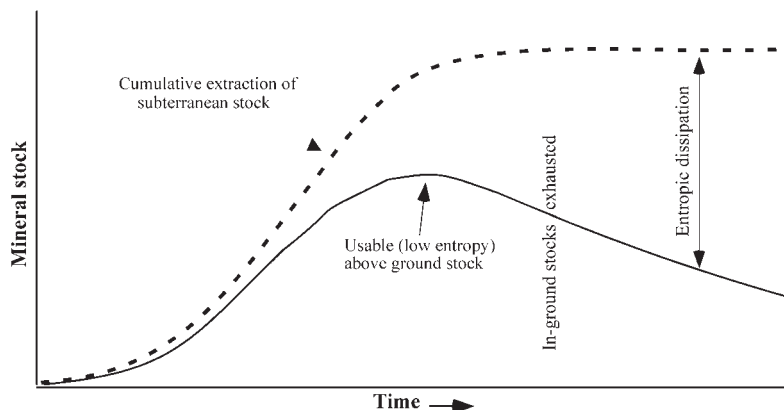


Figure 5.2 • The cumulative extraction of subterranean stock and aboveground stock of minerals over time. The distance between the two curves is a measure of entropic dissipation.

aboveground low-entropy stocks embodied in artifacts in use by (or available to) the economy (solid line) and the cumulative depletion (extraction) of subterranean stocks (dashed line) over time. We assume that initial rates of mineral extraction are low, but increase with economic growth and greater knowledge of reserve location. Eventually, however, stocks become scarcer, the costs of extraction become greater than the benefits, and extraction ceases. The point where this occurs is labeled “In-ground stocks exhausted” on the graph, and cumulative depletion ceases. In the absence of entropy, and if 100% recycling were possible and practiced, the two lines on the graph would be identical. In the real world, some portion of aboveground stocks dissipates into waste every year. The rate of increase in the aboveground stocks is equal to net annual mineral extraction minus entropic dissipation; that is, aboveground stocks are equal to minerals currently in use plus those that can be recycled.

There are two important categories of waste. Much waste is in the form of products that have stopped working, become obsolete, or simply gone out of fashion, and are discarded while still in a relatively ordered state. They are not recycled, because it is either cheaper or more convenient to extract virgin mineral flows from the Earth. For our purposes, this waste returns to the subterranean stock, though with higher entropy than the ore from which it was initially extracted.¹² Eventually, as we deplete the

¹²Some of this material will be in a highly ordered state and have lower entropy than the same amount of mineral in the form of an ore. Georgescu-Roegen distinguishes “garbo-junk” (a bald tire useless as a tire but recyclable) from “pure waste” (the dissipated rubber particles that are not recyclable). For practical purposes, however, large stocks of ore presumably still have an overall lower entropy; otherwise waste material would be processed before the ore.

most concentrated ores, it becomes cheaper to start mining the lowest-entropy waste. For example, slag heaps near old silver mines have been mined again with newer methods. But the slag heaps resulting from the second mining will be harder to mine.

Another type of waste results from entropy in the form of mechanical or chemical erosion of the material in question. Pennies eventually wear out through use—an atom rubbed off here, another there. Other metals rust away. Hence, gross subterranean stocks are never depleted (First Law of Thermodynamics). They simply become stocks of lower and lower entropy, akin to bound energy, and are no longer of use to humans (Second Law of Thermodynamics).

We have some control over the creation of waste in the form of discarded goods, but virtually none over the effects of entropy. The entropic limit to extraction in this case occurs when the extraction process consumes more material than it can provide. As we stated earlier, some people assert that this never happens, others assert it will happen soon enough to make a steady-state economy a pipe dream, and we take the middle road.

As more minerals are brought to the surface and put to use, entropy acts on a larger stock. As more subterranean stocks are extracted, the remainder becomes more difficult to find and extract. Therefore, even before in-ground stocks are exhausted, the rate of dissipation of aboveground stocks must become greater than the net extraction of new material, and aboveground stocks begin to decline. However, even after reaching the entropic limit to extraction, we are still likely to have a large stock of material in the economy that can be reused and recycled. Over time, of course, it must gradually erode away, atom by atom. In Figure 5.2, the distance between the lines depicting cumulative extraction and aboveground stocks measures cumulative dissipation, and eventually the entire aboveground stock must succumb to entropy. This process is probably slow enough that we could achieve a steady state through material recycling for a time sufficient that extinction through evolution would happen before extinction through resource depletion. However, as is the case with oil, the threat to us is probably more from the impacts of the waste itself than from the exhaustion of mineral resources. We'll put this discussion off until we get to the section on waste absorption capacity in Chapter 6.

What points can we draw from our discussion? First, mineral resources are rival goods at a given point in time. If I am using a hunk of steel in my car, it is not available for you to use. But through recycling, most of these resources could be made available for someone else to use in the future. Thus, we can think of *mineral resources as rival goods within a generation, but as partially nonrival between generations*, depending on how much is wasted and how much recycled. Fossil fuels are rival both within and be-

A resource is nonrival between generations if the use by one generation does not leave less of the resource for future generations.

tween generations. Second, stocks of low-entropy mineral ores are finite but can be extracted at virtually any rate we choose. In contrast to fossil fuels, we have control not only over the spigot of extraction, but also over the drain by which extracted materials return to the ecosystem as waste. We open the spigot wider almost every year and do very little to close the drain, but if we shut the drain as much as possible (though it will always leak some), the open spigot matters less to future generations. Third, we could not sustain existing populations or levels of economic production in the absence of these minerals. While it would clearly be impossible to develop substitutes for all minerals, thus far it has been reasonably easy to develop substitutes for specific minerals as they become scarce, and it may be possible to keep this up for some time to come.

■ WATER

Earth is a water planet. Though the stock of water is finite, fully 70% of the Earth's surface is covered in water. Freshwater, however, is far less abundant, accounting for less than 3% of the total, of which less than one-third of 1% is in the form of readily exploited lakes (0.009%), rivers (0.0001%), and accessible groundwater (0.31%). Another 0.01% is found in the atmosphere, 0.31% is deep groundwater, and over 2% is in the polar ice caps and glaciers.¹³ Humans are composed mostly of water, and in addition to drinking, we depend on it for agriculture, industry, hydroelectricity, transportation, recreation, waste disposal, and for sustaining the planet's ecosystem services. Water for different uses has different relevant characteristics that make generalizations difficult.

Water for drinking, irrigation, industry, and waste disposal is clearly a stock-flow resource, but a unique one. In contrast to fossil fuels and mineral deposits, many water resources are renewable as a result of the hydrologic cycle. However, for all practical purposes, many aquifers are "fossil" water, with negligible recharge rates. Many other aquifers are being mined; that is, the rate of water extraction is greater than the rate of replenishment. Even many rivers around the world, including the Colorado and the Rio Grande in North America, the Amu-Dar'ja and Syr-Dar'ja rivers that once fed the Aral Sea in Central Asia, and at times the Yellow, Hai, and Huai rivers in northern China are so heavily utilized (primarily for irrigation) that they never reach the sea.

At first glance, flowing water might appear to be a fund-service resource. In any stream or river at any given time, water is flowing at a specific rate, and the proper unit of measurement is volume/time (volume per

¹³P. Gleick, *The World's Water: The Biennial Report on Freshwater Resources*, Washington, DC: Island Press, 2002.

unit of time), as is the case for fund-service resources. Dams, however, allow us to stockpile flowing water for later use, which is a characteristic of stock-flow resources, and water is “used up” by drinking, irrigation, industry, and waste disposal but never “wears out.”

Perhaps the best way to look at flowing water is to distinguish it from the hydrologic cycle. The water itself is a stock-flow resource that is rapidly renewed by the service (provided by solar energy) of the hydrologic cycle. Hydroelectricity is not produced by water, but rather by the energy transferred to water by the hydrologic cycle—it is solar energy stored in water. Solar energy is generally a fund service, but when stored in water, it can be either a stock-flow or a fund-service resource. When mechanical energy in the water is converted to electric energy by a microhydro power plant that depends on river flow, it is essentially a fund-service resource. However, damming of the river allows the energy to be stockpiled by converting mechanical energy to potential energy, which is a stock-flow resource.

When used for transportation, recreation, or sustaining all other ecosystems on the planet, water functions as a fund-service resource. Atmospheric moisture, as part of the hydrologic cycle, is essentially a fund-service resource.

Like biotic resources, water can be a stock-flow and fund-service resource simultaneously. Unlike biotic resources, however, humans cannot meaningfully affect the total stock of water on the planet. We can and do reduce the stock of usable water, and while it is possible to restore the usability of water, there are no substitutes available for its most important uses.

As one would expect from its dual nature as a stock-flow, fund-service resource, water can be rival or nonrival depending on its use; stock-flow uses are rival, and fund-service uses are nonrival. However, as flowing water is recycled through the hydrologic cycle, it is intergenerationally nonrival. Excludability varies dramatically depending on existing institutions, though rainfall for all practical purposes is nonexcludable by nature.¹⁴

■ RICARDIAN LAND

Ricardian land—land as a physical substrate and location, distinct from its other productive qualities—is also a fund that provides the service of a substrate capable of supporting humans and our infrastructure, and of capturing solar energy and rain (Ricardian land does not include soil or the nutrients in the soil). A hectare of land may be capable of producing

¹⁴The seeding of clouds with silver nitrate can produce rainfall in a specific location, but for practical purposes, this is basically irrelevant.

1000 tons of wheat over 100 years, but one cannot produce that wheat from the same land in an appreciably shorter period, nor would it be possible to accumulate land's capacity as a substrate.

The services provided by land are certainly excludable, and at any given point in time, they are also rival. For example, if used for farming, land provides the service of a substrate for crops. If one farmer uses that service, no one else can in the same time period. Economists often use the term “depletable” as a synonym for “rival,” but the case of land suggests that this is inappropriate.¹⁵ Using Ricardian land does not deplete it. While rival within a generation, it is intergenerationally nonrival and absolutely nondepletable.

THINK ABOUT IT!

Why do you think we distinguish between Ricardian land as a physical substrate and the more conventional definition of land that includes the soil and its mineral content? Who or what creates value in Ricardian land? What makes land in one place more valuable than a similar piece of land elsewhere? Who or what creates value in fertile topsoil?

SOLAR ENERGY

The last abiotic producer of goods and services we will discuss is the sun. It shines to the Earth in 19 trillion tons of oil equivalent (toe) per year—more energy than can be found in all recoverable fossil fuel stocks—and will continue to do so for billions of years.¹⁶ Why then the fuss over the consumption of the Earth's fossil fuels?

While the flow of solar energy is vast, it reaches the Earth at a fixed rate in the form of a fine mist, and hence is very difficult to capture and concentrate. Most of the sunlight that strikes the Earth is reflected back into space.¹⁷ Over the eons, life has evolved to capture enough of this energy to maintain itself and the complex ecosystems that life creates. It would appear that the “order” of the global ecosystem over billions of years has reached a more or less stable thermodynamic disequilibrium. A better term is “meta-stable,” meaning that the global ecosystem fluctuates

¹⁵When “depletable” is used in this sense, it means that one person's use depletes the resource in question. Hence, the ozone layer is *nondepletable* because if I use it to protect me from skin cancer, it is still there for someone else to use. It is certainly possible to deplete the ozone layer with chemicals, but that is not a case of depletion caused by use.

¹⁶Unless otherwise cited, estimates of energy availability are from World Energy Council, *19th Edition Survey of Energy Resources*, London: World Energy Council, 2001. Online: <http://www.worldenergy.org/wec-geis/publications/reports/ser/overview.asp>.

¹⁷N. Georgescu-Roegen, *Energy and Economic Myths: Institutional and Analytic Economic Essays*, New York: Pergamon Press, 1976.

around a steady state rather than settling into one without further variation.¹⁸ Virtually all energy captured from the sun is captured by chlorophyll. In the absence of the evolution of some alternative physiological process for capturing sunlight, it would seem that our planet cannot sustain more low entropy than it currently does for any extended period. Yet through the use of fossil fuels, Americans are able to consume 40% more energy than is captured by photosynthesis by all the plants in the country. We also directly use over half of the energy captured by plants.¹⁹

As fossil fuels run out, we will need an alternative source of low entropy to maintain our economy at its current level of thermodynamic disequilibrium. The sun unquestionably radiates the Earth with sufficient energy to meet our needs, but how do we capture it? Global gross energy consumption is about 9 billion toe per year. Biomass, hydroelectricity, wind, photovoltaics, and wave/ocean thermal energy are all forms of solar energy we could potentially capture. Biomass is widely touted as a substitute for fossil fuels, but as we saw previously, converting *all* of the net primary productivity (NPP) of the United States to liquid fuel would still not meet our liquid fuel needs. Hydroelectricity currently provides 19% of global electricity, but even fully developed it could not supply 60%. Wind currently supplies little energy (about 17,500 MW in 2000), but it is a promising alternative: At current installation rates, capacity is *doubling* every 3 years.

Photovoltaics and wave/ocean thermal technologies still play very minor roles. With all of these technologies, however, large energy investments are required to produce the infrastructure needed to capture solar energy, and in many cases (e.g., photovoltaics), the energy returns on investment may be negligible. At the same time, human activity decreases the surface area of the planet covered in plant life and disrupts the ability of plants to capture sunlight. The net effect is likely to be an annual decrease in the amount of solar energy the Earth captures, and thereby a decrease in the complexity of the systems it is capable of maintaining. Figure 5.1 earlier illustrates the loss of solar energy capture that can be attributed to waste from fossil fuels.

While solar energy will bathe the Earth in more energy than humans will ever use, for practical purposes it is a fund-service resource that arrives on the Earth's surface at a fixed rate and cannot be effectively stored for later use.²⁰ As humans have negligible ability to directly allocate solar

¹⁸E. Laszlo, *Vision 2020*, New York: Gordon and Breach, 1994.

¹⁹D. Pimentel and M. Pimentel, *Land, Energy and Water: The Constraints Governing Ideal U.S. Population Size*, Negative Population Growth, Forum Series, 1995. Online: http://www.npg.org/forum_series/land_energy&water.htm.

²⁰Solar energy can be stored in fossil fuels, in batteries, or in the form of hydrogen for later use by humans, but this energy cannot subsequently be used to power photosynthesis, the most important function of solar energy.

energy, the fact that it is essentially nonrival and nonexcludable is not particularly important.

■ SUMMARY POINTS

Table 5.1 summarizes some of the policy-relevant characteristics of these five abiotic resources. Why are these details important to ecological economic analysis, and what message should you take home from this chapter? The stock-flow/fund-service distinction is important with respect to scale. We have control over the rate at which we use fossil fuels, mineral resources, and water. As the economy undergoes physical growth, it must use ever-greater flows from finite stocks. Because fossil aquifers and fuels are irreversibly depleted by use, and mineral resources may be irreversibly dissipated through use, the finite stock of these resources imposes limits on total economic production over time. Limits to growth are not apparent until the stock is nearly gone, and once gone, it is gone forever. Funds, in contrast, provide services at a fixed rate over which we have no control (though one thing that distinguishes biotic fund services from abiotic ones is that we can damage or even destroy them). Fund-services therefore limit the size of the economy at any given time, but they do not limit total production over time.

■ Table 5.1

SELECTED POLICY-RELEVANT CHARACTERISTICS OF ABIOTIC RESOURCES					
Abiotic Resource	Stock-Flow or Fund-Service	Can Be Made Excludable	Rival	Rival Between Generations	Substitutability
Fossil Fuels (nonrecyclable)	Stock-flow	Yes	Yes	Yes	High at margin, time is important factor, but possibly substitutable
Minerals (partially recyclable)	Stock-flow	Yes	Yes	Partially	High at margin, ultimately nonsubstitutable
Water (solar recycling)	Context-dependent	Context-dependent	Context-dependent	Stocks, yes; funds and recycled, no	Nonsubstitutable for most important uses
Ricardian Land (indestructible)	Fund-service	Yes	Yes	No	Nonsubstitutable
Solar Energy (indestructible)	Fund-service	No	No, for practical purposes	No	Nonsubstitutable

Water sources are a complex mix of stock-flow and fund-service. But even the stock-flow uses of water are completely recyclable—in particular, running water is so closely linked to the fund-service of the solar-powered hydrologic cycle that it acts much like a fund-service, imposing limits on the output of the economy only at a given point in time.

Substitutability is also relevant to scale. If we can develop a substitute for a resource, then the constraints it imposes on scale are less rigid. However, developing substitutes generally relies on technology, and technology takes time to develop. In addition, truly innovative technologies are impossible to accurately predict—we could only predict one if we already knew what it would look like in which case it would not be truly innovative.

Rivalness is relevant primarily to distribution, both within and between generations. All abiotic resources are rival except for water in some of its forms and uses, and solar energy (for practical purposes). One person's use of these rival resources means they are not available for others to use, and we must be concerned about distribution within a generation. People can use the nonrival resources of solar energy and water in its fund-service functions without leaving less for anyone else, and all else being equal, we should, therefore, let anyone use them. When a good is nonrival between generations, we needn't worry about excessive use within a generation. When addressing distribution, we must remember that all natural resources are produced by nature, not humans.

Excludability is primarily relevant to allocation. The market cannot allocate nonexcludable goods, and other allocative mechanisms are required. However, in the case of sunlight and rainfall, allocation by human institutions is simply not feasible.

BIG IDEAS to remember

-
- | | |
|---|--|
| ■ Big ideas from Chapter 4 that recur in Chapter 5 | ■ Aboveground and subterranean mineral stocks |
| ■ Ricardian land | ■ Entropic dissipation |
| ■ Energy return on investment | ■ Rival within versus between generations |
| ■ Recoverable reserves | ■ Garbo-junk versus pure waste |
| ■ P ₁₀ and P ₉₀ reserve estimates | ■ Unique characteristics of water and solar energy |
| ■ Net recoverable energy from fossil fuels | |
-



Biotic Resources

Biotic resources include the raw materials upon which economic production and human life depend, the ecological services that create a habitat capable of supporting human life, and the absorption capacity that keeps us from suffocating in our own waste. As nonrenewable resources are exhausted, human society will come to rely more and more on the self-renewing capacity of biotic resources. It is therefore critically important that we understand the nature of these resources.

As we turn our attention from abiotic resources to biotic ones, we must address a quantum increase in complexity, inevitably accompanied by a quantum increase in ignorance and uncertainty. One level of complexity arises from the intrinsic value we give to living systems. Abiotic resources are almost entirely considered means to various ends, where one of the foremost ends is the sustenance of life, the maintenance of biotic resources. Biotic resources not only enhance human well-being directly, they are also considered by many to be an end in their own right, especially in the case of sentient creatures. Biotic resources are also physically complex in two ways. First, the processes responsible for the sustained reproduction of individuals, populations, or species are highly complex and poorly understood. Second, individuals, populations, and species interact with other individuals, population, and species, as well as abiotic resources, to create an ecosystem. Ecosystems are extraordinarily complex and dynamic, changing over time in inherently unpredictable ways. The differences between these two types of physical complexity bear closer examination.

■ ECOSYSTEM STRUCTURE AND FUNCTION

Ecologists look at ecosystems in terms of structure and function, corresponding to the two types of physical complexity mentioned above. This

distinction is very relevant to economic analysis. Conventional natural resource economics is essentially the economics of ecosystem structure. Environmental economics focuses on certain ecosystem functions. In reality, structure and function are mutually interdependent, and we need an economics that effectively integrates both. Certainly we must understand the distinctions and interactions between the two if we are to incorporate them into economic analysis.

Ecosystem structure refers to the individuals and communities of plants and animals of which an ecosystem is composed, their age and spatial distribution, and the abiotic resources discussed in Chapter 5.¹ Most ecosystems have thousands of structural elements, each exhibiting varying degrees of complexity. Scientists have learned that when enough separate elements are thrown together into a complex system, a sort of spontaneous order results. One property of such systems is their tendency to generate emergent phenomena, which can be defined as properties of the whole that could not be predicted from an understanding of the individual parts, no matter how detailed that understanding. Complex systems are also characterized by highly nonlinear behavior, which means that we cannot predict the outcomes of large interventions based on an understanding of smaller ones. For example, removing 40% of a species stock from an ecosystem may have a qualitatively different impact than removing 20%—that is, not just twice the known impact of removing 20%.

In an ecosystem, the structural elements act together to create a whole that is greater than the sum of the parts. We refer to these emergent phenomena in ecosystems as **ecosystem functions**,² and they include such things as energy transfer, nutrient cycling, gas regulation, climate regulation, and the water cycle. As is typical of emergent properties, ecosystem functions cannot be readily explained by even the most extensive knowledge of system components.³ Variability, ignorance, and uncertainty play an extremely important role in the analysis of ecosystem structure, and a far greater role in the analysis of ecosystem function. We have a very limited understanding of exactly how ecosystem functions emerge from the complex interactions of ecosystem structure, and thus a difficult time

¹It may seem strange to include such things as fossil fuels and mineral deposits as elements of ecosystem structure, but we must not forget that humans are part of the global ecosystem, and these resources affect our ability to thrive.

²Whether a particular element of an ecosystem is part of structure or part of function depends on perspective. Organelles are part of the structural components of a cell that enable the cell to function. Cells are structural components of an individual that enable the individual to function. In the same way, individuals are part of the structure of a population, a population is part of the structure of a local ecosystem, an ecosystem is part of the structure of a landscape, and a landscape is part of the structure of the global ecosystem.

³E. Odum, *Ecology: A Bridge Between Science and Society*, 3rd ed., Sunderland, MA: Sinauer, 1997.

Box 6-1 RISK, UNCERTAINTY, AND IGNORANCE

Whenever we do not know something for sure, we are uncertain, but there are different types of uncertainty. When I throw dice, I cannot say in advance what the outcome will be, but I do know the possible outcomes and their probabilities. This type of uncertainty is referred to as risk. Pure uncertainty occurs when we know the possible outcomes, but cannot assign meaningful probabilities to them. Ignorance or absolute uncertainty occurs when we do not even know the range of possible outcomes.

In economics, Frank Knight pointed out that risk is a calculable or insurable cost, while pure uncertainty is not. In his view profit—the difference between revenue and calculable risk-adjusted costs—is a return for willingness to endure pure uncertainty. However, Knight was discussing the case where the entrepreneur bore the costs of failure and reaped the rewards of success. In economic decisions regarding exploitation of ecosystems, it is often the entrepreneur who reaps the rewards, while society bears the costs.^a

Discoveries in quantum physics and chaos theory suggest that uncertainty and ignorance do not result simply from a lack of knowledge, but are irreducible, inherent properties in certain systems. For example, chaos theory shows that even in a deterministic (i.e., nonrandom) system, extremely small differences in initial conditions can lead to radically different outcomes. This has been popularized as the ‘butterfly effect,’ in which a butterfly flapping its wings over Japan can create a storm in North America.

Change in highly complex systems is characterized by ignorance, especially over long time spans. We cannot predict evolutionary change in organisms, ecosystems, or technologies. For example, while we can predict that computers will continue to get faster and cheaper, we cannot predict what the next big technology will be 50 years from now. Leading experts are often notoriously wrong even when predicting the future of existing technologies. Bill Gates once predicted that no one would ever need more than 540 kilobytes of computer memory.

Estimating stocks of natural resources, or reproductive rates for cultivated species is basically a question of risk. Estimating reproductive rates for wild species is a question of uncertainty, since we cannot accurately predict the multitude of factors that affect these reproduction rates, but we do know the range over which reproduction is possible. Estimating ecological thresholds, conditions beyond which ecosystems may flip into alternative states, is a question of pure uncertainty, since we have limited knowledge of ecosystems and cannot predict the external conditions that affect them. Predicting the alternate state into which an ecosystem might flip when it passes an ecological threshold, and how humans will adapt, are cases of absolute ignorance involving evolutionary and technological change.

^aF. H. Knight, *Risk, Uncertainty, and Profit*, Boston: Houghton Mifflin, 1921; *Library of Economics and Liberty*, Feb. 21, 2002. Online: <http://www.econlib.org/library/Knight/knRUPl.html>.

predicting and managing the impacts of human actions on these functions. Therefore, a great deal of uncertainty attends decision making involving ecosystem functions. How we choose to treat uncertainty in economic analysis is ultimately a normative (ethical) decision, yet another source of complexity. One of the most important issues concerning any analysis of biotic resources is the degree of uncertainty involved.

Concrete examples always help clarify a concept. To illustrate the links between structure and function, and the implications of complexity, let's focus on a wet tropical forest, the terrestrial ecosystem that exhibits the greatest biodiversity of any yet studied. The forest is composed of individual plants (part of ecosystem structure). Each plant alone has little impact on climate, nutrient cycling, and habitat provision, and may even be unable to reproduce. However, when we bring together hundreds of millions of plants, as in the Amazon or Congo basin, these and other ecosystem functions emerge.

The forest canopy filters out about 98% of the sunlight at ground level, dramatically reducing daytime temperatures. It traps air and insulates, increasing night temperatures under the canopy and maintaining high and constant humidity. Trees absorb the energy of tropical storms, aerate the soil to allow water absorption, and slow water flows—all of which prevent soil and nutrients from being washed out of the system. Trees create the microclimate and habitat essential to the soil fauna that help recycle nutrients, facilitating their reabsorption by the system.

On a regional scale, the water retained by forest structure is absorbed and returned to the atmosphere through evapotranspiration, increasing humidity over the forest. Greater humidity increases the frequency of rainstorms. Estimates for the Amazon forest suggest that it generates up to 50% of its own rainfall, enabling the water-dependent species there to thrive. Without the increased absorption capacity of the soil and the evapotranspiration it facilitates, rainfall would simply drain into the rivers and be flushed from the system forever.

On an even larger scale, forests absorb up to 90% of the solar energy that strikes their canopies. Much of this is released through evapotranspiration and carried high up into the atmosphere, where it is carried into the temperate zones, helping stabilize the global climate (a function provided by carbon sequestration as well).

The species and populations in the forest cannot survive without a stable climate and a steady nutrient flow. Loss of forest structure can degrade forest function to the point where the forest spontaneously declines, creating a positive-feedback loop with potentially irreversible and catastrophically negative consequences.⁴ Computer models suggest that

⁴J. Farley, "Optimal" Deforestation in the Brazilian Amazon: Theory and Policy. Ph.D. dissertation, Cornell University, 1999.

regional climate change following extensive deforestation in the Amazon will prevent the regrowth of rainforest.⁵ Other research suggests that deforestation of 50% of the Amazon could lead to no rainfall at all during much of the year, conditions under which the forest may enter spontaneous decline,⁶ or at the very least become dramatically more susceptible to massive fires, such as those that occurred in 1997 in the Amazon, Indonesia, and Mexico.

In other words, ecosystem structure interacts to create ecosystem functions, and the structural elements depend on these functional attributes for their own survival. Owing to the complex nature of the whole system, as structural elements of an ecosystem are lost, in most cases we cannot say for sure to what extent ecosystem functions will be affected. Similarly, as ecosystem functions change in response to human impacts or nonanthropogenic change, we cannot say for certain what the impact will be on ecosystem structure.

THINK ABOUT IT!

In Chapter 4, we asked you to make a list of stock-flow and fund-service resources provided by a local ecosystem. Which of these resources are elements of ecosystem structure? Which are elements of ecosystem function? Do you see any links between these classifications?

Roughly speaking, conventional and natural resource economics has focused on ecosystem structure, while conventional environmental economics has focused on certain elements of ecosystem function, with a major emphasis on waste absorption capacity and the monetary valuation of other functions. In reality (as many conventional economists are fully aware), ecosystem structure and function are mutually interdependent, and conclusions based on the analysis of one dimension may not apply to the multidimensional case. With this caveat in mind, we now turn our attention to specific categories of biotic resources.

Three basic categories of biotic resources deserve attention. First are **renewable resources**, the elements of ecosystem structure that provide the raw materials for economic processes. Second are **ecosystem services**, defined as the ecosystem functions of value to humans and generated as emergent phenomena by the interacting elements of ecosystem structure. Third is **waste absorption capacity**, an ecosystem service that is sufficiently distinct from the others to warrant separate treatment.

⁵R. Monastersky, Amazon Forest Unlikely to Rise from Ashes, *Science News* 11:164 (1990).

⁶I. Walker et al. "Climatic and Ecological Conditions as Key Factors for Ecodevelopment Strategies." In M. Clusener-Godt and I. Sachs, eds. *Brazilian Perspectives on Sustainable Development of the Amazon Region*, Paris: UNESCO and Parthenon Publishing Group, 1995.

■ RENEWABLE RESOURCES

For simplicity, we can treat biological resources as material stock-flow resources, that is, as elements of ecosystem structure. Like nonrenewable resources, biological stocks can be extracted as fast as humans desire, but they are capable of reproduction. Figure 6.1 depicts a renewable natural resource in stock-flow space: the X-axis depicts the stock, or amount of resource that exists, and the Y-axis depicts the flow. The flow in this case can be the rate of reproduction (or biomass increase) that is likely for any given stock, or the rate of extraction (harvest). A 45-degree dashed line shows the theoretically maximum rate at which we can extract a given stock (i.e., we can extract the entire stock at one time; stock = flow). Actual extraction rates must lie on or below this line. We have also drawn a curve that shows the growth rate of each level of stock, which is also the sustainable yield curve. The sustainable yield is the net annual reproduction from a given stock—for every population of a resource, there is an associated average rate of population increase, and that increase represents a sustainable harvest that can be removed every year without affecting the base population.

We must caution here that there is a great deal of uncertainty concerning the position of this sustainable yield curve. Not only do we not know at precisely what rate a given population will reproduce; we are also uncertain of the exact population of any given species, though this is more true for animals than for plants, since plants sit still for the census takers while animals do not. While with careful study and census techniques we can assign reasonable probabilities to population estimates for renewable resource stocks (risk), there is greater uncertainty of a qualitatively different type concerning reproduction rates, particularly because these rates depend on a host of “external” factors such as rainfall, abundance of predator and prey species, disease, and so on. In addition, habitat destruction and degradation, pollution, climate change, and other human impacts can profoundly affect the entire curve, shifting it dramatically over time. Thus, in any given year, the actual rate of increase from a given population stock may be wildly different than the average.

As most people are at least vaguely aware, stocks of plants and animals in nature cannot grow forever. Instead, populations reach a point where they fill an available niche, and average death rates are just matched by average birth rates. Populations “stabilize” around an equilibrium, known as the **carrying capacity**. (We use the term *stabilize* loosely, because populations fluctuate in the short term depending on weather conditions, predator-prey cycles, etc. and in the longer term depending on a wide variety of factors. To paraphrase John Maynard Keynes, in the very long term, all species go extinct.) At carrying capac-

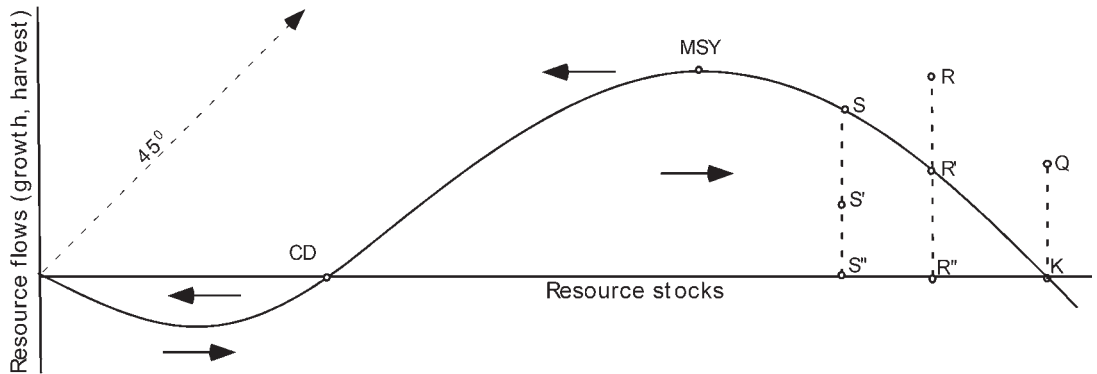


Figure 6.1 • The growth, or sustainable yield, curve.

ity, there is just enough food and habitat to maintain the existing population, and the rate of growth in biomass is zero (point K in Figure 6.1). Obviously, the rate of growth of a stock is also zero when the stock has been driven extinct (the origin in Figure 6.1). In between these two points, things get interesting.

The growth, or sustainable yield, curve for a renewable natural resource indicates the increase in stock over one time period for any given stock (see Figure 6.1). The Y-axis can measure growth or harvest, that is, flows from or to the existing stock depicted on the X-axis. Any harvest up to the total stock is theoretically possible. A harvest at any point on the sustainable yield curve, such as S, is just equal to the growth in the stock, and hence has no net impact on stocks. Harvests above the sustainable yield curve deplete the resource, and harvests below the curve lead to an increase in the stock, as indicated by the arrows. For example, a harvest at R will reduce the stock to S'', and a harvest at S' will allow the stock to increase to R''. CD represents the critical depensation level, which is the level below which a species or stock cannot sustain itself even in the absence of harvest. The **critical depensation** level is equivalent to the **minimum viable population**, and K represents carrying capacity. The graph suggests extremely rapid growth rates—about 30% per time period at **maximum sustainable yield (MSY)**, which is the average maximum catch that can be removed under existing environmental conditions over an indefinite period without causing the stock to be depleted, assuming that removals and natural mortality are balanced by stable recruitment and growth. While this may be appropriate for small, rapidly reproducing species, growth rates for many economically important species are on the order of 1% per year or less.

THINK ABOUT IT!

Prior to the arrival of humans in North America, where on the graph in Figure 6.1 would you locate American bison populations? Ten thousand years later, approximately where on the graph do you think bison harvests occurred? After the introduction of the horse, where do you think they occurred? After the introduction of the rifle? After the settling of the frontier by Europeans? What impact would the conversion of the Great Plains to agriculture have on the sustainable yield curve for bison?

What happens when we remove some fish, for instance, from a population at carrying capacity? In terms of the graph, we represent this as a harvest at point Q. As there is zero net productivity at our starting point K, harvest Q reduces the stock of fish by the quantity Q-K, and we find ourselves at stock R'. At the lower population stock, there are fewer fish competing for available food, shelter, and breeding grounds, and the remaining fish get more of each than they would under more crowded conditions. This greater resource abundance per individual leads to increased growth rates and fertility. With less competition for spawning areas, a higher percentage of eggs are laid in desirable locations, thereby increasing recruitment.⁷ In addition, most species grow fastest in their youth (as a percentage of biomass) and slower as they age. If we harvest larger, older fish, the remaining population has a higher percentage of recruits and potentially faster net growth.

At stock R', we could continue annual harvests forever at point R', just equal to the annual rate of increase of the stock. Any harvest below the sustainable yield curve would be less than the annual growth rate. Flows can accumulate into stocks, and the population would increase. Any harvest above the sustainable yield curve would reduce the stock even further. The arrows above and below the curve indicate the direction of change in stock for harvests in each of these regions. For example, a harvest at point R would reduce the population to stock S'. At S', per-capita resource abundance would be even greater than at R', increasing net annual growth and sustainable harvest from R' to S.

Over a certain range, lower population stocks can generate higher sustainable yields, but this obviously cannot go on forever. Eventually, the breeding population is insufficient to sustain high yields. Insufficient numbers of eggs lead to insufficient recruits, reducing yield in spite of superabundant resources per individual within the population. This means that at some point there is a maximum sustainable yield, MSY on Figure

⁷Note that a harvest at point R decreases the stock in the subsequent period by an amount equal to R-R' (as measured on the Y-axis), and a harvest at point S' increases the next period's stock by S-S'. Thus, R-R' and S-S' as measured on the Y-axis equal S'-R' as measured on the X-axis.

Box 6-2**CRITICAL DEPENDSATION, MAXIMUM SUSTAINABLE YIELD, AND UNCERTAINTY**

Using the notion of maximum sustainable yield to help communicate ideas in ecological economics is a far cry from using the concept as a tool in resource management. In reality, the MSY will vary dramatically from year to year in response to climatic cycles such as the El Niño Southern Oscillation (ENSO), changes in populations of predator and prey species of the species of interest, changes in pollution levels, and a broad array of other ecological changes and cycles. Under the most stable of conditions, natural variability can mask the effects of economic exploitation, and the scale of human impacts is rapidly changing the global ecosystem. Science relies on replication and controls, and neither of these is possible when dealing with a unique species in a highly complex and rapidly changing ecosystem. We cannot scientifically estimate MSY accurately enough for use in resource managements.^a

Where the critical dependsation point lies, and whether or not one exists for a given species or population, is similarly marked by extreme uncertainty. When a population becomes small enough, it is more susceptible to stochastic events and the negative impacts of inbreeding. For the North American passenger pigeon, probably once the most numerous bird on Earth, numbering in the billions but driven extinct in a matter of decades, it appears the minimum viable population was quite high, perhaps because of its colonial nesting habit. At the other extreme, the Mauritius Kestrel has rebounded from a known population of 6 in 1974 to a present population of over 600, though it would have almost certainly gone extinct without substantial conservation efforts, and inbreeding may make it highly susceptible to disease or other stochastic shocks. There is currently a debate over whether or not the blue whale and some populations of North Atlantic cod are below their level of critical dependsation. Obviously, experiments to scientifically determine critical dependsation points could not be replicated, as the first trial could wipe out the species in question.

On the other hand, uncertainty as to where the MSY is does not mean that we are not overshooting it, nor does it relieve policy makers of the responsibility to decide acceptable levels of offtake. As a general rule, the higher the uncertainty, especially in the presence of irreversibility (extinction), the more conservative (i.e., lower) should the allowed offtake be.

^aD. Ludwig, R. Hilborn, and C. Walters. *Uncertainty, Resource Exploitation, and Conservation: Lessons from History*, Science 260:17, 36 (1993).

6.1. Again, we caution that the MSY may vary dramatically from year to year, and there is probably no way we can accurately estimate it in any given year. Although it is useful as a pedagogical device, it has very little value as a calculating device for setting annual quotas.⁸ If harvests continue to be greater than annual yield, the population will continue to fall. Eventually, fish will become too scarce to find each other for breeding, and other ecological mechanisms (many only poorly understood) can break down. This means that at some population above zero, we may reach a minimum viable population, at which point the rate of growth is zero. Below this point the population enters into spontaneous decline; that is, death rates exceed birth rates.

Note that the sustainable yield curve enters into negative numbers below the critical depensation point. This means that in order to sustain the population, a negative harvest is required; that is, recruits have to be placed into the system every year just to maintain the existing stock. Harvest is still possible anywhere below the 45-degree line, but this will simply lead to even more rapid extinction (or extirpation) of the population. Unfortunately, we do not know where the critical depensation point occurs and can only make very rough (and highly contentious) estimates of where the maximum sustainable yield occurs.

The same basic concepts explained here apply to plant species or even plant communities. For example, what happens when we clear some trees from a virgin forest for timber? Light, water, and nutrients become available to the other trees already there, accelerating their growth, and space becomes available for new seedlings to sprout. This process can initially speed up the growth rate and increase the sustainable yield of a forest. Soon, however, sources of new seeds become more distant from the cleared land, and recruitment slows. Nutrients are lost from the soil as trees are removed. Trees of the same species are too far apart to cross-pollinate, resulting in sterile seeds or the problems of inbreeding. The sustainable yield begins to fall. And as we described above, the removal of ecosystem structure can dramatically affect ecosystem function, with the potential of further reducing the capacity of the forest to reproduce itself. Thus, like animal populations, a stock of forest will show a maximum sustainable yield and may exhibit critical depensation.

We will return to this analysis in Chapter 12 when we examine the microeconomics of biotic resource allocation.

⁸D. Ludwig, R. Hilborn, and C. Walters. Uncertainty, Resource Exploitation, and Conservation: Lessons from History, *Science* 260:17, 36 (1993).

Box 6-3**CASSOWARIES AND CRITICAL DEPENDENCE OF ECOSYSTEMS**

In the rainforests of northeastern Australia, up to 100 species of large seeded fruit trees depend almost entirely on a single bird species for distribution. This bird, the cassowary, is a large ratite, an ostrich-like bird that lives in the forest. It is the only animal known in the region capable of swallowing and transporting very large seeds, up to 2 kg of which can be found in a single scat. Evidence suggests that some seeds must pass through the digestive tract of a cassowary before they can germinate.

Cassowaries need large home territories to survive, especially in the highland forests. As forests are cleared in a patchwork pattern, few areas remain that can sustain a viable cassowary population. Without cassowaries, many trees in the region will be unable to disperse, and some may not even be able to germinate. Eventually, these species are likely to go extinct. Other plants and animals depend on these species, and they too will go extinct, igniting a chain reaction of extinction in species that may in turn depend on them. The net result could be a dramatic change in forest composition, leading to a qualitatively different ecosystem. The entire process, however, could take a very long time. It might not be noticed until centuries after it is too late.^a

Such examples of mechanisms for critical dependence are just a few of the possibilities that have been proposed. Again, we must emphasize that we really have little idea where maximum sustainable yields or critical dependence points lie. Ignorance, uncertainty, and variability are our constant companions in the real world.

^aJ. Benrupperbaumer, *Conservation of a rainforest giant*. Wingspan 8(Dec.): 1–2. (1992). Also extensive personal communications.

■ ECOSYSTEM SERVICES

In our discussion of ecosystem structure and function, we explained why forests need the functions generated by forests to survive, but we also hinted at the presence of extensive benefits that ecosystem functions provide for humans. We call an ecosystem function that has value to human beings an ecosystem service. For example, forested watersheds help maintain stable climates necessary for agriculture, prevent both droughts and floods, purify water, and provide recreation opportunities—all invaluable services for watershed inhabitants. But ecosystems provide many more services, of course. Unfortunately, we are unsure exactly how ecosystem structure creates ecosystem services, and we are often completely unaware of the services they generate. For example, prior to the 1970s, most people were unaware that the ozone layer played a critical role in making our

planet habitable.⁹ If we also take into account the tightly interlocking nature of ecosystems, it's safe to say that humans benefit in some way from almost any ecosystem function.

We just described forests as a stock of trees that generates a flow of trees. Now we want to look at the forest as a creator of services; as such, they are very different from a stock of trees. A stock of trees can be harvested at any rate; that is, humans have control over the rate of flow of timber produced by a stock of trees. Trees can also be harvested and used immediately, or stockpiled for later use. Ecosystem services are fundamentally different. We cannot use climate stability at any rate we choose—for example, drawing on past or future climate stability to compensate for the global warming we may be causing today. Nor can we stockpile climate stability for use in the future. Nor does climate stability become a part of what it produces. If timber is used to produce a chair, the timber is embodied in that chair. If climate stability is used to produce a crop of grain, that grain in no way embodies climate stability. Furthermore, climate stability is not altered by the production of a crop of grain (unless perhaps the grain is grown on recently deforested land, but still it is the deforestation and not the grain that affects climate stability).

Intact ecosystems are funds that provide ecosystem services, while their structural components are stocks that provide a flow of raw materials. However, recall that stock-flow resources are used up and fund-service resources are worn out. But when ecosystems provide valuable services, this does not “wear them out.” The fact is, however, that ecosystems would “wear out” if they did not constantly capture solar energy to renew themselves. The ability of ecosystem fund-services to reproduce themselves distinguishes them in a fundamental way from manmade fund-services. Depreciating machines in a factory do not automatically reproduce new machines to replace themselves.

Examples of ecosystem services provided by a forest may help clarify the concept. Costanza et al. describe 17 different goods and services generated by ecosystems.¹⁰ Forests provide all of these to at least some degree. Of these, food and raw materials are essentially stock-flow variables. The remaining fund-service variables included are described in Table 6.1.

⁹As further evidence of the extreme uncertainty concerning ecosystem function and human impacts upon it, in 1973 physicist James Lovelock, famous for the Gaia hypothesis, to his later regret stated that fluorocarbons posed no conceivable hazard to the environment. M. E. Kowalok, Common Threads: Research Lessons from Acid Rain, Ozone Depletion, and Global Warming, *Environment* 35(6): 12–20, 35–38 (1993).

¹⁰R. Costanza et al., “The Value of the World's Ecosystem Services and Natural Capital,” *Nature*, 387:256, Table 2 (1997).

■ **Table 6.1**

EXAMPLES OF SERVICES PROVIDED BY ECOSYSTEMS

Ecosystem Service	Examples from Forests
Gas regulation	Trees store CO ₂ and growing trees create O ₂ ; forests can clean SO ₂ from the atmosphere.
Climate regulation	Greenhouse gas regulation; evapotranspiration and subsequent transport of stored heat energy to other regions by wind; evapotranspiration, cloud formation, and local rainfall; affects of shade and insulation on local humidity and temperature extremes.
Disturbance regulation	Storm protection, flood control (see water regulation), drought recovery, and other aspects of habitat response to environmental variability mainly controlled by vegetation structure.
Water regulation	Tree roots aerate soil, allowing it to absorb water during rains and release it during dry times, reducing risk and severity of both droughts and floods.
Water supply	Evapotranspiration can increase local rainfall; forests can reduce erosion and hold stream banks in place, preventing siltation of in-stream springs and increasing water flow.
Waste absorption capacity	Forests can absorb large amounts of organic waste, and filter pollutants from runoff; some plants absorb heavy metals.
Erosion control and sediment retention	Trees hold soil in place, forest canopies diminish impact of torrential rainstorms on soils, diminish wind erosion.
Soil formation	Tree roots grind rocks; decaying vegetation adds organic matter.
Nutrient cycling	Tropical forests are characterized by rapid assimilation of decayed material, allowing little time for nutrients to run off into streams and be flushed from the system.
Pollination	Forests harbor insects necessary for fertilizing wild and domestic species.
Biological control	Insect species harbored by forests prey on insect pests.
Refugia or habitat	Forests provide habitat for migratory and resident species, create conditions essential for reproduction of many of the species they contain.
Genetic resources	Forests are sources for unique biological materials and products, such as medicines, genes for resistance to plant pathogens and crop pests, ornamental species.
Recreation	Eco-tourism, hiking, biking, etc.
Cultural	Aesthetic, artistic, educational, spiritual and/or scientific values of forest ecosystems.

THINK ABOUT IT!

Of the ecosystem services in Table 6.1, which are rival and which are excludable? Which would be impossible to make excludable?

Again we emphasize that the precise relationship between the quantity and quality of an ecosystem fund and the services it provides is highly uncertain and is almost certainly characterized by nonlinearities, thresholds, and emergent properties. We can say with reasonable confidence that the larger an ecosystem fund and the better its health, the more services it is likely to generate. As we deplete or degrade a complex ecosystem fund, we really cannot predict what will happen with any reasonable probability. Since we have defined service as an anthropocentric concept, we do know that it can be dramatically affected by human presence and use and not just by abuse. For example, a highly degraded forest in an urban setting may offer more water regulation and more recreational and cultural services (as measured by benefits to humans) than a pristine forest remote from human populations. Forests near orchards or other insect-pollinated crops may offer far more valuable pollination services.

Perhaps even more critical for the economic problem of efficient allocation of ecosystem services is their spatial variation. To use an example already described, large tropical forests can regulate climate at the local level, the regional level, and the global level. Flood control and water purification provided by forests may only benefit select populations bordering local rivers and floodplains, and the provision of habitat for migratory birds may primarily benefit populations along the migratory pathways.

Ecosystem services have some other characteristics that make them extremely important economically. Probably most important, it is unlikely that we can develop substitutes for most of these services, including their providing suitable habitat for humans. We scarcely understand how these services are generated, and we are not aware of all of them. At the cost of some \$200 million, a billionaire named Edward Bass initiated the Biosphere Two project in Arizona to see if he could develop substitutes for these services sufficient to sustain only eight people. The project failed. Imagine creating substitutes for billions of people! In addition, most ecosystem services are nonrival—if I benefit from a forest's role in reducing floods, providing habitat for pollinators, or regulating atmospheric gases, it does not diminish the quantity or quality of those services available to anyone else. Many ecosystem services (though certainly not all) are nonexcludable by their very nature as well.

The Relationship Between Natural Capital Stocks and Funds

In review, the structural elements of an ecosystem are stocks of biotic and abiotic resources (minerals, water, trees, other plants and animals), which

when combined together generate ecosystem functions, or services. The use of a biological stock at a nonsustainable level in general also depletes a corresponding fund and the services it provides. Hence, when we harvest trees from a forest, we are not merely changing the capacity of the forest to create more trees, but are also changing the capacity of the forest to create ecosystem services, many of which are vital to our survival. The same is true for fish we harvest from the ocean, except we know even less about the ecosystem services produced by healthy oceanic ecosystems.

The relationship between natural capital stock-flow and fund-service resources illustrates one of the most important concepts in ecological economics: It is impossible to create something from nothing; all economic production requires a flow of natural resources generated by a stock of natural capital. This flow comes from structural components of ecosystems, and the biotic stocks are also funds that produce ecosystem services. Therefore, an excessive rate of flow extracted from a stock affects not only the stock and its ability to provide a flow in the future, but also the fund to which the stock contributes and the services that fund provides. Even abiotic stocks (i.e., elements and fossil fuels) can only be extracted and consumed at some cost to the ecosystem. In other words, production requires inputs of ecosystem structure. Ecosystem structure generates ecosystem function, which in turn provides services. All economic production thus has an impact on ecosystem services, and because this impact is unavoidable, it is completely internal to the economic process.

■ WASTE ABSORPTION CAPACITY

But this is only half the story. The laws of thermodynamics ensure that raw materials once used by the economic system do not disappear but instead return to the ecosystem as high-entropy waste. They also ensure that the process of producing useful (ordered) products also produces a more than compensating amount of disorder, or waste. Much of this waste can be assimilated by the ecosystem. Indeed, waste assimilation and recycling are ecosystem services on which all life ultimately depends. However, as a fund-service, waste absorption occurs only at a fixed rate, while conversion of stock-flow resources into waste occurs at a rate we can choose. Waste absorption capacity is a sink for which we have control over the flow from the faucet, but not over the size of the drain. The removal of ecosystem structure also affects the ability of the ecosystem to process waste. If we discharge waste beyond the ecosystem's capacity to absorb it, we can reduce the rate at which an ecosystem can absorb waste, which makes the waste accumulate more quickly. In time, the waste buildup will affect other ecosystem functions, though we cannot always predict which services will be affected and when.

A specific example can help illustrate these points. When we first begin to dump wastes, such as raw sewage and agricultural runoff, into a pristine lake, they will be heavily diluted and cause little harm. Higher waste loads may threaten humans who use the lake with intermittent health problems from bacteria and noxious chemicals contaminating the sewage, and water becomes unsuitable for drinking without prior treatment. Increasing nutrients allow bacterial and algal populations to thrive, increasing the ability of the system to process waste, but reducing a number of other ecosystem services. Fish will begin to accumulate noxious compounds present in the waste stream and become inedible. Pollution-sensitive species will be extirpated. Yet more waste may make the water unsuitable for drinking even after extensive processing, and eventually it will become too contaminated for industrial use. Excess nutrients eventually lead to eutrophication, where algal and bacterial growth absorbs so much oxygen during the night¹¹ and during the decay process that fish, amphibians, and most invertebrate species die out. Birds and terrestrial animals that depend on the lake for water and food will suffer. With even greater waste flows, even algae may fail to thrive, and we have surpassed the waste absorption capacity of the system. Waste begins to accumulate, further decreasing the ability of algae to survive, and leading to a more rapid accumulation of waste even if the waste flow is not increased any more. The system collapses.

Just prior to the point where waste flows exceed the waste absorption capacity, a reduction in flows will allow the system to recuperate. After that point, it may not. Similar dynamics apply to other ecosystems. If the ecosystem in question provides critical life-support functions, either locally or globally, the costs of exceeding the waste absorption capacity of an ecosystem are basically infinite, at least from the perspective of those humans it sustains.

In general, ecosystems have a greater ability to process waste products from biological resources, and a much more limited capacity to absorb manmade chemicals created from mineral resources. This is because ecosystems evolved over billions of years in the presence of biological wastes. In contrast, products such as halogenated cyclic organic compounds and plutonium (two of the most pernicious and persistent pollutants known) are novel substances with which the ecosystem has had no evolutionary experience, and therefore has not adapted.

In contrast to many ecosystem services, waste absorption capacity is rival. If I dump pollution into a river, it reduces the capacity of the river

¹¹While growing plants are net producers of oxygen and absorbers of CO₂, they also require oxygen for survival. During the day, photosynthesis generates more oxygen than the plants consume, but at night they consume oxygen without producing any. Average oxygen levels may be higher, but the lowest levels determine the ability of fish and other species to survive.

to assimilate the waste you dump in. It is also fairly simple to establish institutions that make waste absorption excludable, and many such institutions exist.

The bottom line is that the laws of thermodynamics tell us that natural resources are economic throughputs. We must pay close attention to where they come from, and where they go.

Table 6.2 summarizes some of the important characteristics of the three biotic resources. We will discuss these characteristics and examine their policy relevance in greater detail in Chapter 12 and Part VI.

The points to take away from this chapter deserve reiteration. First, humans, like all animals, depend for survival on the ability of plants to capture solar energy in two ways: directly as a source of energy, and indirectly through the life-support functions generated by the global ecosystem, which itself is driven by the net primary productivity of plants. There are no substitutes for these life-support functions. Second, every act of economic production requires natural resource inputs. Not only are these inputs being used faster than they can replenish themselves; when these structural elements of ecosystems are removed, they diminish ecosystem function. Third, every act of economic production generates waste. Waste has a direct impact on human well-being and further diminishes ecosystem function. While the removal of mineral resources may have little direct impact on ecosystem function, the waste stream from their extraction and use is highly damaging to ecosystems and human well-being in the long run. As the economy expands, it depletes nonrenewable resources, displaces healthy ecosystems and the benefits they provide, and degrades remaining ecosystems with waste outflows.

Biotic resources are unique because they are simultaneously stocks and funds, and their ability to renew themselves is a fund-service. This means

■ **Table 6.2**

ECONOMIC CHARACTERISTICS OF BIOTIC RESOURCES

Biotic Resource	Stock-Flow or Fund-Service	Can Be Made Excludable	Rival	Rival Between Generations	Substitutability
Renewable Resources	Stock-flow	Yes	Yes	Depends on rate of use	High at margin, ultimately nonsubstitutable
Ecosystem Services	Fund-service	No	For most, no	No	Low at margin, nonsubstitutable
Waste Absorption Capacity	Fund-service	Yes	Yes	Depends on rate of use	Moderate at margin, nonsubstitutable

that ultimately economic scale is determined by the amount of fund-services provided in a given year, where one of those fund-services is the ability of renewable natural resources to renew. Biotic resources have a particularly large impact on scale because they ultimately have no substitutes, and we cannot survive without them.

BIG IDEAS to remember

- | | |
|---|-----------------------------|
| ■ Ecosystem structure | ■ Carrying capacity |
| ■ Ecosystem function | ■ Minimum viable population |
| ■ Ecosystem services | ■ Critical depensation |
| ■ Stock-flow and fund-service resources | ■ Maximum sustainable yield |
| ■ Risk, uncertainty, ignorance | ■ Waste absorption capacity |
-



From Empty World to Full World

Given the undeniable importance of entropy to the economic process, and the resulting fact that “sustainable economic growth” is an oxymoron,¹ how do we explain the unwavering devotion to continuous economic growth by economists, policy makers, and the general public in the face of ecological and natural resource limits? Apparently, people believe that the economic system faces no limits to growth or that the limits are far off. The laws of thermodynamics ensure that there are limits to growth. Now we must briefly address the question of how close those limits are.

Certainly for most of human history, including the time when modern economic theory was being developed, human populations and levels of resource use were quite low. Material and energetic limits to growth appeared so far off that it seemed sensible to ignore them and concentrate on developing a system that efficiently allocated the much scarcer labor, capital, and consumer goods. But since the development of market economies and neoclassical explanations thereof, both human populations and per-capita levels of resource use have been increasing exponentially. The success of the market system reduced the relative scarcity of market goods and increased that of nonmarket goods and services provided by the

¹To reiterate, we do not believe that sustainable growth of the “psychic flux” of satisfaction (what we would call development, not growth) is an oxymoron as long as that flux is not produced by ever-increasing natural resource consumption. However, as far as we know, economic growth as measured by GDP has never occurred without increased throughput. Even when each unit of GNP requires fewer resources, the net outcome has always been greater throughput. While this need not be the case, the free market economy seems poorly suited for promoting the activities that provide improved human well-being without increasing throughput.

Box 7-1 HOW CLOSE ARE WE TO A FULL WORLD?

Exponential growth occurs when a system keeps growing at a certain rate. For example, from 1900 to 2000, the per-capita material output of the global economy grew at about 2.3% per year. As a rule of thumb, one can calculate the doubling time of a given growth rate by dividing 72 by that growth rate. This means that per-capita output doubled more than three times during the twentieth century, an increase of more than ninefold. Over the same period, the human population has increased from 1.6 billion to 6.1 billion, almost a fourfold increase. The total increase in material output has increased more than 36 times this century.^a How many more times can our material output double?

Our situation may have parallels to a well-known riddle. If the area of a petri dish covered in bacteria doubles every hour, you inoculate the dish at noon on day one and it is completely full at noon two days later (and thereafter the population crashes because it has exhausted its food source and inundated the petri dish with waste), when is the dish half full? The answer, of course, is at 11 A.M. on the final day. At 9 A.M., $\frac{1}{2}$ of the resources available for continued growth are still present. The question right now for humans is: How close is it to noon?

Humans, of course, are very different from bacteria, and the Earth is different from a petri dish. Humans can control their rate of reproduction and, to an extent, the quantity of resources they use. The Earth hosts numerous ecosystems capable of providing renewable resources and processing wastes. However, human adaptation to resource scarcity requires taking time to develop new technologies, new institutions, and new ways of thinking—perhaps a great deal of time. Essentially, the closer it is to noon, the less time we have to develop and implement the necessary changes to show that we actually are substantially different from bacteria in a petri dish. (See figure 7.1.)

^aCalculations by author from data found in Chapter 5 of J. B. DeLong, *Macroeconomics*, Burr Ridge, IL: McGraw-Hill Higher Education, 2002.

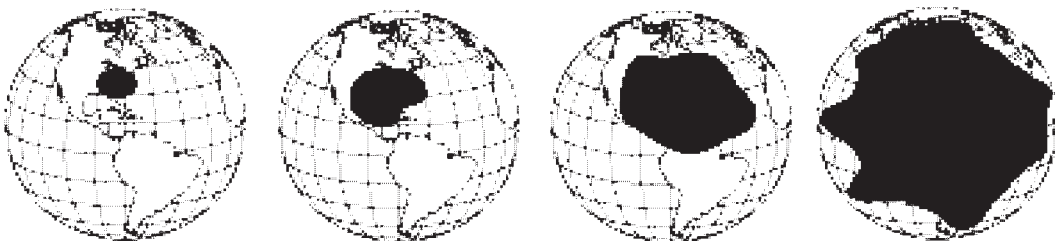


Figure 7.1 • How close are we to a full world?

sustaining system. What follows is a very quick assessment of how full the world is and how close we are to resource exhaustion.

THINK ABOUT IT!

The world is always “full” of some things and “empty” of others. In the “full world,” what is it that the world is relatively full of? Relatively empty of? How is the fullness with respect to some things related to the emptiness with respect to others?

FOSSIL FUELS

As fossil fuels run the world economy and are among the most well studied of the resources required to sustain us, we will assess their limits first. At first glance, exhaustion hardly seems imminent. Economists tell us that price is a measure of scarcity, yet the price of crude oil averaged \$18.63 per barrel between 1869 and the present (in 1996 dollars) and is scarcely any different today.² However, as we mentioned earlier, we can extract fossil fuels at virtually any rate we want, and it is the scarcity of flow that determines prices, not the scarcity of stocks (a point we return to in Chapter 11). In the regions with the vastest reserves, installed extraction capacity, more than the size of underground stocks, determines flow rates. Best estimates suggest that if we continue to extract oil at the same rate, we will exhaust probable stocks in about 40 years, yet the Energy Information Administration estimates that global demand for oil will increase by 50% over the next 20 years.³ As we said above, the net energy returns to fossil fuel exploration are declining dramatically. The same is true for new discoveries, which peaked in 1962 at 40 billion barrels per year⁴ and fell to 6 billion barrels per year in the 1990s. Consumption currently stands at 26 billion barrels per year, exceeding new discovery rates by a factor of 2 to 4.⁵ While the rate of increase in global oil consumption began to decline after 1973, the world still used over twice as much oil since 1970

²WTRG Economics, 2000. Online: <http://www.wtrg.com/prices.htm>.

³Energy Information Administration, 2002. Online: http://www.eia.doe.gov/oiaf/aeo/aeotab_21.htm. Figure of 50% increase is from 2001–2020.

⁴J. J. MacKenzie, Oil as a Finite Resource: When Is Global Production Likely to Peak? World Resources Institute, 2000. Online: http://www.wri.org/wri/climate/jm_oil_000.html.

⁵When a new oil field is discovered, it is very difficult to say exactly how much oil exists. Also, some sources report increased estimates of recoverable oil from a previously discovered source as a new discovery, while others do not; thus the discrepancy in estimates. MacKenzie (ibid.) cites a 2:1 ratio for 1996, and L. F. Ivanhoe cites a 6:1 ratio for major discoveries during the 1990s. Hubbert Center Newsletter # 2002/2, M. King Hubbert Center for Petroleum Supply Studies, Petroleum Engineering Department, Colorado School of Mines, Golden, CO, 2002.

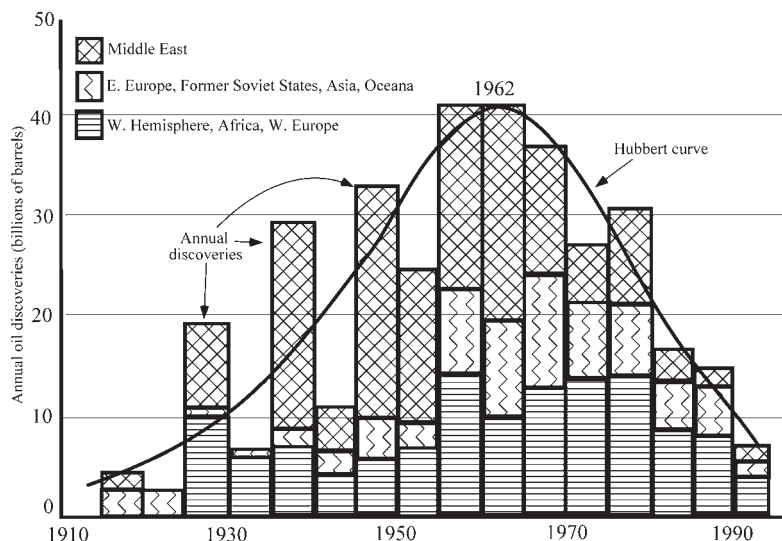


Figure 7.2 • A. Hubbert curve for oil discovery. The bars represent the average amount of crude oil discovered worldwide during each 5-year period from 1912 to 1992. The line known as the Hubbert curve is the weighted average of global oil discovered from 1915 to 1992. (Source: Adapted from L. F. Ivanhoe, King Hubbert, updated. Hubbert Center Newsletter # 97/1. Online: <http://hubbert.mines.edu/news/v97n1/mkh-new2.html>.)

than for all of human history prior to 1970.⁶ What is the net result of all this?

M. King Hubbert, while working as a petroleum geologist for Shell Oil Company, developed a theory of nonrenewable resource extraction, graphically depicted in the **Hubbert curve**. Figure 7.2 shows a Hubbert curve for oil discoveries using actual data, and Figure 7.3 shows a Hubbert curve for oil production using estimates of future production. In 1954, Hubbert used this theory to predict that oil production in the U.S. would peak in between 1967 and 1971—a prediction that was treated with considerable skepticism. In reality, it peaked in 1970. Using this same methodology, leading industry experts are now predicting that oil production will peak sometime between 2003 and 2020. Peak production follows peak discovery with a time lag, and global oil discoveries have been in fairly steady decline since their 1962 peak. As oil demand continues to grow and production peaks then declines, prices will skyrocket. While oil exhaustion may not be imminent, the end of the era of cheap oil is.⁷

⁶The area under the curve in Figure 7.3 shows total oil production, which is almost identical to consumption. You can see that the area under the curve from 1973 to the present is nearly two and a half times greater than that from 1869 to 1973.

⁷C. J. Campbell and J. H. Laherrère, The End of Cheap Oil, *Scientific American*, March 1998.

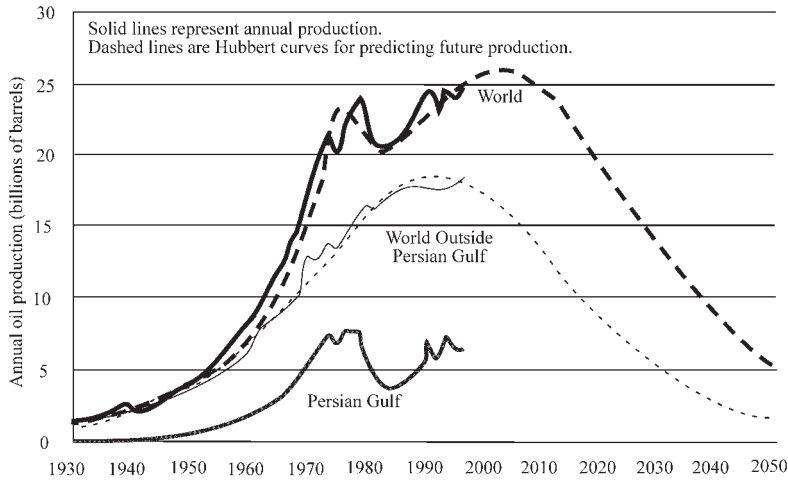


Figure 7.3 • A Hubbert curve for oil production. Global production of oil, both conventional and unconventional (solid black line), recovered after falling in 1973 and 1979. But a more permanent decline is less than 10 years away, according to the authors' model, based in part on multiple Hubbert curves (dashed lines). A crest in the oil produced outside the Persian Gulf region now appears imminent. (Source: Adapted from C. J. Campbell and J. H. Laherrère, *The End of Cheap Oil*. *Scientific American*, March 1998.)

Given the vast supplies of solar power available as a substitute, does it matter if we exhaust oil supplies? Developing solar energy as a substitute will take considerable time, and sophisticated analyses of oil prices accounting for both scarcity and information effects suggest that oil prices may rise suddenly and rapidly.⁸ Also, because solar energy strikes the Earth as a fine mist, large areas of land are needed to capture that energy in significant quantities. With current technologies and without disrupting agriculture, forestry, or the environment, the amount of solar energy that could be captured in the U.S. would meet only 20–50% of our current energy demands. And we could not get around this problem simply by using lower-wattage light bulbs. Food production and transportation in the United States currently consumes three times as much hydrocarbon energy as it provides in carbohydrate energy.⁹ One grain-fed steer

⁸D. B. Reynolds, *The Mineral Economy: How Prices and Costs Can Falsely Signal Decreasing Scarcity*, *Ecological Economics* 31(1): 155–166 (1999).

⁹D. Pimentel and M. Pimentel, *Land, Energy and Water: The Constraints Governing Ideal U.S. Population Size*, 1995. Online: http://www.npg.org/forum_series/land_energy&water.htm. Negative Population Growth, Inc. Forum Series. Georgescu-Roegen rightly pointed out that the notion of producing carbohydrates from hydrocarbons was absurd, and when this was first presented as an option, he also accurately predicted that we would move in the other direction first. N. Georgescu-Roegen, *The Entropy Law and the Economic Process*. Cambridge, MA: Harvard University Press, 1971. However, in the short-term direct conversion of oil to food might be more efficient than Western agriculture.

“consumes” some 284 gallons of petroleum in the process of becoming our dinner.¹⁰ Returning to animal traction to power our farms would require additional land devoted to fodder to feed the draft animals. With existing technologies, less fossil fuel means less food.

■ MINERAL RESOURCES

Mineral resources are also growing scarcer. As noted earlier, the richest, most available ores are used first, followed by ores of decreasing quality. We previously used hematite ore from the Mesabi range in Minnesota, which is about 60% pure iron. That ore is now exhausted, and we must use taconite ore, at about 25% pure iron.¹¹ The situation with other ores is similar. At least metal can be recycled, and other materials are adequate substitutes. If we consider topsoil as a mineral resource, the situation looks more serious. Rates of topsoil depletion in the U.S. are currently 100 times the rate of formation.¹² Globally, experts estimate that 40% of agricultural land is seriously degraded, and this number is as high as 75% in some areas.¹³ Currently, the most widely used substitute for declining soil fertility is petroleum-based fertilizers.

■ WATER

Among the more threatening of the imminent shortages is that of freshwater. While water is the quintessential renewable resource, thanks to the hydrologic cycle, global water consumption has tripled over the last 50 years, and it continues to climb. Humans are pumping rivers dry and mining water from aquifers faster than it can be replenished. While global climate change may lead to a wetter climate overall as increased evaporation leads to increased rainfall, increased evaporation will also dry out the land much more quickly. Many climatologists believe the net result will be intense downpours interspersed with severe drying. In addition, global climate change is likely to affect where water falls, leading overall to greater risks of both flooding and drought.¹⁴

The dominant use of water (70%) is agriculture, so a water shortage

¹⁰M. Pollan, “Power Steer,” *New York Times Magazine*, March 31, 2002.

¹¹J. Hanson, “Energetic Limits to Growth,” Online: <http://www.dieoff.com/page175.htm>. Also appeared in *Energy Magazine*, Spring 1999.

¹²D. Pimentel and M. Pimentel, *Land, Energy and Water: The Constraints Governing Ideal U.S. Population Size*, 1995. Online: http://www.npg.org/forum_series/land_energy&water.html.

¹³World Resources Institute, *People and Ecosystems: The Fraying Web of Life*. Washington, DC: WRI, 2000.

¹⁴C. J. Vörösmarty, P. Green, J. Salisbury, and R. B. Lammers, “Global Water Resources: Vulnerability from Climate Change and Population Growth,” *Science* 289: 284–288 (July 14, 2000).

will likely translate into hunger before thirst. The estimated water deficit (extraction of water greater than the recharge rate) in northern China is 37 billion gallons, which produces enough food to feed 110 million people.¹⁵ The Ogallala aquifer in the United States has turned the arid western plains into a breadbasket, but it is being depleted at a rapid rate. In some areas the irrigation potential has already been depleted, and tens of thousands of additional acres will lose irrigation over the next 20 years at current depletion rates.¹⁶

The use of river water for irrigation has already led to one of the planet's worst environmental catastrophes in the Aral Sea. In every continent, important aquifers are falling at rates between 2–8 m per year.¹⁷ Currently, about 1.2 billion people lack access to potable drinking water,¹⁸ less than one-third of the world's population enjoys abundant water supplies,¹⁹ and some studies suggest that nearly 50% of the world's population will be living in water shortage areas by 2025.²⁰ The World Bank warns that continued reduction in aquifers could prove catastrophic.²¹ *Fortune* magazine suggests that water shortages will make water the oil of the twenty-first century, “the precious commodity that determines the wealth of nations.”²²

Projections concerning future water supplies are highly uncertain. First, we lack adequate data.²³ Second, consumption patterns and technology can dramatically change the demand for water. Third, as mentioned, climate change can have serious impacts on the hydrologic cycle, increasing evaporation rates and changing rainfall patterns.²⁴

¹⁵L. Brown, *Water Deficits Growing in Many Countries: Water Shortages May Cause Food Shortages*. Earth Policy Institute, Eco-Economy Updates, August 6, 2002.

¹⁶M. Glantz, ed., 1989: *Forecasting by Analogy: Societal Responses to Regional Climatic Change*. Summary Report, Environmental and Societal Impacts Group NCAR, 1989. Summary based on a study by D. A. Wilhite, Center for Agricultural meteorology and Climatology, University of Nebraska, Lincoln. Online: <http://www.meteor.iastate.edu/gccourse/issues/society/ogallala/ogallala.html>; also <http://www.uswaternews.com/archives/arcsupply/7staco14.html>.

¹⁷L. Brown, *Water Deficits Growing in Many Countries*. Earth Policy Institute, Eco-Economy Updates, August 6, 2002. Online: <http://www.earth-policy.org/Updates/Update15.html>.

¹⁸U.N. Department for Policy Coordination and Sustainable Development, *Critical Trends: Global Change and Sustainable Development*. New York: United Nations, 1997.

¹⁹Vörösmarty et al., op. cit.

²⁰L. Burke, Y. Kura, K. Kassem, C. Revenga, M. Spalding, and D. McAllister. *Pilot Analysis of Global Ecosystems: Coastal Ecosystems*. Washington, D.C.: WRI, 2000.

²¹Brown, op. cit. Online: <http://www.earth-policy.org/Updates/Update15.html>.

²²N. Currier, *The Future of Water under Discussion at “21st Century Talks.”* United Nations Chronicle, on-line edition Volume XL, Number 1 2003. Online: http://www.un.org/Pubs/chronicle/2003/webArticles/013003_future_of_water.html.

²³*Science* 297(5583): 926–927 (August 9, 2002).

²⁴Vörösmarty et al., op. cit.

■ RENEWABLE RESOURCES

The fact that we live in a full world is even more obvious when it comes to “renewable” resource stocks. For virtually every renewable stock of significance, the rate of extraction is limited by resource scarcity, not by a lack of adequate infrastructure. It is a shortage of fish, not fishing boats, that has stagnated fish harvests over the last few years. The Food and Agriculture Organization (FAO) of the United Nations estimates that 11 of the world’s 15 major fishing areas and 69% of the world’s major fish species are in decline and in need of urgent management. For instance, cod catches dropped by 69% from 1968 to 1992. West Atlantic bluefin tuna stocks dropped by more than 80% between 1970 and 1993.²⁵ Similarly, it is a shortage of trees, not chainsaws, that limits wood production. As commercially valuable species are depleted, we turn to harvesting others that were formerly considered trash. As a result, for both fish and timber, the number of commercially valuable species has increased dramatically over recent decades.

Many economists cite this ability to substitute one species for another as evidence that there are no limits to potential harvests. However, when one fish species is exhausted because too many boats are going after too few fish, the whole fishing fleet is available to deplete any new stocks we may identify. Having virtually exhausted rapidly reproducing species such as cod, we now pursue species such as orange roughy, which may take as long as 30 years to reach sexual maturity. We run the risk of harvesting such species to extinction before we even acquire sufficient data to estimate their sustainable yields.²⁶

While there is serious cause for concern for the resource exhaustion of raw material inputs into the economy, these may pale in insignificance when compared to the dangers presented by the depletion and destruction of ecosystem services. Ecosystem services are destroyed directly by the harvest of their structural components, primarily the renewable resources of which they are composed, and less directly by waste emissions. Forest cover is currently being depleted in the poorer countries at the rate of about 140,000 km² per year,²⁷ and if the World Trade Organization’s

²⁵FAO of the U.N. Focus: Fisheries and Food Security, 2000. Online: <http://www.fao.org/focus/e/fisheries/challeng.htm>.

²⁶For example, one study found a 60–70% decline in total biomass of one stock of orange roughy in less than 10 years of fishing. P. M. Smith, R. I. C. C. Francis, and M. McVeigh. Loss of Genetic Diversity Due to Fishing Pressure. *Fisheries Research* 10: 309–316, 1991. Cited in M. Lee and C. Safina, The Effects of Overfishing on Marine Biodiversity. A SeaWeb Background Article originally published in *Current: The Journal of Marine Biology*, 1995. Online: <http://web.tri-countynet.net/~cgclark/marinebiodiversity.htm>.

²⁷World Resources Institute, People and Ecosystems: The Fraying Web of Life, Washington, DC: WRI, 2000.

efforts to liberalize trade in forest products goes forward as planned, the rate of deforestation is expected to increase.²⁸ The Ramsar convention on wetlands is an intergovernmental treaty providing a framework for the conservation of wetlands and their resources, yet 84% of the wetlands supposedly protected by the treaty are threatened.²⁹ While we understand marine ecosystems less than terrestrial ones, it seems unimaginable that healthy fish populations do not play a vital role in these ecosystems and the scarcely understood mechanisms by which they provide ecosystem services. For example, we know that oysters provide the valuable service of filtering entire estuaries and have seen this service drastically reduced in the Chesapeake Bay. Virtually all other ecosystems confront similar threats through depletion of their component stocks.

■ WASTE ABSORPTION CAPACITY

People have worried about resource exhaustion at least since the time of Malthus, but concern over the excessive accumulation of waste is more recent. Every economic activity produces waste. As humans overwhelm the waste absorption capacity of ecosystems at local and global levels, we suffer in two ways. First, accumulating toxins have direct negative effects on humans. Second, the toxins damage ecosystems and degrade the ecosystem services on which we depend. Accumulating evidence suggests we are overwhelming the waste absorption capacity of the planet for several classes of wastes.

The most prominent category of waste in the news today is CO₂ emissions. In spite of an impressive ability of ecosystems to absorb CO₂, there is irrefutable evidence that it is currently accumulating in the atmosphere, and almost consensus in the scientific community that this will lead to global climate change. International recognition of the seriousness of the problem has led to international discussions, but at the time of this writing, the world's worst emitter of greenhouse gases has refused to participate in international accords. Even if the United States did participate, the reductions proposed under the Kyoto protocol would fail to limit CO₂ emissions to the waste absorption capacity of the environment, and would therefore at best merely slow the rate of global warming.³⁰ In the absence of major changes in human behavior, global warming will have dramatic

²⁸P. Golman and J. Scott, et al. *Our Forests at Risk: The World Trade Organization's Threat to Forest Protection*. Earthjustice Legal Defense Fund, 1999.

²⁹M. Moser, C. Prentice, and S. Frazier. *A Global Overview of Wetland Loss and Degradation* Conference Proceedings, vol. 10. Online: http://www.ramsar.org/about_wetland_loss.htm.

³⁰R. Watson et al. (2001) *Climate Change 2001: Synthesis Report*. Summary for Policy Makers. Intergovernmental Panel on Climate Change. Online <http://www.ipcc.ch/>.

impacts on global ecosystems. This is particularly true because so many remaining ecosystems are islands in a sea of humanity, and the species they contain will be unable to leave their islands in response to changing conditions.

No less threatening than global warming are the waste emissions from mineral resources. Heavy metals are highly toxic to humans. As these metals are elements, there is no waste absorption capacity *per se*; once in the environment or in our aquifers, they remain indefinitely. These elements are normally highly diluted in nature or out of reach of living systems; humans have extracted and purified them and released them into the environment in dangerously high concentrations. Many of them tend to bioaccumulate; when ingested, they are not released, so predators retain all that has been consumed by their prey. Many fish species have dangerously high levels of mercury and other metals, which cause human birth defects and worse when consumed, not to mention their impacts on other species.

Nuclear wastes are also elements and far more toxic than the other heavy metals. Nuclear wastes do break down, but not on a human time scale. Plutonium, one of the most toxic substances known, has a half-life of 24,300 years. At minimum, we must sequester such waste for ten times that long—nearly fifty times as long as civilization has existed.

Halogenated hydrocarbons are another class of particularly dangerous manmade mineral wastes. Chlorofluorocarbons are the best known and they are now banned. However, other ozone-depleting halogenated hydrocarbons are still being produced. The ozone layer is still growing thinner, and it is not expected to recover until 2050.³¹ Ozone depletion threatens not only human health, but also global plant and animal life. The Antarctic ozone hole poses a particularly serious threat to phytoplankton production in the southern seas. In addition to its key role at the bottom of the oceanic food chain, phytoplankton may play an important role in sequestering carbon dioxide, and its depletion may contribute to global warming.³²

Other halogenated hydrocarbons are classified as persistent organic pollutants (POPs). International negotiators are currently calling for a ban on the most notoriously harmful POPs. These chemicals are now found in every ecosystem on Earth. Among their other negative traits, some of them seem to mimic hormones and are capable of affecting the reproduc-

³¹U.S. Environmental Protection Agency, Questions and Answers on Ozone Depletion. Online: http://www.epa.gov/docs/ozone/science/q_a.html.

³²R. C. Smith, B. B. Prezelin, K. S. Baker, R. R. Bidigare, N. P. Boucher, T. Coley, D. Karentz, S. MacIntyre, H. A. Matlick, D. Menzies, M. Ondrusek, Z. Wan, and K. J. Waters. Ozone Depletion: Ultraviolet Radiation and Phytoplankton Biology in Antarctic Waters, *Science* 255: 952–959 (1992).

tive capacity of many species. As their name implies, POPs will continue to persist in the environment for many years to come, in spite of the ban. In the meantime, industry is busy introducing new chemicals, many with a very similar structure to the most toxic ones, at the rate of over 1000 per year. We frequently do not become aware of the negative impacts of these chemicals for years or even decades. And while it may be possible to perform careful studies about the damage caused by a single chemical, outside of the laboratory, ecosystems and humans will be exposed to these chemicals in conjunction with thousands of others.³³

Pollution in some areas is becoming so severe that it threatens human health, ecosystem function, and even large-scale climate patterns. For example, a recent study has shown that a 3-km-thick layer of pollution over South Asia is reducing the amount of solar energy striking the Earth's surface by as much as 15% in the region, yet preventing heat from the energy that does pass through from leaving. In addition to threatening hundreds of thousands of premature deaths, the pollution cloud is likely to increase monsoon flooding in some areas, while reducing precipitation by as much as 40% in others.³⁴

In summary, it would appear that the global sink is becoming full more rapidly than the global sources of natural resources are being emptied. This is understandable in view of the fact that sinks are frequently freely available for anyone to use, that is, they are rival and non-excludable. In contrast, sources are more often rival and excludable resources that are under the discipline of property, either privately or publicly owned and managed.

A rapid assessment of the resources on which the human economy depends suggests that we are now in a full world, where continued physical expansion of the economy threatens to impose unacceptable costs. Whereas historically people have been most worried about resource depletion, the source problem, it appears that the most binding constraint on economic growth may be the waste absorption capacity of the environment, the sink.

BIG IDEAS to remember

- | | |
|----------------------|---------------------------------------|
| ■ Exponential growth | ■ Source and sink limits |
| ■ Doubling time | ■ Measures of “fullness” of the world |
| ■ Hubbert curve | |

³³A. P. McGinn, Why Poison Ourselves? A Precautionary Approach to Synthetic Chemicals. World Watch Paper 153. Washington, D.C.: World Watch, 2000.

³⁴P. Bagia, Brown Haze Looms Over South Asia, *Science* 13 (2002).



Conclusions to Part II

In Part II we have examined the scarce resources upon which all economic production depends, applying insights from physics and ecology. The economic system, like all other known systems, is subject to the laws of thermodynamics. As it is impossible to create something from nothing, and equally impossible to create nothing from something, then economic production must deplete natural resources and generate waste. On a finite planet, economic growth (and human population growth) must eventually come to an end. The real question is whether this must happen soon, or in the far distant future. The previous summary suggests that we are indeed approaching a “full world,” and that ecological economists’ concern with scale is abundantly justified. While the ultimate scarce resource is low-entropy matter-energy, the different forms of low entropy have fundamentally different characteristics. In Part III, we will see how these characteristics affect the allocation process.

Part III will involve a shift in focus from the whole to the part, from the natural earth ecosystem to the human-dominated subsystem, from concepts of physical and biological science to concepts of social science, especially economics. But even as we focus on supply and demand, prices, national income, interest rates, and trade, we will not forget that all of these economic activities take place in an increasingly full world, and that the whole system is governed more by thermodynamics and photosynthesis than by prices and GNP.

PART III

Microeconomics





The Basic Market Equation

The purpose of microeconomics is to understand how decentralized decisions by thousands of independent firms and households are communicated, coordinated, and made consistent through prices determined by supply and demand in markets. In the absence of central planning, markets make it possible to have spontaneous order. This seems a bit miraculous, but an analogy will clarify. Even though no one designed a language, nevertheless a language is logically structured and ordered. Although the capacity for language is no doubt part of our genetic inheritance, an actual language is an ordered structure that evolved spontaneously through use. We can, after the fact, analyze the logical structure and grammar of a language, but that is description, not design. Even Esperanto, which is a designed language, was largely copied from Spanish, a natural spontaneously evolved language. Few people speak Esperanto. Like language, the market is also a communications system, though much more limited in what it can communicate. But the market, too, has a “grammar,” a logic that can be analyzed even though it arose in a spontaneous, unplanned way. And the market does more than communicate—it allocates.

What are the grammar rules of markets? What is being communicated? In what sense is the market allocation efficient, and why? Is the outcome fair, or sustainable? These are microeconomic questions. Our approach here is to develop the basic answers, the big ideas and conclusions, without getting lost in details, and without resorting to calculus or indeed to any mathematics beyond ratios and simple equations. To do this we need five definitions and three principles. On the basis of these definitions and principles, we will derive a “basic market equation” and then interpret its meaning concretely. So a little patience will be required for several pages before we see just how all the concepts and principles cohere into the main conclusion of price theory. Once we have the main conclusion and

big picture, we'll show how basic supply and demand derive from that picture and will present the most basic "grammar rules" that govern supply and demand in individual markets. Then we will look behind supply and demand, first at the production function that underlies supply, then at the utility function that underlies demand. Finally, we will see the implications of ecological economic analysis with respect to the production and utility functions.

■ COMPONENTS OF THE EQUATION

Let's begin with the five definitions of concepts we'll need:

1. MU_{xn} = marginal utility of good x to consumer n . The marginal utility is the extra satisfaction one gets from consuming one more unit of the good, other things being equal. If x is pizza and you are the consumer, MU_{xn} is the amount of utility you get from consuming another piece of pizza.
2. P_x = the market price of good x (goods are x, y). What's the market price of a slice of pizza, \$2.50?
3. P_a = the market price of factor a (factors are a, b). What factors of production go into making a pizza? The kitchen and stove are the capital (fund-service), the cook is the labor (fund-service), and the flour, tomatoes, and cheese are raw material stock-flows yielded by natural capital, with the assistance of cultivation. The cook and oven transform the raw ingredients into pizza.
4. MPP_{ax} = the marginal physical product of factor a when used to make good x . The marginal physical product is the extra output produced as a result of using one more unit of a factor as an input, all other inputs remaining constant. For example, if adding one more cook increases the pizza output of a pizzeria by 20 pizzas per night, then the marginal physical product of 8 hours of labor in terms of pizza is 20 pies.
5. Competitive market. A **competitive market** is a market in which there are many small buyers and sellers of an identical product. "Many" means "enough that no single buyer or seller is sufficiently large to affect the market price." Put another way: Everyone is a price-taker; no one is a price-maker. Everyone adjusts his or her plans to prices; no one has the power to adjust prices to plans. Since everyone treats price as a parameter (a given condition) rather than a variable (something one can change), this condition is sometimes called the parametric function of prices.¹

¹See O. Lange, "On the Economic Theory of Socialism," *Review of Economic Studies*, for an excellent exposition of markets prices and parametric function.

Here are the three principles:

Law of diminishing marginal utility. As one consumes successive units of a good, the additional satisfaction decreases, that is, total satisfaction increases, but at a decreasing rate. The marginal utility of one's first slice of pizza on an empty stomach is great. The marginal utility of the fifth slice is much less. How do we know that this principle is generally true? One way to make an argument is to assume the contrary and show that it leads to absurdity. Suppose that there were a law of increasing or even constant marginal utility—the utility of the fifth slice of pizza were equal or greater than the first. Then what would we observe? A consumer would first purchase the good that gave him the highest marginal utility per dollar. But then the second unit of that good would, under constant or increasing marginal utility, give him the same or even more satisfaction per dollar than the first, and so on. With a law of constant or increasing marginal utility, consumers would spend all their income on only one good. Since the contrary of diminishing marginal utility leads to absurdity, we have indirectly established the reasonableness of diminishing marginal utility.

Law of diminishing marginal physical product. As a producer adds successive units of a variable factor to a production process, other factors constant, the extra output per unit of the variable factor diminishes with each addition, that is, total output increases at a decreasing rate. This is sometimes called the law of diminishing returns. Again we can convince ourselves of its reasonableness by assuming the contrary and showing that it leads to absurdity. Assume a law of increasing marginal physical product. We have a 10-acre wheat farm. We add one more laborer, and his marginal product is greater than that of the previous laborer. So we add another, and so on. The result is that all agricultural labor will be employed on a single farm. Indeed, we could grow the whole world's wheat crop in a single flowerpot. This is absurd, and thus we have indirectly established the reasonableness of the law of diminishing marginal physical product.²

The **equimarginal principle of maximization** has been referred to

²The law of diminishing marginal product should not be confused with economies or diseconomies of scale. An economy of scale occurs when a 1% increase in *all* the factors of production together leads to more than a 1% increase in output. This does not contradict the law of diminishing marginal physical product. With an economy of scale, we couldn't necessarily produce the world's wheat supply in one flowerpot, but we would want to grow it all on one very large farm. In reality, economies are likely to occur over a limited range of production, usually followed by diseconomies of scale. Over very limited ranges of production, an additional unit of a factor of production may have a higher marginal physical output than the first. For example, four carpenters building a house may finish a house more than four times faster than a single carpenter, as a single person simply cannot lift a large wall or maneuver a truss, but 16 carpenters on the same house are unlikely to get it built four times faster than four.

earlier as the “when to stop” rule.³ When does a consumer stop reallocating her income among different goods? When she has found an allocation that maximizes her total satisfaction or total utility. That point occurs when the marginal utility per dollar spent on each good is equal. Again, suppose the contrary—the marginal utility of good *a* per dollar spent was greater than that for good *b*. Then our consumer could increase her total utility by reallocating a dollar from *b* to *a*. Only when utilities were equal at the margin would it no longer be possible to increase total utility by reallocation of expenditure. Furthermore, the law of diminishing marginal utility guarantees that each reallocated dollar brings us closer to the optimum—buying more *a* reduces the marginal utility of *a*, buying less *b* raises the marginal utility of *b*, moving us toward equality at the margin.⁴

In a simple economy of shoes and pizzas, how would you spend your money? You buy pizza if a dollar spent on pizza provides more pleasure than a dollar spent on shoes, and you buy shoes if a dollar spent on shoes provides more pleasure than a dollar spent on pizza. To maximize pleasure, the last dollar spent on pizza must supply the same pleasure as the last dollar spent on shoes.

Similar logic applies to the producer who is maximizing her output by choosing a combination of factors such that the marginal product per dollar spent on each factor is equal. If the MPP of each factor were not equal, the same total output could be produced with fewer inputs, hence at lower total cost, thus yielding higher profit.

Think in terms of a pizza parlor. The owner can hire another cook for \$1600/month who can produce an additional 20 pizzas per day, generating \$1700 in additional net monthly revenue. Alternatively, if the owner spends \$1600/month on payments for a better kitchen, she would only be able to produce 18 more pizzas per day. The owner will keep hiring cooks as long as the additional pizzas they produce generate net profits, and

³This basic rule in economics does have an important limitation. The rule says that the way to get to the top of the mountain is to take any step that leads upward. When you can no longer do that, when any step you take will move you downward, you know you are at the top of the hill, the maximum point. Or do you? If there is a temporary dip on the hillside, you might mistake a local maximum for the global maximum. The laws of diminishing marginal utility and diminishing returns are thought to guarantee hillsides with no dips, but the caveat is important to keep in mind.

⁴The “when to stop” rule assumes people are only concerned with maximizing their own utility and do not allocate resources for the sole purpose of making others happy. It also assumes people always make the utility-maximizing choice. Mainstream economic theory assumes that rational self-interest guides all allocation decisions, and someone who acts in this way is known as *Homo economicus*. Empirical studies and common observations show that in reality, people are not always “rational,” as economists define the term, and sometimes act selflessly (helping others with no gain to themselves), vengefully (harming others even when it harms themselves as well), or in other “irrational” ways. Though the individual might maximize his or her own utility by harming or helping someone, applying the equi-marginal principle of maximization to such actions will not lead to optimal outcomes for society.

those profits are greater than the profits from spending an equivalent sum on the kitchen. However, as she hires more cooks, the cooks get in each other's way, there are insufficient ovens, and their marginal productivity goes down. In contrast, with more cooks available, the marginal productivity of a better kitchen might increase. If the productivity of a bigger kitchen per dollar invested becomes greater than the productivity of another cook (and produces enough extra pizza so their sales will cover the costs of improvement), the owner will invest in the kitchen. The point is that to maximize profits in the pizzeria, the last dollar spent on hiring cooks should generate the same profit as the last dollar spent on expanding the kitchen.⁵

Now let's imagine a shoe store moves in next door to the pizzeria. Business grows, and the owner estimates that an assistant would generate an additional \$1800 in net monthly revenue. The town is small, and the only available workers are employed by the pizza parlor. Because the shoe store owner profits more from an extra worker than the pizzeria, he can afford to pay more. One of the pizza cooks comes to work for him (assuming the skills are transferable). As the shoe store owner hires away more pizza cooks, his store gets crowded and production per laborer goes down, while the pizzeria gets less crowded and production per laborer goes up. The pizzeria owner can therefore raise wages to retain her employees. As long as the shoe store owner can make more profit from another laborer than the pizzeria owner, he can hire laborers away from her, but the decreasing productivity of more workers for him and the increasing productivity of fewer workers for her means that this cannot go on forever. Eventually, the marginal product valued in dollars of profit from a pizza cook and a shoe store assistant become equal, and neither store owner can outbid the other in hiring extra labor. Hence, the marginal physical product of labor for shoes and pizzas, as valued in dollars, will be equal. The same holds true for all factors of production across all industries.

Now we are in a position to state the basic market equation, then show why it must hold, and then interpret just what it means. The basic market equation can be written as:

$$MU_{xn}/MU_{yn} = P_x/P_y = MPP_{ay}/MPP_{ax}$$

First, notice the central position of relative prices. On the left are conditions of relative desirability, reflecting the upper or ends part of the ends-means spectrum (see Figure 3.1). On the right are conditions of

⁵Of course, there are real differences between hiring cooks and improving a kitchen. Cooks can be hired or fired, can work longer or shorter hours, and in general may allow greater flexibility than kitchen construction. Improving a kitchen is a lumpier investment, harder to do in discrete units. The arguments presented here would make more sense if the restaurant owner could invest in cooks and kitchens in very small units.

relative possibility, reflecting the lower or means part of the ends-means spectrum. The intermediary role of prices is to bring about a balance between ends and means, an efficient allocation of means in the service of ends.

But how do we know that the basic market equation holds? We know the left-hand equality holds because it is just a restatement of the consumer's allocation rule of equal marginal utility per dollar, usually written as:

$$MU_{xn}/P_x = MU_{yn}/P_y$$

The marginal utility per dollar spent on pizza should equal the marginal utility per dollar spent on shoes for consumer n . If you think about it and review the section on the equimarginal principle of maximization, you'll see that if the first ratio is larger than the second, n will increase her total utility by buying more pizza and fewer shoes.

Similarly, the right-hand equality must hold because it is the producer's equimarginal principle of maximization—all factors are employed in quantities such that the price of each factor equals the value of its marginal product. For all firms using a to produce x , we have

$$P_a = P_x (MPP_{ax})$$

The price of labor equals the price of pizza times the number of pizzas an additional unit of labor can produce with all other factors of production held equal.

THINK ABOUT IT!

In real life, can more labor produce more pizza without additional pizza ingredients? We will return to this question.

If the marginal unit of labor costs more than the value of the pizzas it can produce, the pizzeria owner would earn higher profits by employing less labor.

Likewise, for all firms using factor a (e.g., labor) to produce good y (e.g., shoes) we have

$$P_a = P_y (MPP_{ay})$$

Since P_a is the same for all firms, and things equal to the same thing are equal to each other, it follows that

$$P_x (MPP_{ax}) = P_y (MPP_{ay})$$

The value of the additional pizzas a worker could produce in a pizzeria will equal the value of the additional shoes a worker can produce in a shoe store.

Reorganizing terms, we get

$$P_x/P_y = MPP_{ay}/MPP_{ax}$$

This is the right-hand equality in the basic market equation.

■ WHAT DOES THE MARKET EQUATION MEAN?

Now that we have derived the basic market equation, what does it mean? So what?

First, note that x and y are any pair of goods, a is any factor of production used by any firm, and n is any individual. The equation holds for all pairs of goods, all factors, all firms, and all individuals. We could string out marginal utility ratios to the left, one for each individual in the economy, not just n . For each individual, the ratio of his marginal utilities between x and y would equal the price ratio. Does this mean that all individuals consume the same amounts of x and y ? Certainly not! People have different tastes, and in order to get equality at the margin, different consumers have to consume different total amounts of x and y . Unless each consumer is consuming amounts such that the ratio of marginal utilities is equal to the price ratio, that consumer is not maximizing his utility.

Likewise, we could string out marginal physical product ratios to the right, one for each factor of production (a , b , c , etc.) used by any firm in making x and y . Does the equality of each marginal ratio imply that all firms use the same total amounts of factor a or b in producing x and y ? No, because different firms have different production processes. But unless the ratio of MPP equals the price ratio, the firm in question is not maximizing profits.

We could write a similar basic market equation for every other pair of goods—one for x and z , one for y and z , and so on. So the equation holds for all relative prices.

The central role of P_x/P_y is worth emphasizing. It brings about an equality of the marginal utility ratios with the marginal productivity ratios. Things equal to the same thing are equal to each other—prices serve as a kind of sliding fulcrum on a seesaw that balances the weight of relative possibility with the weight of relative desirability, of means with ends (Figure 8.1). The rate at which consumers are *willing* to substitute one good for another (psychological rate of substitution) is equal to the rate at which they are *able* to substitute goods by exchange (market rate of substitution), and also equal to the rate at which producers are able to produce one good rather than another (essentially “transform” one good into another) by reallocating resources between them (technical rate of substitution or transformation).

Relative prices serve as a sliding fulcrum to bring about balance or equality between the utility and productivity ratios. But once that equality is achieved, what does it mean? To understand better, leave out the

The basic market equation is:

$$MU_{xn}/MU_{yn} = P_x/P_y = MPP_{ay}/MPP_{ax}$$

MU is the marginal utility of good x or good y to person n , and MPP is the marginal physical product of factor a used to produce good x or good y .

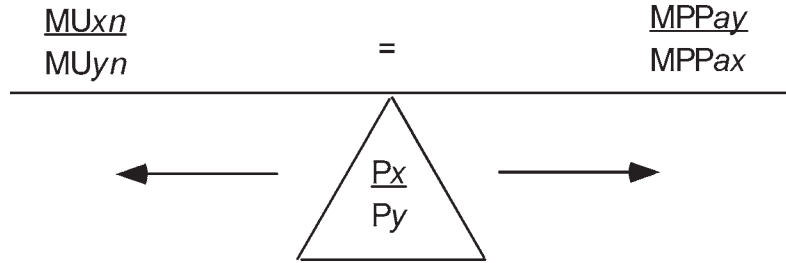


Figure 8.1 • The parametric or fulcrum function of relative prices.

intermediate price ratio and consider only the resulting equality of the marginal utility ratio and the marginal productivity ratio:

$$MU_{xn}/MU_{yn} = MPP_{ay}/MPP_{ax}$$

Rewrite this equation as:

$$MU_{xn} \times MPP_{ax} = MU_{yn} \times MPP_{ay}$$

This states that the marginal utility derived from factor a when allocated to good x is just equal to the marginal utility derived from factor a when allocated to the production of good y , as judged by consumer n ,⁶ or in more concrete terms, the amount of pizza a worker produces in an hour provides the same utility to consumer n as the amount of shoes the worker could produce in an hour (assuming the same wage). Because n is any consumer, x and y are any goods, and a any factor, it follows that no consumer would want to reallocate any factor between any pair of goods. In other words, the **basic market equation** defines an optimal allocation of resources, one in which no one would want to reallocate any factor to any alternative use because doing so would only decrease that person's total satisfaction. No firm would want to reallocate any factor to any other use because doing so would lower profit.

Perhaps this seems a very big rabbit to pull out of a very small hat! We will come back to that, but for now, let's appreciate the result. Prices in competitive markets lead to an efficient allocation of resources in the sense that no one can be made better off in his own judgment by reallocating resources to produce a different mix of goods. Of course, individual n could be made better off if income or wealth were redistributed from individual m to himself. We all could be better off if we had a larger resource endowment. But this analysis assumes a given distribution of income and wealth among people, and a given total resource endowment. The optimal

⁶A review of the units: MU_{xn} is utility per unit of good x , or U/x . MPP_{ax} is units of good x per unit of factor a , or x/a . The units of the product are $U/x \times x/a = U/a$. The units are utility per unit of factor a when it is used to make x (utility as experienced by individual n). In other words the utility yielded by factor a in its x -use is equal to that yielded in its y -use.

allocation of resources is what economists call a **Pareto optimum**: Everyone is as well off as they can be without making someone else worse off.

These qualifications shrink the rabbit back down to the dimensions of the hat, but it is still a nice trick to bring about a balanced adjustment of relative possibilities with relative desirabilities, to communicate and mutually adjust means to ends in an efficient way without central coordination. The technical possibilities of transforming one good into another by reallocating resources are balanced with the psychic desirabilities of such transformations as judged by individuals.

The key thing about this result is that it is attained by an unplanned, decentralized process. The problem solved by the price system is in the words of F. A. Hayek, “the utilization of knowledge not given to anyone in its totality.”⁷ The psychological rates of substitution, the terms on which consumers are willing to substitute commodities, are known to the different individual consumers only. The technical terms on which producers are able to substitute or transform commodities are known only to various production engineers and managers. Yet all this piecemeal, scattered knowledge is sounded out and communicated, and used by the price system in the allocation of resources. No single mind or agency has to have all of this information, yet it all gets used.

A *Pareto optimum* occurs when no other allocation could make at least one person better off without making anyone else worse off. This is also known as a Pareto efficient allocation (see Chapter 1), Pareto efficiency, or simply efficiency.

■ MONOPOLY AND THE BASIC MARKET EQUATION

The parametric or fulcrum function of prices depends on pure competition. It fails if there is a monopoly. Suppose the producer of good x is a monopolist while y is still produced in a competitive market. The equimarginal rule of maximization tells both monopolist and competitive firms to produce up to where **marginal cost**—the additional expenditures required to produce one more unit—equals **marginal revenue**—the additional income from selling one more unit.⁸ For the competitive firm, **marginal revenue** is equal to price (price is constant and the extra revenue from selling one more unit of x is P_x). But the monopolist is the only supplier and is definitely not too small to influence price by the amount he can produce. When the monopolist supplies more, it causes the price to fall. But the monopolist’s marginal revenue is not equal to the price times the extra unit. Instead it is equal to the new lower price times the extra unit *minus* the fall in price times all previously sold units. For the

⁷F. Hayek, “The Use of Knowledge in Society,” *The American Economic Review* 35(4), 520 (1945).

⁸The pizzeria owner will keep hiring cooks as long as their marginal cost (their wage, plus the cost of the additional ingredients they use) is less than the marginal revenue they generate (the price of the pizzas they produce).

monopolist, marginal revenue is less than price, that is, $MR_x < P_x$. If we substitute MR_x for P_x in the basic market equation on the right-hand side, we have:

$$MU_{xn}/MU_{yn} = P_x/P_y > MR_x/P_y = MPPay/MPPax$$

Therefore,

$$MU_{xn}/MU_{yn} > MPPay/MPPax$$

And further

$$MU_{xn} \times MPPax > MU_{yn} \times MPPay.$$

This means that the marginal utility yielded by factor a in its x -use is greater than that yielded in its y -use. Consumer n would like to see some of factor a reallocated from y to x . But this is not profitable for the monopolist. The monopolist finds it profitable to restrict supply below what consumers would most desire. He does this to avoid losing too much revenue on previously sold units of x as a result of lowering the price a bit in order to sell another unit of x . The fulcrum is split, the balance between ends and means is broken, the invisible hand fails.

Neoclassical economics deserves a hearty round of applause for the interesting and important demonstration that competitive markets result in an optimal allocation of resources. If we do not rise to our feet in a standing ovation, it is only because conventional economists sometimes forget the assumptions and limitations of the analysis that led to the conclusion. A short list that we will address later includes the limiting assumptions that the analysis is independent of the distribution of income and wealth, that all goods are market goods (i.e., rival and excludable), that factors of production are substitutes for each other, that external costs and benefits are negligible, that information is perfect,⁹ and that all markets are competitive.

■ NON-PRICE ADJUSTMENTS

We need to consider two more results from the basic market equation: First, what it tells us about making adjustment by means other than price, as well as price adjustment; and second, how it relates to supply and demand, the most basic rules of market grammar.

In the seesaw diagram (Figure 8.1), the desirability conditions (MU ratios) can be altered only by substitution, by reallocation of consumers' expenditure (not by a fundamental change in preferences); likewise, the

⁹Perfect information requires that buyers and sellers can acquire information about a product at negligible cost, and that one side in a negotiation does not have more information than the other (i.e., information is cheap and symmetrical).

possibility conditions (ratios of MPP) are alterable only by substituting factors (not by a fundamental change in technology). Prices, the sliding fulcrum, coordinate the substitutions and reallocations necessary to attain balance between the first and last terms of the equation. But what really defines the optimum is the equality of the first and last terms. Prices play only an adjusting and accommodating role.

Suppose that relative prices were fixed. Could we ever attain balance? Suppose the fulcrum position on a seesaw were fixed. How would we attain balance? By directly adjusting the weights on both sides. We could directly change the conditions of relative desirability by altering peoples' preferences through advertising. We could also directly alter the conditions of relative possibility by technological innovation. Vast sums of money are spent on advertising and on technological research. These efforts may be thought of as **non-price adjustments**. But regardless of whether adjustments to equality are by price or non-price mechanisms, the resulting equality of the end terms defines a Pareto optimum.

There are many Pareto optima, one for each distribution of income, set of technologies, and set of wants or preferences. However, if wants are created and preferences altered through advertising, and if advertising is a cost of production of the product, then production begins to look like a treadmill. If we produce the need along with the product to satisfy it, then we are not really making any forward motion toward the satisfaction of pre-existing needs. The producer replaces the consumer as the sovereign. Then, the moral earnestness of production, as well as the concept of Pareto optimal allocation of resources in the service of such production, suffers a loss.

Indeed, even under price adjustment we have the parametric function of prices—each individual takes prices as given and adjusts his plans to prices, rather than adjusting prices to his convenience. Yet the market price does change as a result of the market supply and demand conditions that result from each individual treating price as given. Yet, if no individual can change the price, then how do prices ever change in the real world? Someone, somewhere has to be a price-maker rather than a pure price-taker if prices are ever to change. This puzzle has been met in two ways by economic theorists. One is to assume an auctioneer who takes bids and changes the price. This is fine for auction markets, but most markets are not auctions. The other solution is to say that markets really cannot be 100% competitive in the sense of total compliance with the parametric function of prices, or they would never be able to adjust prices. Someone has to have a bit of market power if prices are ever to change. So some real resources have to be dedicated to price adjustment, whether it be the salary of an auctioneer or the temporary monopoly profits of a price leader. But the point to remember is that there is

non-price market adjustment as well as adjustment by prices, and the basic market equation helps to analyze all three types of adjustment: price adjustment, psychological adjustment, and technological adjustment. All will get us to Pareto optima, albeit different ones. There are many, many different Pareto optima. While it is good to know that the market will get us to some Pareto optimum, it is vital to remember that that is not enough. Some Pareto optima are heavenly, others are hellish. There is more to welfare than efficient allocation—there are distribution and scale, for example.

■ SUPPLY AND DEMAND

How do supply and demand relate to the basic market equation? This is important because supply and demand are the most useful tools of market analysis. Let's begin with demand. Take the left-hand side of the basic market equation, rewritten as:

$$M_{uxn}/P_x = MU_{yn}/P_y$$

Let commodity y be money. Then MU_{yn} becomes M_{umn} the marginal utility of money, and P_y becomes P_m the price of money, which is unity—that is, the price of a dollar is another dollar. Then we have:

$$M_{uxn}/P_x = M_{umn}/P_m$$

or

$$MU_{xn}/P_x = M_{umn}$$

or

$$P_x = MU_{xn}/M_{umn}.$$

This is the condition for the consumer to be on his demand curve. To be on the demand curve is to be maximizing utility by substituting good x for money to the point where utility is a maximum—where the marginal utility of a dollar is just equal to the marginal utility of a dollar's worth of good x .

From the relation $P_x = MU_{xn}/M_{umn}$, and the law of diminishing marginal utility, we can see that the quantity of x demanded by the consumer will be inversely related to price.

If the consumer is always maximizing utility according to the relation $P_x = MU_{xn}/M_{umn}$, and if diminishing marginal utility rules, then we can see that the relation between P_x and the amount of x demanded (Q_x) will be inverse, as depicted in Figure 8.2. Suppose the left-hand side of the equation, P_x , falls. Then to reestablish equality the right-hand side will have to fall. The consumer makes the right-hand side fall by buying more x . More x means the numerator, M_{uxn} , will decline thanks to the law of

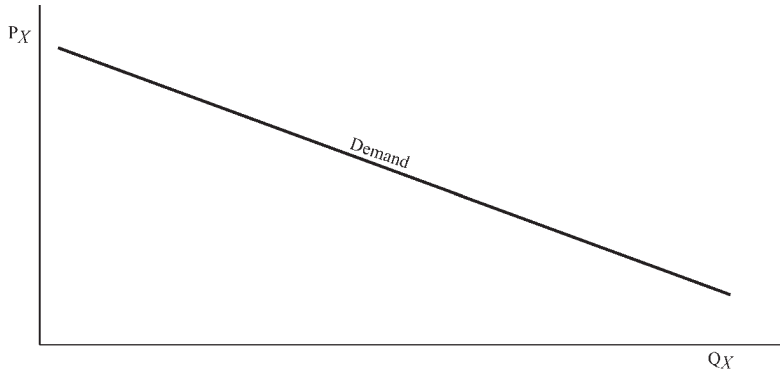


Figure 8.2 • The demand curve.

diminishing marginal utility. That reduces the ratio. Also, as the consumer buys more x she has less money, so the marginal utility of money rises, increasing the denominator and again reducing the right-hand ratio.

To go from the demand curve of the individual consumer to the demand curve of the whole market, we just add up all the individual demand curves; that is, for each price we add up all the q 's demanded at that price by each consumer. Thus the market demand curve will be downward sloping, just like that for each individual. Only the units on the horizontal axis have changed, now Q instead of q .

Turning to supply, we know that at all points on the producer's supply curve he must be offering an amount at each price that would maximize profits. That would be when $P_x = MC_x$, where MC_x is the marginal cost of producing x .¹⁰ By definition, MC_x is the cost of producing an additional unit of x . We can produce an additional unit of x by using an additional amount of factor a or b or some other factor. The marginal cost of producing x by using more factor a is P_a/MPP_{ax} —that is, the dollars spent to get one more unit of a (which is P_a) divided by the extra x that was produced by the extra unit of a gives the extra cost of a unit of x , or the marginal cost of x . We could do the same in terms of factor b , and so on. Whichever turns out to be the cheapest way to produce another unit of x (using more a or more b) is the marginal cost of x .

THINK ABOUT IT!

As you might guess, the marginal cost of x in terms of each factor would tend to equality. Can you explain why?

¹⁰Remember, we're assuming here that the marginal cost of producing x is increasing. As long as the marginal cost of production is less than the price, then it pays to produce more. Therefore, the producer stops producing when the marginal cost of production equals the price.

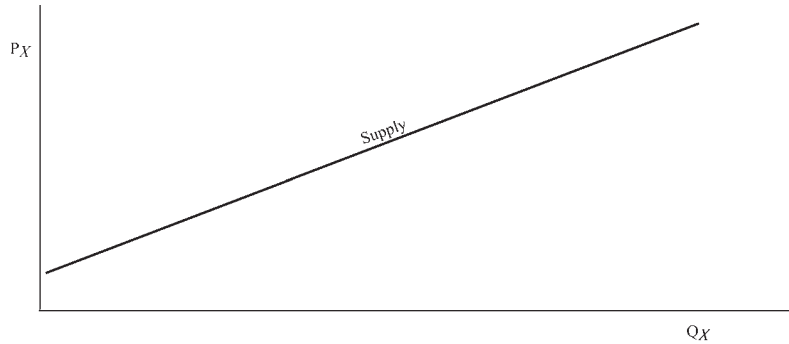


Figure 8.3 • The supply curve.

So for the profit maximizing producer to be on her supply curve the condition that must hold is

$$P_x = MC_x = P_a/MPP_{ax} \dots = P_b/MPP_{bx} \dots$$

Let's consider only $P_x = P_a/MPP_{ax}$ as the condition. From it we see that the relation between P_x and Q_x supplied must be positive, as shown in Figure 8.3. Suppose P_x rises; then P_a/MPP_{ax} will also have to rise to maintain equality. To bring this about, the producer makes more of x . By the law of diminishing marginal physical product, MPP_{ax} declines. Since it is the denominator of the right-hand ratio, that ratio rises to restore equality with the increased P_x .

Once again we move from the individual's supply curve to that of the market simply by adding up the amounts supplied by each individual at each P_x . The upward slope of the curve remains, and we simply have larger units on the horizontal axis, Q instead of q .

If we put the supply and demand curves together, their intersection will give the combination of P_x and Q_x such that buyers and sellers are both happy (Figure 8.4). Sellers are maximizing profits and buyers are

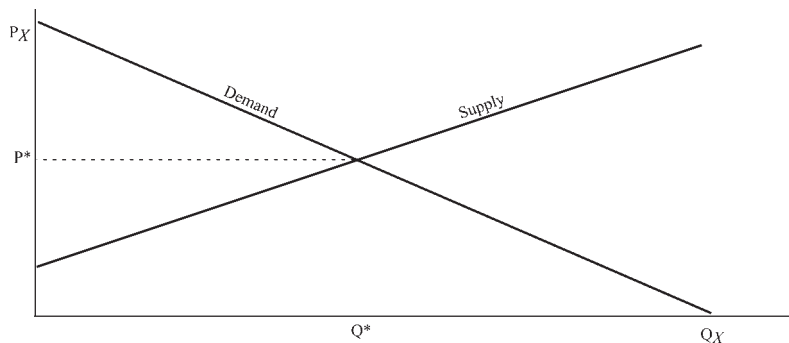


Figure 8.4 • Supply and demand determine P^* and Q^* .

maximizing utility. It is as if the market were solving two simultaneous equations in two unknowns by a process of trial-and-error approximation.

At the equilibrium point we have the following:

$$MU_{xn}/MU_{mn} = P_x = P_a/MPP_{ax}$$

The left-hand equality assures that the buyer is on his demand curve. The right-hand equality guarantees that the seller is on his supply curve. The point of intersection satisfies both buyer and seller, so the market equilibrium is at P^* and Q^* .

We could do exactly the same analysis with the market for good y , and we would end up with similar supply and demand curves, and a similar equation:

$$MU_{yn}/MU_{mn} = P_y = P_a/MPP_{ay}$$

If we divide each term in the x -market equation by each term in the y -market equation, we again arrive at the basic market equation previously derived. The terms for the marginal utility of money and the price of factor a cancel out, and we have again the basic market equation:

$$Mux_n/Mu_{yn} = P_x/P_y = MPP_{ay}/MPP_{ax}$$

Finally, it is worth commenting on the fact that the basic market equation states that the ratio of marginal utilities of good x and y are *inversely* proportional to the ratio of marginal physical products of factor a when used to make x or y . Why this inversion? What does it mean in terms of economics? Why not a direct proportionality instead of an inverse one? Actually there is a direct proportionality between the ratio of marginal utilities of x and y and the marginal costs of x and y . Prices, in effect, measure both the ratio of marginal utilities and the ratio of marginal costs. But we have to realize that the true marginal cost of x , the *opportunity cost* of a unit of x , is in fact the amount of y that has to be sacrificed to get the extra x . Thus MPP_{ay} is the amount of y given up when we use a to produce x instead of y . The amount of y sacrificed (MPP_{ay}) is, in real terms, the marginal opportunity cost of x . The best alternative sacrificed (in this case the only alternative) is by definition the opportunity cost. Therefore, the direct proportionality between prices and marginal opportunity costs is at the same time an inverse proportionality between prices and marginal physical products. The true marginal cost of x is *precisely* the amount of y you have to sacrifice to get it.

To summarize: We have derived a basic market equation; shown why it defines and how it brings about an optimal allocation of resources in the sense of Pareto (everyone as well off as they can be without making someone else worse off, i.e. without redistributing income or wealth); and shown how supply and demand derive from the equation. Since x and y

represent any pair of goods, a and b any pair of factors, and n any consumer, our conclusions hold for all pairs of goods, all pairs of factors, and all consumers. In other words, we get a good insight into the meaning of general equilibrium, yet without having to confront all the complexities of a general equilibrium model.

We have also repeatedly applied the equimarginal principle of maximization in determining how far consumers should substitute one good for another in their shopping basket, and how far producers should substitute one factor for another in their production processes. We have taken it for granted that goods can be substitutes in the minds of consumers, and that factors can be substitutes in the production processes of firms. This is sometimes referred to as the **principle of substitution**. Goods (and factors) are not always related as substitutes. Sometimes they are complements, meaning that more of one makes the other more desirable (useful) rather than less. These relations of substitution and complementarity will play an important role in later chapters.

BIG IDEAS to remember

-
- | | |
|--|--------------------------------------|
| ■ Basic market equation | ■ Pareto optimal allocation |
| ■ Competitive market | ■ Sliding fulcrum function of prices |
| ■ Law of diminishing marginal utility | ■ Monopoly and misallocation |
| ■ Law of diminishing marginal product | ■ Marginal cost and marginal revenue |
| ■ Equimarginal principle of maximization | ■ Non-price adjustments |
| | ■ Supply and demand |
-



Supply and Demand

Since supply and demand are the most important rules of market grammar, it is worth saying more about them—how they are used as tools of analysis, and something about the theories of production and consumption that underlie them. Supply and demand are so important that some wags have said you could turn a parrot into an economist just by training it to repeatedly squawk “supply and demand”—no need to worry about grammar. However, as you will see in this chapter, there is a good deal more to supply and demand than even a genius parrot could master. But fortunately, it is not too difficult for human beings.

■ A SHIFT IN THE CURVE VERSUS MOVEMENT ALONG THE CURVE

The amount demanded of a good can change for many reasons, and economists classify these causes into two categories: a change in the price of the good, and everything else. The effect of a change in price on the quantity demanded is shown by a movement along the demand curve. The effect of all other causes of a change in amount demanded is shown by a shift in the entire curve. What are these other causes? The most important are the consumer’s income, his tastes, and the prices of related goods.

If the consumer’s income rises, he will likely purchase more of every good including x , at every price. His demand curve for x shifts to the right. For a fall in income, it would shift to the left.

Goods are related as substitutes (ham and bacon) or as complements (bacon and eggs). If the price of bacon goes up, people will substitute ham and buy more ham at every price. The demand curve for ham will shift up. If the price of eggs goes up, the demand curves for both ham and

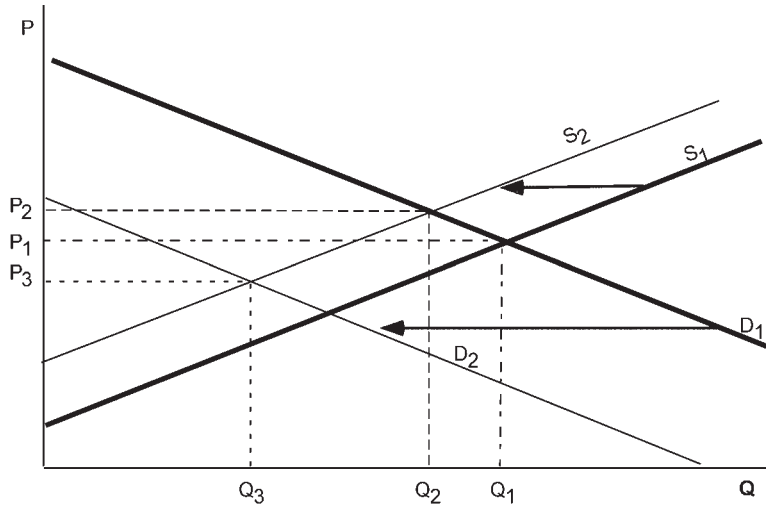


Figure 9.1 • Shifting demand and supply curves.

bacon are likely to shift down as people buy fewer eggs and eat bacon and eggs less often.

A change in tastes or information can shift the demand curve. If people worry more about cholesterol, they will buy fewer eggs at all prices (a shift downward in the demand curve).

To summarize: The demand curve is a relationship between P and Q . Within that given relationship, a change in P causes a change in Q , or vice versa, along the demand curve. But the whole relationship between P and Q can also change. That is a shift in the entire curve. The curve may change its shape and position, but it will always be downward sloping.

Similar things can be said about a movement along the supply curve and a shift in supply. The discovery of a more efficient technology, or a new deposit of resources would, for example, shift the supply curve outward so that more would be offered at each price.

The impact of shifting supply and demand curves on equilibrium price and output are shown in Figure 9.1. A massive recall of *E. coli*-tainted beef and a closing of the guilty processing plant might shift the supply curve for beef from S_1 to S_2 . In response to the shift, prices would rise along the D_1 curve, from P_1 to P_2 , and supply would drop from Q_1 to Q_2 . The recall leads to a series of investigative reports on conditions within meat-packing plants, showing that crowded conditions, rapid processing, and poor inspections make bacterial infection a regular and recurring problem. In response, the demand for beef might shift from D_1 to D_2 . As a response to this shift, the amount suppliers would be willing to provide falls along the S_2 curve from Q_2 to Q_3 , and price falls from P_2 to P_3 .

A mild winter would cause the demand for natural gas to shift down.

An increase in the price of electricity would cause it to shift up. Supply and demand analysis is not just finding the intersection, but knowing where the curves are!

■ EQUILIBRIUM P AND Q, SHORTAGE AND SURPLUS

The intersection of supply and demand is important because it defines the equilibrium in which both buyers and sellers are satisfied, and, as we saw in the basic market equation, resources are optimally allocated.

In Figure 9.2, P^*Q^* is the equilibrium price at which both buyers and sellers are satisfied (maximizing utility and profits, respectively). It is called an “equilibrium” because once at it, there is no tendency to change, because everyone is satisfied.

Any movement away from equilibrium sets in motion forces pushing us back toward equilibrium (the equilibrium is stable). Suppose the price were P_1 . Then at P_1 quantity supplied would be P_1B and quantity demanded would be P_1A . The excess of quantity supplied over quantity demanded (AB) is called a *surplus*, and if there is a surplus, the market is not in equilibrium. There are many eager sellers and few eager buyers. The many unsatisfied sellers will begin to compete with each other to sell to the few buyers. How? By offering a lower price. The price will be bid downward until the surplus disappears.

At a price of P_2 we have a quantity supplied of P_2C and a quantity demanded of P_2E . The distance CE represents a *shortage*. It, too, is a disequilibrium situation. We have many willing buyers facing few willing

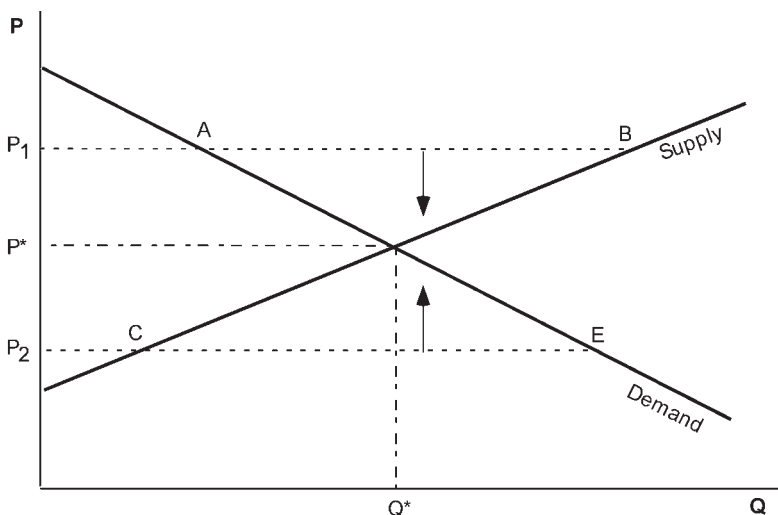


Figure 9.2 • Shortage, surplus, and equilibrium.

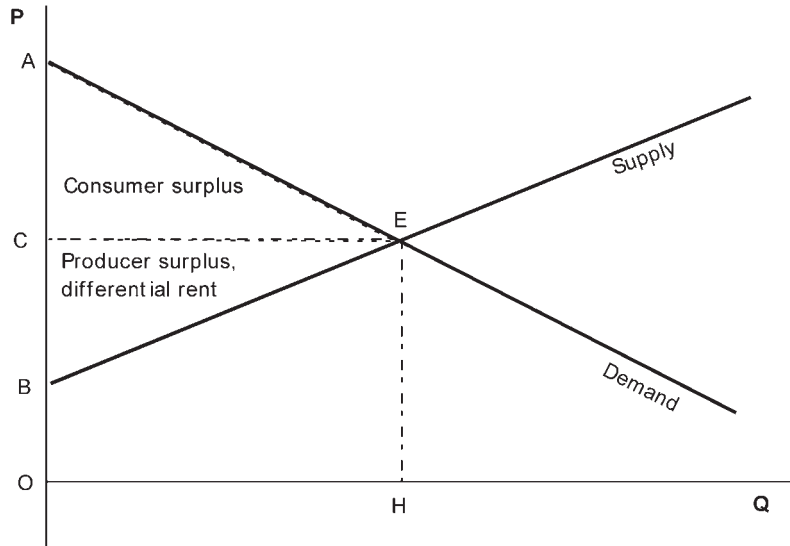


Figure 9.3 • Consumer surplus, differential rent.

sellers. Unsatisfied buyers will begin to compete with each other for the limited sellers. How? By offering a higher price. The price will rise, and as prices rise, producers will produce more, until the shortage disappears.

Note that shortage and surplus are defined with reference to a given price. When people complain about a shortage of petroleum or of labor, what they mean is a shortage of *cheap* petroleum or *cheap* labor. This has led some to claim that the free market can solve all problems. There can be no shortage of resources in a free market, nor any surplus of labor either. All we need to do is to let the price be free to find its equilibrium, and *viola*, no shortage, no surplus—and some would go on to say no resource scarcity, no involuntary unemployment! Hardly comforting if the shortage of a key resource only disappears at a near infinite price, or if the surplus of labor only disappears at a wage below the level of subsistence. The market is a wonderful institution, but not a magic charm.

In Figure 9.3, the equilibrium price is OC, and the equilibrium quantity supplied and demanded is CE. Notice that buyers pay the amount OC for *every* unit of x they buy, not just the last or marginal unit. Buyers would have been willing to pay a price of OA for the first unit of x rather than do without it. And for the second unit they would have been willing to pay almost as much, and similarly for the third, and so on. But for the last unit at CE, they were only willing to pay OC. The maximum the consumers would have been willing to pay rather than do without is AEHO. What they actually pay is OCEH. The difference, the triangular area ACE, is consumer surplus. It results from the law of diminishing marginal util-

ity. The consumer enjoys the higher marginal utility on all infra-marginal units purchased, but has to pay a price equal to the lowest marginal utility, that of the last unit purchased.

For necessities, consumer surplus is enormous. For example, I buy water at a price that equals my marginal utility of water—say, washing my car. But I get the benefits of the higher marginal utility uses on the infra-marginal units of water used to keep me from dying of thirst, from just being thirsty, from going without a bath, and so on. Many environmental goods and services are necessities that have a large or even infinite consumer surplus.

The producer is also getting a producer surplus, sometimes called **rent**, or differential rent. The producer sells all CE units at his marginal cost for making the last unit, or OC. But for the first unit he would have been willing to accept only OB rather than not sell it. And he would have been willing to accept just a bit more for the next unit, and so on. So thanks to the law of diminishing marginal physical product (law of increasing marginal costs) the producer can sell all his low-cost infra-marginal production for the same price as his highest marginal cost final unit. For example, the price of coal will be equal to the marginal cost of the most expensive coal that can be sold. That marginal coal will come from the worst, leanest, most inaccessible coal mine. But the coal from the rich easy to dig coal mines will sell at the same high price (equal to marginal cost at the worst mine), so the good mines earn a surplus or rent. This is over and above the normal profits of the operation. Normal profit is reflected in the costs underlying the supply curve. **Normal profit** is defined as the opportunity cost of the time and money the entrepreneur has put into the enterprise—that is, what he could have earned from his time and money in his next best alternative. Rents are especially important in extractive industries.

The usual definition of rent in economics is payment over and above minimum necessary supply price. The term is usually associated with land because any payment for the use of land is over and above its minimum supply price in the sense of its cost of production. The cost of production of land is zero. That doesn't mean that land is not scarce, and that no charge should be levied on its use. But it does mean that such rent is unearned income, as the tax accountants so frankly call it. It is better to tax unearned income than earned income from the point of view of both fairness and economic efficiency, as Henry George argued over a century ago. Ecological economists have followed Henry George, generalizing his insight a bit to advocate Ecological Tax Reform—shifting the tax base from value added, and on to that to which value is added, namely the throughput flow. In bumper sticker form: “Tax bads, not goods!” The bads are depletion and pollution (throughput), and the goods are value added by labor and capital, that is, earned income. More about that later.

Rent (also known as differential rent or, in the case of natural resources, scarcity rent) is equivalent to producer surplus and is defined as payment over and above the minimum necessary supply price.

■ ELASTICITY OF DEMAND AND SUPPLY

How sensitive is a change in quantity demanded to a change in price? Does it take a big change in price to get even a small change in quantity demanded? Or does even a tiny change in price cause a big change in quantity demanded? Clearly it could be either, or anywhere in between. **Elasticity** is a measure of the responsiveness of a change in quantity demanded to a change in price.

The numerical measure of elasticity is defined as follows: Price elasticity of demand = percentage change in quantity demanded, divided by the percentage change in price. Or,

$$ED = (\Delta q/q) \div (\Delta p/p)$$

An exactly analogous definition holds for the price elasticity of supply.

Figure 9.4 shows the extreme values of elasticity of demand and helps give you a feel for the concept. In the case on the left, elasticity is infinite because even the smallest percentage change in price would cause an infinite percentage change in quantity demanded. This is the way the demand curve looks to a pure competitor, a pure price-taker (i.e., one whose production is too small relative to the market supply to have any noticeable effect on price). In the case on the right, even an infinite percentage change in price would cause no change in quantity demanded, which might approximate the demand for essential goods (e.g., water, food, vital ecological fund-services) when in short supply. Most demand curves are neither horizontal nor vertical, but are negatively sloped somewhere in between. For these in-between curves, elasticity varies along the curve in most cases. Demand is said to be elastic when a 1% change in price gives rise to a more than a 1% change in quantity demanded. If a 1% change in price causes a 1% change in quantity then the formula gives an elasticity of 1, and consequently this case is called unitary elasticity.

A classic case of the importance of elasticity is the demand for agricultural crops in general. People need to eat no matter how high the price,

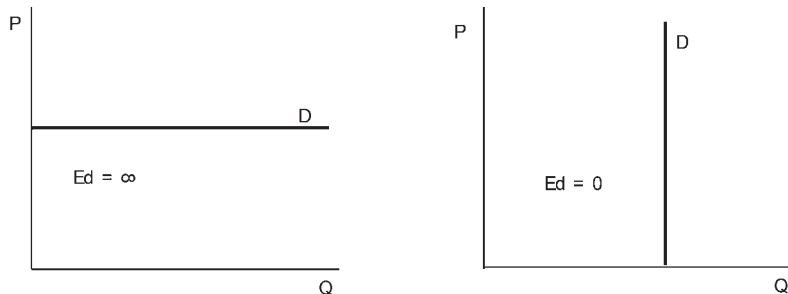


Figure 9.4 • Price elasticity of demand.

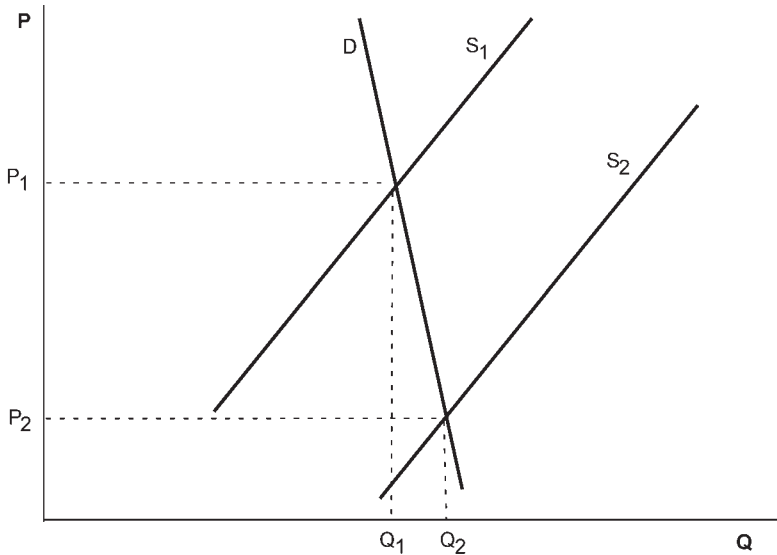


Figure 9.5 • Inelasticity of demand in agriculture. A small shift in quantity supplied leads to a large change in price.

and they will not eat much more than their fill no matter how low the price. The demand for food in general is rather inelastic, as shown in Figure 9.5. This means that the price (and total revenue) change drastically with small changes in quantity, putting the farmer in a risky position. This is one reason why governments frequently subsidize agriculture.

Figure 9.5 shows the impact of a shift in supply on the equilibrium price and quantity of a good with inelastic demand, using food as an example. At a harvest of Q_1 the price is P_1 and the total revenue to the farmers is the area of the rectangle $P_1(Q_1)$. Next year comes a bumper crop due to good weather, and the supply shifts out to S_2 with demand staying the same. The farmers' total revenue has fallen to $P_2(Q_2)$, a much smaller amount even though the harvest has increased from Q_1 to Q_2 . The reason is that with inelastic demand, a small increase in quantity caused a large decrease in price. (Elasticity works in reverse, too!) When demand is elastic, price and total revenue move in the opposite directions; when demand is inelastic, they move in the same direction.¹

THINK ABOUT IT!

What do you think would happen if the World Bank loaned money to lots of developing countries to produce and export food crops?

¹Note the implications of this for GNP (to be discussed at length in Chapter 13). If agricultural production goes down, GNP could go up as a result, and if water becomes sufficiently scarce, GNP could skyrocket!

What determines whether demand is elastic or inelastic? Mainly the necessity of the good for human well-being, and the number of good substitutes available. If ham is easily available, the demand for bacon is likely to be elastic. Since there is no good substitute for food in general, the demand for agricultural goods in general is going to be inelastic.

The longer the time period of the analysis (Q on the X-axis is really a flow, Q/t), then the more time consumers have to adjust their habits, and the more elastic demand will be.² The time period also greatly affects the elasticity of supply. In the very short run, say daily, the fish supply is totally inelastic once the fishing boats have returned with the day's catch. Weekly supply is more elastic because fishermen can respond to a higher price by staying out longer, or taking more crew. Over a year, elasticity is still greater because new fishing boats can be built. But then the supply could become totally inelastic as the limits of the fishery are reached. There is a tendency to neglect the last case in most microeconomics texts, but it is critical in ecological economics.

■ THE PRODUCTION FUNCTION

The **production function** is a relation that shows how factor inputs are converted into product outputs. It is basically a technical recipe for producing a good or service, or rather for *transforming* labor, capital, and resources into a good or service.

$$Q = F(a, b, c \dots)$$

In plain English, the quantity produced is a function of the inputs (factors of production) a , b , and c . We earlier spoke of the marginal physical product of factor a . In terms of the production function, it is the increase in Q resulting from adding one more unit of factor a to the production process, holding factors b , c , and everything else constant. The main thing we know about production functions is that they follow the law of diminishing marginal physical product, as depicted in Figure 9.6. In the graph, the marginal physical product of a is the change in Q resulting from a change in a ($\Delta Q/\Delta a$), which is also the slope of the Q curve, when Δa is one unit. In Figure 9.6, Δa is the same in both cases, but the corresponding ΔQ is much smaller when a is larger. The slope diminishes as we increase a , and it finally becomes zero. A similar curve could be drawn for factors b and c . From the basic market equation we already know the significance of the law of diminishing marginal product, as well as the basic argument for why it is true. Another way of stating the relationship is to say that for each equal increment of Q we need to employ increasing

²Of course, it can be very difficult to adjust demand for food, water, and other things essential to life.

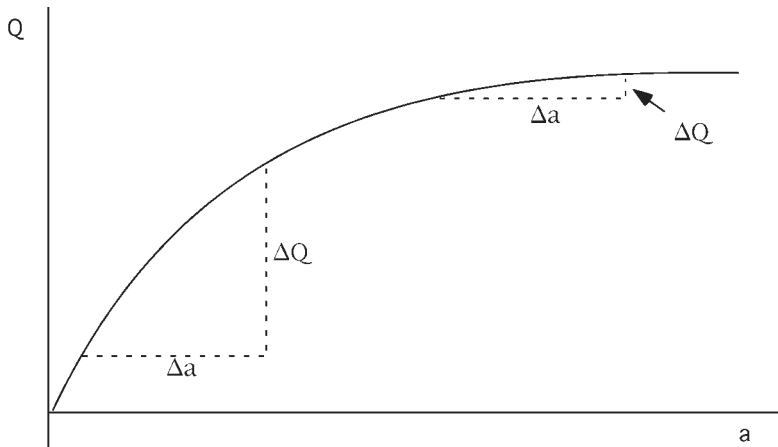


Figure 9.6 • Diminishing marginal physical product.

amounts of a . Stated this way, the relation is the law of increasing marginal cost.

The assumption behind the curve is that a is a substitute for b , c , etc.—all the other factors held constant, but that it is an imperfect substitute. Therefore, the more we try to substitute a in the recipe, the less successful the substitution becomes in terms of producing more Q . The poorer a substitute factor a is for the other factors held constant, the sooner MPP_a falls to zero. And if a is really a complement to the other factors held constant, then MPP_a is zero from the start.³ Production functions exhibit both **substitutability** and **complementarity** between factors. But standard economists tend to see mostly substitution, whereas ecological economists emphasize complementarity. Why is this so?

Probably it is because ecological economists put different things in the production function and assign different qualitative roles to the different factors in the production process. For instance, neoclassical economists treat all inputs the same—labor, capital, and resources. Ecological economists insist on a qualitative difference. Labor and capital are transforming agents, funds that transform the flow of resources into a flow of product, but which are not themselves embodied physically in the product. Labor and capital are agents of transformation (efficient causes), while resources are that which is being transformed (material causes). The neoclassical production function abstracts from the difference between material and efficient cause of production, and considers both to be equivalent. Ecological economists insist on the distinction. It is clear that while one **material cause** (resource) can often be substituted for another, and one

³Note that the relations of complementarity and substitutability apply to factors of production as well as to goods used by consumers.

efficient cause (labor, capital) can often be substituted for another, the relation of efficient cause to material cause, of agent of transformation to material undergoing transformation, is mainly one of complementarity. Ecological economists emphasize this latter relation; neoclassical economics misses it entirely.

The ecological economics production function embodies the fund-flow distinction:

$$Q = F(N, K, L; r)$$

K and L are funds of labor and capital, and r represents flows of natural resources. Funds and flows are basically complements. Substitution takes place within each category, but usually not between them. N stands for natural capital, which exists both as a stock that yields a flow of resources (a forest yielding cut timber) and as a fund (a forest yielding the service of watershed protection or wildlife habitat). The stock function of yielding a flow of resources is already captured in r, so N then should be taken here as representing the fund function of providing an indirect service that contributes to the transformation of r into Q, much as K and L provide direct services. For example, forest cover providing the service of water catchment to recharge aquifers used in irrigated agriculture is as much a capital fund-service to agriculture as is the fund of pipes and sprinklers used in irrigation.

In most neoclassical economics textbooks, the production function is written as

$$Q = F(K, L)$$

In other words, resources are neglected entirely! A flow of output is seen as a function of two funds, or two stocks that are not decumulated, or drawn down.

Is it possible for stocks (or funds) that do not decumulate, by themselves, to yield a flow? An economist's first reaction is to say of course a stock by itself can yield a flow; consider the stock of money in the bank yielding by itself a flow of interest (the principal is not decumulated yet the interest flow continues, perhaps even in perpetuity). True, but that is a convention of finance, not a physical process of production. How about a stock of cattle yielding a flow of new cattle in a sustained-yield fashion. Isn't that a physical stock yielding a physical flow? Not really. It is a stock (livestock) converting a *flow* of inputs (grass, grain) into a *flow* of outputs (new cattle, and waste products). The resource inflow of grass and grains, and so on is *transformed* into a product outflow of new cattle (and replacement to the livestock herd for natural mortality or "depreciation"), plus waste. The correct description of "production" is transformation of a resource inflow into product outflows, with stocks (funds) of capital and

labor functioning as the transforming agents. This is true not only for the living transformers in agriculture (plants and animals), but even more obviously for industrial processes of production where the transformation is visible within the factory at every stage.

Why are such basic facts excluded from neoclassical economics? Why does it choose to ignore them, to exclude natural resources from theoretical analysis right from the start? Perhaps it is a case of “money fetishism”—assuming that what is true for money in the bank, the symbol and measure of wealth, “must” be true for the wealth that it symbolizes.⁴

Recall that the production function is a recipe. Real recipes in real cookbooks always begin with a list of ingredients. They do not just say “take the labor of a cook and the capital equipment in a standard kitchen and make cherries jubilee.” Real cookbooks give us a list of ingredients, followed by instructions about how to combine and transform the ingredients into the product. The cook and her kitchen are not physically transformed into an edible dish! They are the transformers, not the transformed.

Perhaps this latter consideration has led some neoclassical economists to include r in their production function. But they have done it in such a way as not to solve the problem. Most production functions are multiplicative forms—that is, the relationship F among the factors is one of multiplication (e.g., the Cobb-Douglass production function⁵). After all, what could be more natural than “multiplying” together things that we call “factors” to produce something that we call a “product”? Unfortunately there is nothing in the real-world process of production that corresponds at all to multiplication. There is only transformation. This means that substitutability is built into these production functions from the beginning as a mathematical artifact, including substitutability between r and K , and r and L (between funds and flows). In these multiplicative production functions, we can make one factor as small as we wish, while keeping the product constant, if we increase the other factor sufficiently. The only restriction is that no factor can be reduced to zero, but it can approach zero. But according to this logic, if our cook is making a 5-pound cake, he can increase it to a 1000-pound cake with no extra ingredients—just by stirring harder and baking longer in a bigger oven! The First Law of Thermodynamics (conservation of matter and energy) has been totally ignored.

⁴One current text includes resources among its factors of production in discussion, but on the next page wrote the production function as above with only K and L as variables, and from then on forgot about resources. J. M. Perloff, *Microeconomics*, 2nd ed., Reading, MA: Addison Wesley, 2000.

⁵The Cobb-Douglass production function (in its simplest form, $Q = K^\alpha L^\beta$) states that production equals capital (raised to an exponent) times labor (raised to an exponent), and the sum of the exponents ($\alpha + \beta$) equals one. The important point for us is simply that capital and labor are multiplied, and if R (resources) is added as a third factor, it too is multiplied in like manner.

Ignoring the necessary role of natural resources in production is part of a pattern in neoclassical economics that has the effect of denying that nature has any role in economic life. Value in their view is only value added—added by labor and capital. But added to what? In the neoclassical view, that to which value is added is thought to be merely inert stuff having no independent value of its own. Ecological economics recognizes the contribution of nature in supplying that to which value can be added. It is by no means inert stuff. It is scarce low-entropy matter and energy, as we discussed in Chapter 4.

Moreover, since resource flows are complementary with manmade funds, they can become the limiting factor on production when they become more scarce than the manmade fund factors. This is exactly what is happening in the real world. Economic logic tells us to maximize the productivity of the limiting factor in the short run, and to invest in its increase in the long run. Economic logic has not changed, but the pattern of scarcity has. As we move from the empty to the full world economy, natural resource flows and services generated by natural capital stocks and funds become the limiting factor. Fish catches are no longer limited by the manmade capital of fishing boats, but by remaining natural capital of stocks of fish in the sea and the natural funds that support their existence. We need to economize on and invest in the limiting factor. Economic logic has not changed, but the identity of the limiting factor has.

Because our basic neoclassical theory of production, when it considers natural resources at all, cannot distinguish funds from flows or recognize complementarity between them, we have been slow to recognize this change.

■ THE UTILITY FUNCTION

The **utility function** relates utility or want satisfaction to the flow commodities (goods and services) consumed by the individual:

$$U = F(x, y, z, \dots)$$

In plain English, our happiness depends on what we consume (and not on our freedom, creativity, social relationships, etc.). The main thing we know about the utility function is the law of diminishing marginal utility—that as we consume increasing amounts of a single good, other things being equal, our additional satisfaction from each additional unit of the good in question declines—that is, total utility increases but at a decreasing rate.

THINK ABOUT IT!

Do you remember the argument demonstrating why it is reasonable to believe that this law is true?

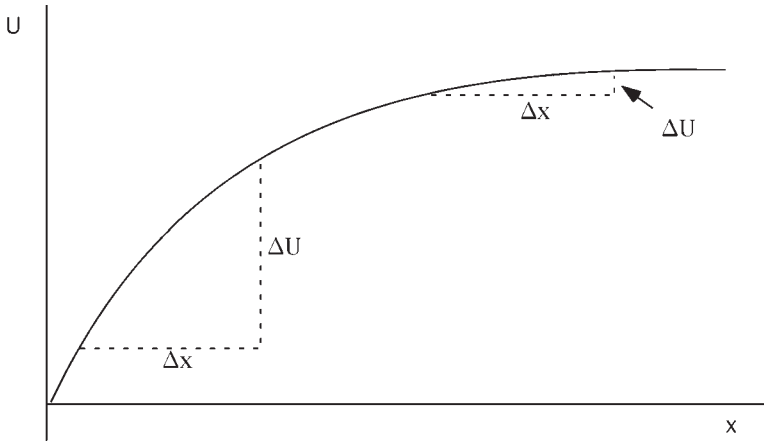


Figure 9.7 • The law of diminishing marginal utility.

The graph in Figure 9.7 simply shows that marginal utility declines with more units of x consumed in a given period, even though total utility increases (but ever more slowly). The first ice cream cone on a hot day is a delight, the third is not. To the extent that x is complementary with other goods that are being held constant, the marginal utility of x will decline faster, or may even be zero from the start. For example, a third ice cream cone without a complementary glass of water is not so enjoyable. And a right shoe without a left shoe has zero marginal utility to most people.

The neoclassical production function only contains flows of goods or services as variables. A more complete picture of consumer satisfaction would include the services of directly enjoyed natural capital funds—the provision of breathable air by a well-functioning atmosphere, drinkable water by a well-functioning hydrologic cycle, and so on. As we did previously, let's refer to these natural funds as N and include them in the utility function in a way analogous to our inclusion of the transforming funds L and K in the production function. Direct service from N is not just the pleasure of a beautiful landscape, although that is included. It also and primarily refers to the basic fitness of our environment to our organism that supports our life, and consequently supports our consumption and want satisfaction from commodities. We have evolved over millions of years into a relation of fitness to our environment. But fitness is a reciprocal relation. If we are fit for our environment, then our environment is fit for us. The relation of fitness can be destroyed by either a change in us or a change in our environment.⁶ This fitness of the environment implies

⁶Fitness-destroying changes in the environment can be induced by humans (e.g., global warming) or autonomously caused by nature (e.g., a volcanic eruption or earthquake). H. J. Henderson, *The Fitness of the Environment*, Gloucester, MA: Smith, 1913.

a relation of complementarity between N and ourselves, including most of our artifacts, which are extensions of ourselves.⁷

Let's rewrite the utility function as:

$$U = F(N; x, y, z, \dots)$$

If x is a pair of hiking boots, then its utility depends on places worth hiking in (N). If y is a snorkeling mask, its utility depends on reefs and clean water (N)—not to mention prior dependence on breathable air, drinkable water, sunlight filtered of enough of its UV rays so we won't get melanoma if we go snorkeling, and so on. N provides a complementary service without which the utilities of most consumer goods are not very great.

Consumers may be able to maintain the same level of satisfaction in the face of a reduction in x by simply consuming more y and z , which to some degree are substitutes. But N is a complement to x , y , and z , and their increase will usually not compensate for a decline in N . In fact, their utility will fall with a decline in N . For example, you won't enjoy your new hiking boots very much if there are no pleasant hiking trails.

The production of more x , y , and z may not make us any better off if accompanied by a decline in N . Indeed, it may make us worse off. The usual assumption is that N is superabundant and x , y , and z are scarce. But that was back in the empty world. We are now in a full world. The use of N as a stock that yields r is likely being pushed beyond sustainable limits. Consequently we may lose not only N in the stock sense, but also the services of N in the fund sense. And those services are complementary with most consumer goods, so reduced N will mean reduced utility from x , y , and z .

We may argue that for rich people with a low MU for goods in general, the relative importance of N is high, but that for the poor, the MU of goods still outweighs the MU of N . That may be the case in certain instances, and it is certainly what institutions like the World Bank consider realistic. But it need not be. Even food bought by the poor has low marginal utility if the poor are being choked by poisonous air and fried by UV radiation.

Microeconomics is a big subject, and this section could turn into a book by itself. We have not tried to cover the whole topic, but rather to explain the basic grammar of markets, why markets are efficient, what conditions have to hold for markets to be efficient (and by implication what might possibly go wrong), and to give the rudiments of supply-and-demand analysis. We have also tried to point out those places where we think the ecological economics perspective improves microeconomics. We

⁷Cars and bicycles are extensions of our legs, clothes are extensions of our skin, telephones extend our ears, computers extend our brains, etc.

will now focus our attention more directly on N, the goods and services provided by nature, to determine how well they meet the criteria for efficient allocation by market forces.

BIG IDEAS to remember

- | | |
|---|---|
| ■ Shift in the supply or demand curve versus movement along the curve | ■ Production function |
| ■ Shortage, surplus, equilibrium | ■ Utility function |
| ■ Consumer surplus | ■ Diminishing marginal utility |
| ■ Producer surplus, rent | ■ Diminishing marginal physical product |
| ■ Elasticity | ■ Production as transformation |
| ■ Complementarity and substitutability | ■ Fitness of the environment |
| | ■ Natural capital's role in both utility and production functions |
-



Market Failures

In the previous chapters, we saw how markets utilize individual self-interest to efficiently allocate resources (means) among alternative ends via the pricing mechanism. However, markets only function efficiently with a narrow class of goods. We have already shown how monopolies undermine the ability of the market to efficiently allocate, but monopolies are a type of market failure generated by a structural problem—the absence of competition. Markets also fail because of inherent characteristics of certain types of resources, or because there are no institutions clearly defining property rights. The purpose of this chapter is to describe the characteristics that make a particular resource a market good, and to examine what happens when resources do not have these characteristics. As we will show, none of the goods and services provided by natural capital has all of the characteristics required for efficient allocation by the market.

For simplicity, in this chapter we will boil down the various conditions for arriving at Pareto optimal allocations, described in Chapter 8, to the generic equation “marginal cost equals marginal benefit,” or $MC = MB$. In addition, we will often use the term *Pareto efficient* or just *efficient* as the equivalent of Pareto optimal.

■ CHARACTERISTICS OF MARKET GOODS

Excludability

We first defined and discussed **excludability** in Chapter 4 and briefly review the concept here in recognition of its importance. An excludable good is one for which exclusive ownership is possible; that is, a person or community must be able to use the good or service in question and prevent others from using it, if so desired. Excludability is virtually synonymous with property rights. If a good or service is not owned exclusively by

someone, it will not be efficiently allocated or produced by market forces. The reason for this is obvious. Market production and allocation are solely dedicated to profits. If a good is not excludable, someone can use it whether or not any producer of the good allows it. If people can use a good regardless of whether or not they pay for it, they are considerably less likely to pay for it. If people are unwilling to pay for a good, there will be no profit in its production, and in a market economy, no one will invest in producing it, or at least not to the extent that the marginal benefit to society of producing another unit is equal to the marginal cost of production.

Of course, many nonexcludable goods, such as fish in the ocean, are produced by nature, not by humans. In this case, “investment” is simply leaving smaller fish to grow larger, or maintaining a high-enough population stock to ensure future production. The “cost” of investment is opportunity cost—the profit that would have been earned by catching those fish today. If a fisherman throws back a small fish to let it grow larger, it is more than likely that a different fisherman will catch that larger fish, and in a market economy, people rarely invest when others will reap the returns.

Excludability is the result of institutions. In the absence of institutions that protect ownership, no good is truly excludable unless the possessor of that good has the physical ability to prevent others from using it. Some type of social contract, be it government or less formal social institutions, is required to make any good excludable for someone who lacks the resources to defend her property. Excludability, therefore, is not a property of the resource *per se*, but rather of the regime that controls access to the resource. It is fairly easy to create institutions that provide exclusive property rights to tangible goods such as food, clothing, cars, and homes. Slightly more complex institutions are required to create exclusive property rights to intangibles such as information. Patent laws protecting intellectual property rights are ubiquitous in modern society, but it remains difficult to enforce such property rights. For example, have you ever recorded copyrighted music or installed unpurchased, copyrighted software on your computer?

However, many goods and services, such as the majority of the fund-services produced by ecosystems, have physical characteristics that make it almost impossible to design institutions that would make them excludable. As we suggested in Chapter 6, it is pretty much impossible to conceive of a workable institution that could give someone exclusive ownership of the benefits of the ozone layer, climate regulation, water regulation, pollination (by wild pollinators), and many other ecosystem services. It is often possible to establish exclusive property rights to an ecosystem fund (e.g., a forest) while at the same time impossible to establish such rights to the services the fund provides (e.g., regional climate regulation). If, like a forest, the fund is simultaneously a stock that can supply a flow (e.g., of timber), market allocation will only account for the

stock-flow benefits of the resource. When there are no excludable property rights to a good or service, that good or service is nonexcludable.

■ RIVALNESS

A second characteristic that a good or service must have if it is to be efficiently produced and distributed by markets is **rivalness**. We defined a rival good or service in Chapter 4 as one for which use of a unit by one person prohibits use of the same unit at the same time by another. Rivalness may be qualitative, quantitative, or spatial in nature. Again, food, clothing, cars, and homes are rival goods.

A nonrival good or service, therefore, is one whose use by one person has an insignificant impact on the quality and quantity of the good or service available for another person to use. Among nonrival goods produced by humans, streetlights, information, and uncrowded roads come to mind (though roads do wear out faster if more people use them). Climate stability, the ozone layer, beautiful views, and sunny days are a few of the nonrival goods produced by nature.

Note that all stock-flow resources are quantitatively rival. If I eat food (a stock), there is less for you to eat. In contrast, fund-service resources may be rival or nonrival. When a fund-service is rival, it is spatially rival at each point in time, and qualitatively rival over time. If I wear clothes, drive a car, use a machine that makes cars, or use a house (all fund-service), they are not available for you to use at the same time I do, and if you use them afterwards, they are just a bit more worn out. As we pointed out in Chapter 4, all nonrival resources are fund-service.

As discussed in Chapter 9, market efficiency requires that the marginal cost to society of producing or using an additional good or service be precisely equal to the marginal benefit. However, if a good is nonrival, an additional person using the good imposes no additional cost to society. If markets allocate the good, it will be sold for a price. If someone has to pay a price to use a good, he or she will only use the good until the marginal benefit is equal to the price. A price is by definition greater than zero, while the marginal cost of additional use of nonrival goods is zero.¹ Therefore, markets will not lead to efficient allocation of nonrival goods, or conversely, a good must be rival to be efficiently allocated by the market.

Nonrival But Congestible Resources

There are actually two types of nonrival goods and services. Some nonrival services, such as UV protection by the ozone layer, are not affected by

¹This does not necessarily imply that providing a nonrival good free of charge is efficient either. We will return to this topic later.

the number of people using them. For other nonrival goods, use by too many people can seriously diminish the quality of the good or service. For example, if I drive my car down an empty road, it does not diminish your ability to drive down that same road. However, if thousands of people choose to drive down the same road at the same time, it results in traffic jams, and the ability of the road to move us from point A to point B is seriously diminished. Such goods are nonrival but congestible, and for shorthand will simply be referred to as congestible. Note that **congestibility** is an issue of scale—as scale increases, as the world becomes more full, congestion leads some nonrival goods to acquire attributes of rival goods.

The Interaction of Excludability, Rivalness, and Congestibility

What happens when goods and services are nonrival, nonexcludable, or both? The simple answer is that market forces will not provide them and/or will not efficiently allocate them. However, we need to be far more precise than this if we are to derive policies and institutions that will lead to the efficient allocation and production of nonrival and/or nonexcludable resources. Effective policies must be tailored to the specific combination of excludability, rivalness, and congestibility that characterizes a particular good or service. The possible combinations are laid out in Table 10.1, and described in some detail next.

■ **Table 10.1**

THE MARKET RELEVANCE OF EXCLUDABILITY, RIVALNESS, AND CONGESTIBILITY

	Excludable	Nonexcludable
Rival	Market goods; food, clothes, cars, houses, waste absorption capacity when pollution is regulated	Open access regimes (“tragedy of the commons”), e.g., ocean fisheries, logging of unprotected forests, air pollution, waste absorption capacity when pollution is unregulated
Nonrival	Potential market good, but if so, people consume less than they should (i.e., marginal benefits remain greater than marginal costs); e.g., information, cable TV, technology	Pure public good, e.g., lighthouses, streetlights, national defense, most ecosystem services
Nonrival but congestible	Market goods, but greatest efficiency would occur if price fluctuates according to usage; e.g., toll roads, ski resorts	Nonmarket good, but charging prices during high-use periods could increase efficiency; e.g., non-toll roads, public beaches, national parks

Adapted from A. Randall, “The Problem of Market Failure.” In R. Dorfman and N. Dorfman, eds. Economics of the Environment, 3rd ed. New York: Norton, 1993, pp. 144–161.

■ OPEN ACCESS REGIMES

The first class of goods and services we will examine are open access resources—those that are nonexcludable but rival. The use of such goods commonly leads to what Garret Hardin has called “the tragedy of the commons.”² The classic example Hardin used was the grazing commons once common in England. Say a village has a plot of land that anyone in the community can use for grazing cattle. There are 100 households in the community, and the plot of land is sufficient to support 100 head of cattle indefinitely without being overgrazed. In the terminology of Chapter 6, if we think of cattle as grass harvesters, then 100 cows will harvest the maximum sustainable yield of grass (see Figure 6.1). If one person adds one more cow to the commons (as might happen when there is no institution preventing her from doing so), not only does the grass need to be shared among more cows, but the grass yield declines, and each cow will be just a bit thinner. One person will gain the benefits of having two cows but will share the costs of all the cattle being thinner with everyone else in the community. If everyone thinks in the same manner, households will keep adding cattle to the commons until its productive capacity has declined, and it is no longer capable of generating the biomass it once did. Each person acting in what appears to be rational self-interest degrades the commons, and everyone is worse off than if he or she had stuck with one cow per person. Under these circumstances, rational self-interest does not create an invisible hand that brings about the greatest good for the greatest number, but rather creates an invisible foot that kicks the common good in the rear!

It is extremely important to note that the “tragedy of the commons” is a misnomer. Common property is property for which a community, not an individual, controls the property rights. Those who are not members of the community are not allowed access to the resource. In many cases, communities have developed institutions that prevent individuals within the community from overexploiting the resource, and there is no problem with the “tragedy of the commons.” A better term, therefore, is the “tragedy of open access regimes” or simply “the open access problem.”

There are many goods characterized by the open access problem. Hardin originally wrote his classic article to describe the problem of population growth. Especially in labor-driven agrarian economies, large families can be great assets. However, if everyone has a large family, the land must be divided up among all the children, and it eventually grows too scarce to sustain the population. People harvest soil nutrients faster than

²G. Hardin, *The Tragedy of the Commons*, *Science* 162: 1243–1248 (1968).

the system can restore them, and sustainable production declines. Another resource plagued by this tragedy is ocean fisheries, and the Atlantic cod is a classic example.

THE ATLANTIC COD FISHERY: BOX 10-1 A TRAGEDY OF THE OPEN ACCESS REGIME

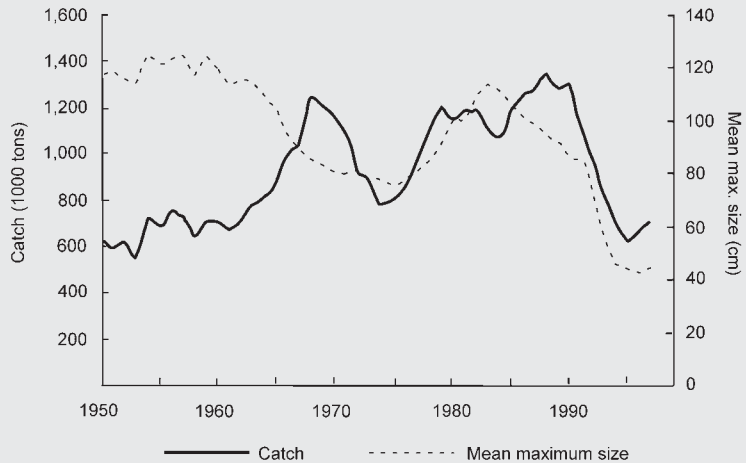


Figure 10.1 • Time series of total catch and average maximum size of species in catch for Canada, Northwest Atlantic. Note the trend to smaller species in the early 1990s. (Source: FAO Statistics. Online: http://www.fishbase.org/manual/fishbasefao_statistics00002679.htm.)

The Atlantic cod fishery was once capable of providing enormous sustained yields at a low harvest cost, but years of overharvesting have been blamed for dramatically reduced yield.^a Sustainable yields may now be close to zero,^b and scarcity has made those that remain more expensive to harvest. For cod and many other commercial species, selective fishing pressure has also drastically reduced the age and size at maturity. Smaller mature females produce far fewer and smaller eggs, threatening the ability of the species to re-establish.^c All the fishermen as a group would clearly have been better off if they had limited their harvest to sustainable levels. However, in any given year, any given fisherman was better off continuing to harvest more fish as long as a profit was to be made. Lacking institutions to keep harvests sustainable, if one fisherman reduced his catch, those cod would simply be caught by another fisherman.

In the case of renewable resources, reducing harvests in one year is passive investment in future production. The costs of investment are the opportunity costs of not harvesting now. As we pointed out in our dis-

cussion of excludability, if a good or service is nonexcludable, the market provides no incentive to invest in it. In the case of Atlantic cod, each fisherman pursuing his own rational self-interest has virtually wiped out the stock, and other fishermen are rapidly following the same path for the majority of fish species worldwide.^d Economists tell us that if we wipe out one species, we can always replace it with another. Fishermen have done this around the world, harvesting fish lower and lower on the trophic chain (i.e., lower-level predators), yet in spite of this substitution, harvests in many places are still plunging.

^aL. Burke, Y. Kura, K. Kassem, C. Revenga, M. Spalding, and D. McAllister, *Pilot Analysis of Global Ecosystems: Coastal Ecosystems*. Washington, D.C.: WRI, 2000.

^bIt is actually quite possible that cod populations in the North Atlantic have fallen below the minimum viable population stock (point of critical depensation) and will gradually diminish, even in the absence of further harvests.

^cF. Saborido-Rey and S. Junquera, *Spawning Biomass Variation in Atlantic Cod (*Gadus morhua*) in Flemish Cap in Relation to Changes in Growth and Maturation*, *Journal of Northwest Atlantic Fishery Science* 25: 83–90 (1999).

^dAs mentioned earlier, the U.N. estimates that 11 of the world's 15 major fishing areas and 69% of major commercial fish species are in decline. The only reason that total harvests have not dropped accordingly is that favored species high on the food chain have been replaced by others lower down (Burke et al., *op. cit.*).

Many economists have correctly pointed out that the open access problem results from a lack of enforceable property rights (i.e., excludability). If the English commons in the first example had been divided up into 100 equally productive excludable private lots, the rational individual would graze only one cow in each lot, and the tragedy would be avoided. Unfortunately, for many of the resources of concern to us, the ability to bestow individual property rights is more the exception than the rule, and in other cases we will describe later, property rights will not lead to efficient outcomes.³

For now, we will draw attention only to the difficulty of establishing property rights in the fairly simple case of open access resources. Analysis of oceanic fisheries provides a good starting point. Most of the oceans are international waters over which there is little or no institutional control. There are treaties limiting harvests, prohibiting certain harvest techniques or prohibiting harvests of certain species all together, but countries can choose whether or not to sign those treaties, and little in the way of enforcement is available even when they do sign. For example, most nations have agreed to cease or drastically reduce the harvest of many species of whale, but countries such as Norway, Japan, and

³Nonetheless, sloppy analysis and a lack of rigor on the part of too many economists have led to a widespread belief that establishing property rights is the answer to most, if not all, of our environmental problems.

Iceland do not always follow these regulations, and little can be done to force them to do so.⁴

Nations now enjoy 200-mile zones of exclusion in coastal waters, where they can prohibit boats from other nations from harvesting marine species and physically enforce this exclusion if necessary. Exclusion zones at least allows the potential for the regulation of fisheries within these waters, and we will discuss effective mechanisms in Chapter 20. Unfortunately, fish are generally pretty disrespectful of such boundaries, and once outside of those bounds, they are fair game to all. In addition, many species of fish migrate from the coastal exclusion zone of one nation to the exclusion zone of its neighbor. This is the case with many salmon populations off the coasts of Canada and the U.S. These two nations, which enjoy some of the best relations of any two nations in the world, are in the midst of a bitter dispute over who is entitled to what share of the catch. In the meantime, salmon populations continue their rapid decline.

If we are unable to establish defensible property rights to a resource such as fish, how are we going to address the far more difficult “tragedy of the commons” problem of overpopulation?⁵

■ EXCLUDABLE AND NONRIVAL GOODS

A second class of goods of great interest is those that are excludable but nonrival and noncongestible. The prime example of this type of good is information. In the not-too-distant past, most information was relatively nonexcludable as well as being nonrival. In Adam Smith’s time, firms would jealously guard their trade secrets, but if such secrets got out, there was nothing to prevent others from using them. As Smith⁶ pointed out, trade secrets were equivalent to monopolies, and “the monopolists by keeping the market constantly understocked, by never fully supplying the effectual demand, sell their commodities much above the natural price. . . . The price of monopoly is upon every occasion the highest which can be got” (p. 164). “Monopoly, besides, is a great enemy to good manage-

⁴The regulations actually put a moratorium on harvesting certain species of whales for commercial purposes, but still allow harvesting for scientific research. Japan now harvests endangered whale species for “scientific research,” selling the carcasses commercially afterwards. CNN.com, Japan Whaling Fleet Returns Home Amid U.S. Dispute, *Nature* (2000). Online: <http://www.cnn.com/2000/NATURE/09/21/whaling.japan.reut/> posted September 21, 2000.

⁵Kenneth Boulding actually proposed a solution to the overpopulation problem based on awarding all women the “property right” to 2.1 children (replacement fertility level) in the form of tradable permits. Needless to say, many people object to such a system. Can you suggest a better solution? See K. Boulding, *The Meaning of the Twentieth Century*, New York: Harper & Row, 1964.

⁶A. Smith, *The Wealth of Nations: Books I–III* (with an introduction by Andrew Skinner), Harmondsworth, Middlesex, England: Penguin Books, 1970.

ment . . .” (p. 251). (We will say more about patents, which are monopolies on information, in our discussion of globalization in Chapter 18.) In more recent times, of course, trade secrets have been protected by patents,⁷ an institution that makes them legally excludable, and hence marketable. The justification for this is the assumption that without excludable property rights, people would not profit from inventing new things. Inventors would have no incentives, and the rate of advance of technology would slow, to the detriment of society. Once a patent expires, the knowledge embodied in it becomes a pure public good.

The problem is that one person’s use of information not only has no negative impact on someone else’s use, it can actually lead to improvements in quality—in the words of one computer programmer, “the grass grows taller when it’s grazed on.”⁸ Intellectual progress is invariably a collective process. In academia, people have freely shared and built upon each other’s ideas for centuries. The Internet and much of its associated software were primarily the result of freely shared knowledge. In many ways, the free flow of information and ideas creates an “efficiency of the commons,” not a tragedy.

Patents, on the other hand, may slow the rate at which we develop new knowledge and use it. Existing knowledge is the most important input in the production function of new knowledge. Keeping existing knowledge artificially expensive during the life of the patent also makes the production of new knowledge more expensive. In addition, corporations often patent scientific methodologies and even mathematical algorithms, thereby making it much more expensive to conduct research using those methodologies. Many researchers are engaged in research for the sake of advancing knowledge and not for making profits, and any additional costs are likely to reduce their ability to advance knowledge. For example, a new virus-resistant strain of rice cannot be distributed because there are as many as 34 separate patent holders with competing claims on the knowledge that went into its invention.⁹

The costs of intellectual property rights have become a serious issue. For example, the U.S. Constitution authorizes Congress to issue copyrights and patents “for limited times” to “promote the progress of science

⁷Trade secrets do still exist in the traditional form. Patents only provide exclusive ownership to information for a fixed time period. To avoid making information public knowledge at the end of this time span, some companies prefer not to patent certain processes or recipes, instead keeping them hidden from potential competitors. See J. E. Stiglitz, “Knowledge as a Global Public Good.” In I. Kaul, I. Grunberg, and M. A. Stern, eds. *Global Public Goods: International Cooperation in the 21st Century*, New York: Oxford University Press, 1999.

⁸Cited in D. Bollier, *Silent Theft: The Private Plunder of our Common Wealth*, London: Routledge, 2002.

⁹Ibid.

Box 10-2**THE LINUX OPERATING SYSTEM AND
OPEN SOURCE: THE EFFICIENCY OF THE
INFORMATION COMMONS**

Dr. Ferdinando Villa, pioneer in eco-informatics, writing open-source code on a Linux operating system.

Much information is covered by patents designed to make it excludable. Open-source software, in contrast, is protected by licenses designed to keep it nonexcludable. While many open-source licenses allow people to sell open-source software, they insist that it also be legal to redistribute the software for free. Even more important, the licenses must allow distribution of the source code, that is, the computer program written in a language intelligible to humans, compared to the compiled binary code intelligible only to computers. This practice enables other programmers to find and remove bugs in the software, to modify the software, and to incorporate the software into their own work on the condition that that work remain open-source as well. The philosophy behind this approach is that when many programmers are free to improve the source code for a piece of software, it evolves and improves at an astonishing rate. The grass grows taller the more it's grazed on.

While some economists would tell us that invention requires the incentive of the profit motive, empirical evidence suggests otherwise. Take the example of the Linux operating system. Linux is an open-source operating system for computers invented by Linus Torvald. Computer experts around the world have worked on this operating system free of charge, and as a result it has become stable, powerful, and adaptable.^a IBM has contributed to the Linux code, and both IBM and HP use the

Linux operating system on their high-end mainframe servers. The use of Linux continues to increase rapidly. Apple has turned to open-source for its Mac OS-X operating system. Open-source Apache runs more than half of the world's Web servers, and hundreds of thousands of other open-source packages exist.^b Certainly this proves that neither profits nor patents are always required to spur innovation.

^a*The Great Giveaway*. New Scientist (2002). Online: <http://dsl.org/copyleft/dsl.txt>; D. Bollier, *Silent Theft: The Private Plunder of Our Common Wealth*, London: Routledge, 2002.

^bOpen Source Initiative Website (<http://www.opensource.org/>).

and the useful arts.” The initial Copyright Act awarded copyrights for 14 years, with the possibility of a single 14-year extension if the author was still alive. Under pressure from corporate lobbies, Congress has gradually increased copyright longevity, and corporate copyrights are now good for 95 years, while individual copyrights are good for the life of the individual plus 70 years. The latest extension was brought before the Supreme Court, with the argument that it actually deters the progress of science.¹⁰

Patents can also generate serious inefficiencies for other reasons. Consider the case of AIDS medicine. A currently available drug cocktail can dramatically reduce the level of HIV in the human bloodstream, potentially decreasing the risk of transmission. The benefits of controlling contagious and deadly diseases are nonexcludable. Currently drug companies hold patents on these medicines, making them prohibitively expensive for Third World countries,¹¹ decreasing their ability to control the disease, and increasing the risk of everyone contracting it. Of course, from the perspective of corporations that profit from these medications, total elimination of the disease would be a very unprofitable outcome. The argument in favor of patents is that without profits, corporations would not have the incentives to invent new drugs. The irony is that patent rights are protected in the name of the free market, yet patents simply create a type of monopoly—the antithesis of a free market.

So we see that while there may be a solid rationale for allowing patents, there also exist compelling arguments against them. If information is free, it will presumably be used until the marginal benefits of use are just equal to the marginal costs of additional use, which is zero. This is a prerequisite for efficient allocation. On the other hand, if a good is nonexcludable,

¹⁰L. Greenhouse, Justices to Review Copyright Extension, *New York Times*, February 20, 2002.

¹¹In spring 2001, a number of drug companies agreed to drop their suit against the South African government and its policy of producing and selling the drug without paying full royalties. Numerous other drug patents exist that still illustrate the basic principles explained here.

the market provides no incentive to invest in it. Patent laws recognize this problem by imposing artificial excludability on information, at least for the time period of the patent. Nonetheless, Linux (see Box 10.2) and many other examples show that patents are not necessary to spur invention, so the belief that patents will result in a faster rate of technological advance is little more than an assertion. Widespread recognition of this problem has led to the “copyleft” movement, “a copyright notice that permits unrestricted redistribution and modification, provided that all copies and derivatives retain the same permissions.”¹²

Still, goods such as information and knowledge present difficult issues. We will return to these issues later in this chapter, and again in our discussion of trade and development in Chapter 17, with the issue of so-called trade-related intellectual property rights.

What about excludable, congestible goods? As discussed earlier, congestible goods are nonrival at low levels of use and rival at high levels of use. We used roads and traffic jams as an example, and recreational resources such as beaches, swimming pools, parks, and wilderness hiking trails are similar (though for the gregarious, crowding may actually add value). When goods or resources have these properties, positive prices may produce efficient outcomes for high levels of use, while at low levels of use, pricing will lead to inefficient outcomes. This suggests that under certain circumstances, it may be reasonable to treat congestible goods as market goods during peak usage and nonmarket goods at other times.

Multi-tier pricing structures are one possible solution. **Multi-tier pricing** involves charging different prices at different times or for different users. In this case, prices could be charged when congestion occurs (e.g., rush hour tolls on a bridge), but the good or service would remain free while uncongested. Such pricing structures can be expensive to implement, and whether the strategy is reasonable generally depends on the specific case. Whether the strategy is possible depends on excludability.

■ PURE PUBLIC GOODS

As most economists readily admit, the market is not capable of optimally producing or efficiently allocating pure public goods, which are both nonrival and nonexcludable. Pure public goods are both nonrival and nonexcludable. We add the adjective “pure” only because many people are careless in their use of the term “public goods.” As we explained in Chapter 8, in a market setting, each person can purchase a good or service until the marginal benefit from purchasing one more unit of that good or service is just equal to the marginal cost. As long as anyone is willing to pay

¹²M. Stiltz, Copyleft and the Information Renaissance (2002). Online: <http://dsl.org/copyleft/>.

more for a good than it costs to produce that good, the supplier will supply an additional unit. If a public good exists, however, anyone can use it regardless of who pays for it. An additional unit of a market good is worth producing only as long as at least one individual alone is willing to pay at least the cost of producing that unit. In contrast, a public good is worth producing as long as all individuals together are willing to pay the cost of producing another unit.¹³ Look again at Figure 9.4 for supply and demand. When we moved from the individual demand curve to the market or social demand curve, we added up the quantities each individual would be willing to pay at a given price because we were talking about market goods. This is because the goods were rival, and what one person consumed another could not. However, public goods are nonrival, so one person consuming the good does not leave any less for others. In this case, we obtain the social demand curve by adding up the prices each individual is willing to pay for a given quantity to find out how much society as a whole is willing to pay for that quantity.

For market goods, each person consumes exactly as much as they purchase, so people's consumption preferences (weighted by their income, of course) are revealed by how they spend their money in the market. For public goods, in contrast, each person consumes as much as all of society purchases. This leads to problems.

For example, assume a nice forested park in the middle of a big city would cost \$100 million for land purchase, landscaping, and infrastructure. Imagine that if we added together everyone's demand curve for the park, we would find that for a park of the proposed size, society is cumulatively willing to pay \$150 million. Therefore, if everyone in the city contributed two-thirds as much money to building the park as he or she thought it would be worth, the park would be built. The problem is, how do we get everyone to contribute toward the park the amount that the park would be worth to him or her? Would market forces (i.e., the private sector) build it? Assume a corporation builds the park, fences it off, then charges admission. Knowing that the average person should be willing to pay \$150 for a lifetime pass to the park, the corporation decides to sell such passes to recoup its investment. But problems arise. Not everyone values the park equally. Some people would be willing to pay much more than \$150 for the pass, while others would be willing to pay very little. Those who are only willing to pay less than \$150 will in effect no longer value the park at all if there is a \$150 fee, and the corporation fails to recoup its investment. The corporation runs into similar problems if it charges an entrance fee for each use, say \$1. In this circumstance, even

¹³P. Samuelson, The Pure Theory of Public Expenditure, *Review of Economics and Statistics* 36: 387–389 (1954).

those who value the park the most will use the park less than they would have if it were free. Since they will use it less, they will value it less. Again, the corporation will be unable to recoup its investment, and the park will not be built. If the park were free, more people would use it, increasing the total welfare of society while imposing no additional costs on society, but then, of course, the corporation would not build the park. Therefore, the market will not provide the park even as a private good, and if it did, it would not be efficiently utilized.

Box 10-3 THE FREE-RIDER EFFECT

What would happen if some institution solicited voluntary donations to build a public park in my neighborhood? I am trying to decide how much to donate. If I meet standard neoclassical economic assumptions, I want to maximize my own utility. I live close to the proposed park site and would value it more than most people. I decide that I am indifferent between a park of the proposed size and \$1000, and prefer the park over any cost less than \$1000. However, I rationalize that if I contribute nothing to the park and others contribute what it's worth to them, that will only reduce the size of the park by one one-hundred-thousandth. I would vastly prefer a park 99.999% of the size of the proposed park at zero cost to myself than the proposed full-size park for \$1000. Alternatively, if I contribute what the park is worth to me and others contribute less, the resulting park will be smaller because of insufficient funds, and therefore no longer worth \$1000 to me.

From this narrow perspective of self-interest, my best strategy, regardless of what others choose, is to contribute nothing and instead rely on the contributions of others. Unfortunately, if everyone else also makes a similarly rational calculation in his or her own self-interest, the city ends up with no park whatsoever, and everyone is worse off than they would have been if the park had been built. This is known as the **free-rider effect**, and it is a serious obstacle to the provision of public goods. In this case, rational self-interest has created an invisible foot that kicks the common good in the rear!

Let's examine another example of the clash between markets and public goods. A small sharecropper in southern Brazil is kicked off his land share so that the landowner can grow soybeans under a heavily mechanized system requiring little labor. The soybeans are exported to Europe as cattle feed for higher profits than the landowner could make using sharecroppers to produce rice and beans for the local market. The sharecropper heads to the Amazon and colonizes a piece of land. Researchers have "guess-timated" the value of the ecosystem services sustainably pro-

duced by this land at roughly \$1660/hectare/year.¹⁴ These ecosystem services are primarily public goods. If the colonist deforests the land, he may make a one-time profit of \$100/hectare for the timber (the timber is, of course, worth much more on the market, but the market is far away, and middlemen and transportation costs eat up the profits) and an estimated \$33 annualized net profits per year from slash and burn farming.¹⁵

In terms of society, there is no doubt that the annual flow of \$1660/year far outweighs the private returns to the farmer. However, the ecosystem services are public goods that the farmer must share with the entire world, and there is no realistic way of giving the farmer or anyone else meaningful private property rights to the ecosystem services his forests supply.¹⁶ In contrast, the returns to timber and agriculture are market goods that the farmer keeps entirely for himself, and existing institutions give him the right to do as he pleases with his private property. Clearly, both the farmer and society could be better off if the beneficiaries of the public goods paid the farmer to preserve them. As long as the farmer receives more than \$150/year he is better off, and as long as global society pays less than \$1660/ha/year, it is better off.

Unfortunately, a number of serious obstacles prevent this exchange from happening, and we'll mention three. First, most people are ignorant about the value of ecosystem services (more on this later). Second, the free-rider effect means that many beneficiaries of public goods will pay little or nothing for their provision. Third, we currently lack institutions suitable for transferring resources from the beneficiaries of ecosystem services to the farmer who suffers the opportunity cost of not deforesting. Thus, from the farmer's point of view, in a market economy, deforestation is clearly the rational choice, and society suffers as a result.

Public Goods and Scarcity

Anyone who accepts the basic premise that global ecosystems create life-sustaining ecosystem services must believe that public goods are critically important. Yet market economic theory offers little advice concerning the production and allocation of public goods.

¹⁴R. Costanza, R. d'Arge, R. de Groot, S. Farber, M. Grasso, B. Hannon, S. Naeem, K. Limburg, J. Paruelo, R. V. O'Neill, R. Raskin, P. Sutton, and M. van den Belt, The Value of the World's Ecosystem Services and Natural Capital, *Nature* 387: 253–260 (1997). The land also produces a number of goods, such as timber and marketable nontimber forest products, which are valued in the cited paper, but those values are not included in this estimate.

¹⁵A. Almeida and C. Uhl, Developing a Quantitative Framework for Sustainable Resource-Use Planning the Brazilian Amazon, *World Development* 10 (1995).

¹⁶This does not mean we cannot develop mechanisms for compensating the farmer for providing ecosystem services; it simply means that if the farmer provides them, they are provided for one and all.

As we have repeatedly stressed, it is impossible to make something from nothing, and nothing from something. The production of market goods requires raw materials and generates waste. Raw materials are stock-flow resources taken from ecosystem structure, which therefore deplete ecosystem fund-services. Waste returned to ecosystems further depletes these services. Thus, if our economic system provides incentives solely for producing and allocating market goods, it will systematically undermine the production of absolutely invaluable public goods—and life-sustaining functions of our planet. One of the underlying assumptions of ecological economics is that many of the scarcest and most essential resources are public goods (services provided by natural resource funds), yet the existing economic system only addresses market goods.

Let's return now to the question of knowledge and information, presented above. If information is a private good, it will not be efficiently allocated; if it's a public good, it will not be produced in sufficient quantity by market forces. If we set theory aside momentarily and simply look at the rate of technological progress, we might believe we have little to complain about. Technological progress is extremely rapid. While it is true that patents create legal monopolies, they only do so for a limited time, after which knowledge becomes a public good. It is not difficult to believe that it is the lure of temporary monopoly profits that brings new inventions onto the market faster than would otherwise be the case. Why worry about a system that works?

One reason is that the creation of this knowledge imposes an opportunity cost on society. There is a limited pool of resources (e.g., money, scientists, laboratories) for conducting research, and if it is being used in one task, it is simply not available for another. If new inventions are driven primarily by the pursuit of profits, then we have a serious bias against the invention of public goods or technologies that preserve or restore public goods. For example, the pharmaceutical industry employs legions of scientists and spends billions on research and development for noncommunicable diseases afflicting the wealthy.

On the other hand, the control of communicable diseases is a public good, and from a societal perspective, we should channel resources toward it. An excellent example is found with the case of tuberculosis treatments. Tuberculosis is a highly contagious disease that is difficult to treat. Effective treatments were developed in the 1950s, but they require close monitoring of patients for 6 months to a year. Many people who suffer from tuberculosis are not sufficiently responsible to treat themselves and governments throughout the world have spent enormous amounts of public money to track down people and force them to take their medicine. In response to declining infection rates, federal funding in the United States targeted for tuberculosis treatment was slashed in the 1970s, and public

health expenditures suffered further cuts in the 1980s. As a result, many tuberculosis sufferers did not receive treatment, or began to take their medicine only erratically. This contributed to a resurgence of tuberculosis in the 1980s, including multiple-drug-resistant varieties. In New York City alone, it cost over \$1 billion in government spending to bring this epidemic back under control.

Tuberculosis primarily affects the poor, which reduces the profitability of any cures and explains the lack of investment in new treatments by drug companies.¹⁷ (It is no coincidence that only 13 of the 1240 new drugs licensed between 1975 and 1996 dealt with lethal communicable diseases that primarily afflict people from developing countries.¹⁸) However, even if drug companies did develop new treatments, they would need to patent the medicine and sell it for a profit to recoup their investments. Patents increase the prices of medicines to cure contagious diseases, while from the perspective of society their cost to patients should actually be *negative*. In other words, it would be efficient for the government to pay people to use such medicines because their use provides positive benefits to the rest of society.

Most research scientists working today are employed by the private sector, which retains rights to whatever they produce. The private sector is increasingly responsible for funding research in universities as well. It will logically concentrate on research with market potential. Corporate scientists would presumably work for a public organization for the same salaries. In this case, the resulting knowledge could be free for all to use, a prerequisite for efficient use (as defined by neoclassical economics) of nonrival goods. We are not suggesting here that all research be government funded.¹⁹ But unless some nonmarket institutions fund research into public goods, technological advance will tend to ignore nonmarket goods.

As the great Swiss economist Sismondi argued long ago, not all new knowledge is a benefit to humanity. We need a social and ethical filter to select out the beneficial knowledge. Motivating the search for knowledge by the purpose of benefiting humanity, rather than by securing monopoly profit, provides a better filter—a filter more likely to give us a cure for

¹⁷L. Geiter, “Ending Neglect: The Elimination of Tuberculosis in the United States,” Committee on the Elimination of Tuberculosis in the United States, Division of Health Promotion and Disease Prevention, 2000. Online: <http://www.nap.edu/books/0309070287/html/>.

¹⁸Garret, Laurie (2000). *Betrayal of Trust: the Collapse of Public Health*, New York: Hyperion.

¹⁹It is worth noting that the government currently does fund enormous amounts of primary research with taxpayer dollars, yet subsequently allows private corporations to establish patents to products derived from that research. This allows corporations to earn monopoly profits from taxpayers on research paid for by those very taxpayers.

AIDS or tuberculosis or malaria, than a new liposuction or heart transplant technique.

If the market is extremely effective at producing market goods but very poor at producing or preserving public goods, then over time, public goods inevitably become more scarce relative to private goods, giving rise to a problem of what we call **macro-allocation**, which is the allocation of resources between market and nonmarket goods and services.

Public Goods and Substitution

In previous chapters, we discussed the issue of substitution, pointing out that ecological economists believe we cannot substitute efficient cause for material cause, except at a very small margin, while neoclassical economists (NCEs) argue that manmade capital is essentially a perfect substitute for natural capital. After all, haven't people been arguing that resource exhaustion is imminent since the time of Malthus, and haven't they consistently been proven wrong? NCEs (and many other people) argue that as a resource grows scarce, the price increases, encouraging the invention and innovation of substitutes. It is true that some civilizations in the past appear to have disappeared from exhaustion of their natural capital, but NCEs assert that the market has averted such collapses since the advent of capitalism. This, of course, is tantamount to claiming that the profit motive is more powerful than the survival motive.

One can certainly find numerous examples where the profit motive has apparently produced substitutes for scarce resources, but that's no guarantee that there will be adequate substitutes for every vital resource. Moreover, even if the profit motive does provide a marvelous spur to our creative processes, what happens when the resources becoming increasingly scarce are public goods? Such goods have no price, and there will therefore be no price signal telling our entrepreneurs that we need substitutes, nor is there any profit to be made by creating such substitutes.²⁰ What happens then? Conventional market economics does not address this question.

The Distribution of Public Goods Through Space

Another complication arises with some public goods, particularly those produced by ecosystem function, which is highly relevant to policy choices. We pointed out earlier that ecosystems can provide different public goods and services for different populations. For example, water regulation and storm surge protection provided by intact mangrove forests are

²⁰Not to mention, as we pointed out in Chapter 6, that it is probably much easier to create substitutes for ecosystem structure (stock-flow resources, raw material) than for ecosystem services provided by the wickedly complex interaction of structural elements in an ecosystem.

local public goods, the role of mangroves as a fishery nursery is a regional public good, and global climate stability promoted by forest carbon storage is a global public good. Individuals are ultimately responsible for how ecosystem stock-funds are treated, they will prefer market flows over public good services, and the two are often mutually exclusive. Unlike individuals, society in some circumstances should prefer public goods over the production of private goods that deplete them. However, local communities may show little concern for providing national public goods. Sovereign nations may show little concern for providing global public goods. Thus, decision makers at different levels (individual, local, national, international, intergenerational) will have different incentives for preserving or destroying ecosystem function, and these incentives must be understood in order to develop effective policies that meet differing needs at all levels. Unfortunately, political systems are largely based on the nation-state or smaller political units, and hence are inadequate for addressing global issues.

The inadequacy of existing political and economic systems for managing public goods is particularly problematic in light of the fact that many ecosystem services are public goods that provide vital services. On the global level, such functions include protection from excessive solar radiation, global climate regulation, and the role of biodiversity in sustaining the web of life. On the local level, ecosystems provide microclimate regulation (often critical for successful agricultural production), buffering from storms, and maintenance of water quality and quantity, all of which may be essential for community sustenance.

■ EXTERNALITIES

Another important type of market failure is known as an externality. An **externality** occurs when an activity or transaction by some parties causes an unintended loss or gain in welfare to another party, and no compensation for the change in welfare occurs. If the externality results in a loss of welfare, it is a negative externality, and if it results in a gain, it is positive. The **marginal external cost** is the cost to society of the negative externality that results from one more “unit” of activity by the agent.

The classic example of a negative externality is a coal-fired utility plant that moves in next door to a laundry service that air-dries its wash. The soot from the coal plant dirties the laundry, and the laundry service receives no compensation from the coal utility. Both air and water are great conveyors of externalities. If a farmer allows his cattle to defecate in a stream flowing through his property, all those downstream from him suffer the negative externality of polluted water. Alternatively, a farmer might reforest his riparian zone, reducing access by cattle. The canopy shades

the stream, killing in-stream vegetation. Water can now run faster, allowing it to scour sediments out of buried springs in the stream, thereby increasing water flow.²¹ Shaded water is cooler, reducing the ability of some harmful bacteria to thrive, thereby increasing water quality. Downstream landowners benefit from these positive externalities. Similarly, if I plant a beautiful flower garden that all my neighbors enjoy yet I receive no compensation from them, I have created a positive externality. A final example is the pollution we spew every time we drive a car, which decreases air quality and contributes to global warming.

Because the agent conducting the activity in question is not compensated for positive externalities and pays no compensation for negative ones, she does not take into account these costs or benefits in her decision to pursue the activity. In the case of negative externalities, the agent carries the activity too far. With positive externalities, the agent engages in too little of the activity. If the agent conducting the activity were to be appropriately compensated or charged, there would be no more externality; the activity would be carried out until marginal benefits equaled marginal costs, not only for the agent conducting the activity but also for society.

As in the case of public goods, economists have suggested that assigning property rights will eliminate the externality problem. If the laundry has the right to clean air, then the coal utility will be forced to pay the laundry service for dirtying its laundry.²² Once compensation is paid, the externality is gone. Alternatively, it would be possible to assign the right to pollute to the coal utility. In this case, the laundry would have to pay the coal utility not to pollute.²³ As Ronald Coase showed in perhaps the most widely cited article ever written on externalities, under certain circumstances, it doesn't matter whether the utility is assigned the right to pollute or the laundry is assigned the right to clean air.²⁴ In either case, the negotiated outcome will lead to an identical amount of pollution, precisely at the level marginal costs of pollution to the laundry are just equal to the marginal benefits to the utility. The implication is that the external-

²¹Note that the outcomes of reforestation are highly dependent on both the system in question and the techniques and species used for reforestation. In some cases, reforestation may reduce water quantity.

²²In reality, this will not necessarily lead to an efficient solution in a dynamic setting. For example, if the payment makes the laundry service profitable, another laundry may locate nearby, which would also be profitable with a subsidy from the utility. For fairly obvious reasons, it is inefficient if the promise of a subsidy from the utility attracts businesses that are otherwise harmed by the utility's presence.

²³In this case, we would have to look at installation of pollution reduction equipment as generating a positive externality for which the laundry service must compensate the public utility.

²⁴R. Coase, The Problem of Social Cost, *Journal of Law and Economics* 3: 1–44 (October 1960).

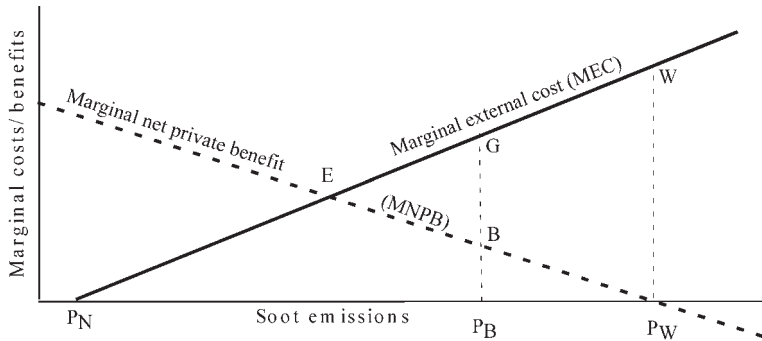


Figure 10.2 • “Optimal” pollution levels. “Optimal” pollution levels are theoretically determined by the intersection of the marginal external cost (MEC) curve and the marginal net private benefit (MNPB) curve.

ity issue requires no government intervention—market forces are perfectly capable of sorting it out. This is known as the **Coase theorem**.

A graphic analysis may help make this a bit clearer. Figure 10.2 shows pollution on the X-axis and marginal costs and benefits on the Y-axis. The coal-fired utility benefits from polluting, while this pollution imposes costs on the laundry service. There are several technologies available for reducing pollution. If the problem is simply to reduce local pollution, the utility can install higher smokestacks at a reasonably low cost. The higher the smokestack, the more it reduces local pollution, but the more expensive it is to install. To reduce pollution even further, it is necessary to install scrubbers, which are more expensive than higher smokestacks. To really reduce pollution, the coal-fired utility would have to convert to natural gas at a very high cost, and to eliminate pollution, the plant would have to close down,²⁶ imposing yet higher costs. It is clear, then, that the marginal cost of reducing pollution is increasing, which is the same as saying that the marginal net private benefits (MNPB) of pollution are decreasing. This is depicted by the MNPB curve, the dashed line sloping down to the right.

For the laundry service, the cost of a small amount of pollution is negligible. As the amount of pollution increases, however, drying the laundry outside makes it noticeably dirtier. Initially, this might result in fewer

The *Coase theorem* states that the initial allocation of legal entitlements does not matter from an efficiency perspective so long as they can be exchanged in a perfectly competitive market.²⁵

²⁵R. Cooter, “Coase Theorem.” In *The New Palgrave: A Dictionary of Economics*, New York: Macmillan, 1987, pp. 457–459.

²⁶Reducing pollution has real energy costs attached to it. For example, studies by Cutler Cleveland have shown that unregulated coal-fired utility plants generate 9 units of energy for every unit of energy required to build and maintain the plant, feed the workers, mine and transport the ore, etc. Scrubbers and other antipollution devices consume considerable energy in manufacturing, installation, and maintenance, and reduce the energy efficiency of plants from 9:1 to 2.5:1. Cited by C. Flavin in *Worldwatch*, January/February 2001, response to letter, p. 6.

customers and reduced profits. If the pollution gets worse, the laundry service will have to move its laundry lines indoors or install electric dryers. Very severe pollution might even reduce air quality inside the laundry service, imposing the need for an air filtration system. Each of these options is more expensive than the last. The basic point is that the marginal cost of pollution to the laundry is increasing. We depict this with line MEC, the marginal external cost curve, sloping up and to the right. In reality, it is unlikely that either the MNPB or the MEC curves would be smooth. Technologies are “chunky”—one cannot purchase smokestacks one row at a time, or scrubbers in small units. External costs often exhibit thresholds beyond which costs increase dramatically. The assumption of smooth curves, however, does not affect the discussion.

Economic efficiency demands that $MNPB = MEC$ (a variant of our basic $MB = MC$ rule of efficiency). Let's assume initially that there are no laws preventing the coal-fired utility from polluting. The utility will produce until the MNPB of additional pollution is zero, which occurs at point P_W . However, at pollution level P_W , the laundry service suffers very high costs from the soot—the area $EW P_W$ measures the net loss to society. The laundry service could increase its profits by paying the coal utility anything less than $P_W W$ to reduce pollution by one unit from point W . The coal utility will increase its profits as long as the laundry service pays it anything at all to reduce pollution. If the laundry pays the coal utility to reduce pollution to P_B , there is still room for mutually beneficial exchanges. At this point, if the laundry pays any sum less than $P_B G$ it will benefit, and if the utility receives any amount greater than $P_B B$ it will benefit. The possibility for mutually beneficial exchanges (payments from the laundry service to the coal utility to reduce pollution) continues until we reach point E , where $MNPB = MEC$ and we have achieved a socially efficient outcome in the absence of government intervention. The same result applies if the laundry has the right to clean air, and we start at pollution level P_N . In this case, the coal utility will keep paying the laundry service for the right to pollute until reaching point E .

A serious problem with this analysis is that it assumes that both the laundry and the utility are able to pay (i.e., that there are no wealth effects) and that there are no real transaction costs. If the laundry earns insufficient profits to pay the utility to decrease pollution and must go out of business if the utility is assigned the right to pollute, that is an example of the *wealth effect*. A **transaction cost** is simply the cost of thrashing out an agreement, which can include legal fees, the cost of gathering information, locating the interested parties, the time cost of bargaining and so on. Transaction costs may not be that high in the case of one laundry service and one utility. However, pollution from a coal-fired utility affects many people in many locations. For example, a coal-fired utility in Ohio

causes local pollution in the surrounding community and also contributes to smog and acid rain at least as far away as New York and Canada; it contributes to global warming, which affects everyone on the planet.

Similarly, a farmer who pollutes his water supply may be one of many upstream farmers affecting thousands of downstream neighbors. Bringing all the relevant agents together to the negotiating table would range from difficult to impossible, and even if it could be achieved, free-riding would become a problem. For example, if I live on the stream polluted by upstream farmers and my neighbors agree to pay the farmers some sum to reduce pollution, I may prefer that level of reduction for free to even more reduction at a positive cost to myself.

Coase explicitly recognized that high transaction costs could justify government intervention, though this caveat seems to have been lost on many of his devoted followers. In the definition of the Coase theorem supplied earlier, we must take “perfect market” to mean no transaction costs and perfect information. Even then, however, the presence of wealth effects undermine the theorem. Remember that efficiency of allocation is defined only for a given distribution. Since vesting property rights in the polluter implies a different distribution (wealth effect) than does vesting them with the pollutee, the two cases are not strictly comparable in terms of efficiency—we simply cannot say that the two situations envisaged by Coase are “equally efficient,” because they are based on two different distributions of wealth. “Efficiency” in the Coase theorem is from the viewpoint of society, and different legal entitlements also have significant direct impacts on the well-being of the parties involved.

In reality, transaction costs are likely to be very important any time an externality impacts more than a very few agents, which is the general rule rather than the exception. Many externalities, in fact, affect the provision of public goods and confront the same obstacles as the provision of private goods. Yet again we must stress that all economic production requires raw material inputs and generates waste outputs, thus depleting ecosystem services. All economic production inevitably generates “externalities.” Indeed, “externalities” is a misnomer, since there is an unbreakable link (throughput) between resource depletion, production, and waste emissions, so these “externalities” are actually 100% *internal* to the economic process. If converting a forest to farmland imposes negative externalities at the local, national, and global levels, transaction costs for an efficient solution would be prohibitively expensive. When externalities affect future generations, we must accept that transaction costs between generations are infinite, and the market will not solve the “externality” problem unaided.

We must also recognize that perfect information concerning human impacts on ecosystems is unobtainable. Uncertainty and ignorance are

such ubiquitous obstacles to market efficiency that they can be considered market failures in their own right. It is hard to equate MC and MB when we do not know the probabilities of possible outcomes, and it is virtually impossible to do so when we are ignorant even of the possible outcomes.

■ MISSING MARKETS

For a market to function optimally, everyone who would want to produce or consume the goods being marketed must be able to participate. For example, if the *Mona Lisa* were to be auctioned off and only people from Waco, Texas, were allowed to participate, it might not fetch as high a price as it would on the international market. Yet the fact is that future generations cannot possibly participate in today's markets, and hence today's market prices will not reflect their preferences. The market can therefore only "efficiently" allocate resources if we assume that future generations have no rights whatsoever to the resources being allocated.

How could we provide future generations with property rights to resources? One way to bring this about would be to impose sustainability criteria. For example, we might decide that the rights of future generations to certain resources, such as the ecosystems responsible for generating life-support functions, are inalienable²⁷—much like human and political rights, where entitlements are not decided by efficiency criteria. As we deplete nonrenewable resources, we could invest a sufficient percentage of the profits in renewable substitutes to replace the depleted resource (we'll describe this option in greater detail later). For renewable resources, we could make sure that they were never depleted beyond their capacity to regenerate. If renewables were depleted below their maximum sustainable yield (MSY), we would need to bequeath some substitute that would compensate for the reduction in future harvests. Or we could lower offtake (passive investment in natural capital) enough to replenish the renewable resource to at least the MSY level.

How we handle intergenerational gambles with unknown reward structures is an ethical issue, but it would certainly seem that most ethical systems would demand at the very least that we do not risk catastrophic outcomes for the future in exchange for nonessential benefits today. Given our ignorance of ecosystem function, this means we would have to stay well back from any irreversible ecological thresholds. Such sustainability criteria would essentially distribute resources between generations, and the market could then function to allocate them within a generation.

²⁷D. Bromley, *Environment and Economy: Property Rights and Public Policy*, Oxford, England: Blackwell, 1991.

Alternatively, we could just continue to act on the ethical assumptions of neoclassical economics. If we are indeed rational maximizers of self-interest, and Pareto efficiency is an objective criterion for allocation, then the rights of future generations can be completely ignored. After all, as Kenneth Boulding once asked: What have future generations ever done for us? We certainly cannot increase our own well-being by redistributing resources to the future.

In reality, conventional economists do not disregard future generations entirely, but in their analyses they do systematically discount any costs and benefits that affect future generations. In Chapters 11 and 12, we'll look at how the convention of discounting can affect decisions concerning natural resource use.

Intertemporal Discounting

Do conventional economists really ignore future generations? In a standard economic analysis where they have to compare costs and benefits in the future with costs and benefits in the present, conventional economists will systematically discount any costs and benefits that affect future generations. There are some very plausible reasons for giving less weight to resources in the future than resources in the present, and we will explore the topic in some detail in Chapter 15. Here we offer only a brief introduction to help you understand, in the following chapters, how the convention of **intertemporal discounting** can affect decisions concerning natural resource use. When evaluating present and future values, intertemporal discounting is the process of systematically weighting future costs and benefits as less valuable than present ones.

Why should resources in the future be worth less than resources today? If I have \$100 today, I can invest it in some profit-making venture, and I will have more than \$100 next year. In perhaps the simplest example, if I can safely invest money in the bank at 5% real interest (i.e., at an interest rate 5% greater than inflation), then I will always prefer \$100 today to anything less than \$105 a year from now for the simple reason that if I have the money today I have the option to spend it now or allow it to become \$105 next year. Next year, of course, I would again have the option to spend the money, or leave it again to grow at 5% to become \$110.25, then \$115.76, then \$121.55, and so on indefinitely. Conversely, \$100 in the future is worth less than \$100 today because of the *opportunity cost* involved (the lost opportunity to invest), and the farther in the future we look, the less the money is worth. Most conventional economists assume that money is an adequate substitute for anything, and therefore anything in the future is worth less than the same thing today. In general, the present value (PV) of a sum of money t years in the future, X_t , when the interest rate is r , will be given by

$$PV = X_t / (1 + r)^t$$

If we have a stream of money at different dates in the future, we can calculate the PV for each yearly amount, and sum them. This is basically what is done in the more complicated formula below.

A standard cost-benefit analysis (CBA) will tell us the **net present value (NPV)**—the value to us today—of a given stream of costs and benefits through time. The farther off in time that a cost or benefit occurs, the more we discount its present value. The basic equation is:

$$NPV = \sum_{t=0}^T (Benefits_t - costs_t) \left(\frac{1}{1+r} \right)^t$$

The discount rate is r , and the discount factor is $1/(1 + r)$. If we let $r = 5\%$, as in the earlier example, then the discount factor is $1/1.05$, which is less than one. The letter t represents time, and $benefits_t - costs_t$ is simply net benefits in period t . As t increases, the discount factor is raised to a larger and larger power, and, because it is less than one, raising it to a higher power makes it ever smaller, reducing the net present value by ever more the farther in the future the benefit or cost is. The symbol Σ tells us to sum together the net benefit stream from time 0 to time T .

THINK ABOUT IT!

Do you think the convention of estimating net present value ignores the rights of future generations? Why or why not?

SUMMARY POINTS

What are the most important points you should take home from this chapter? Markets only balance supply and demand, possibility with desirability, under a very restrictive range of assumptions. Among others, goods and resources must be both excludable and rival (where excludable implies well-defined and enforced property rights), market actors must be able to make transactions with zero cost (which would automatically eliminate most transactions), and people must have perfect information concerning all the costs and benefits of every good. Even if all of these conditions are met, markets will not account for future generations. In reality, these conditions are never met, though many excludable and rival goods meet these criteria well enough that the market is a very useful allocation mechanism. When resources are nonrival and/or nonexcludable, the specific combination of these characteristics has much to tell us about how the resources should be allocated. You should clearly understand the implications of these various combinations. Remember also that social institutions are required to make resources excludable, but some resources

are nonexcludable by their very nature, and rivalness is a physical property.

In particular, we must recognize that the “optimal” production of pure public goods cannot be based on the criterion of Pareto efficiency. The public good problem appears to be beyond the scope of market allocation. You might think about policies and institutions that could be effective mechanisms for allocating public goods and the ecological fund-services that provide many of them. One possibility worth considering is a participatory democratic forum that captures a broader spectrum of human values than self-interest, and does not weight participant values solely by the purchasing power at their disposal.

BIG IDEAS to remember

-
- | | |
|----------------------------------|-----------------------------|
| ■ Excludable | ■ Coase theorem |
| ■ Rival | ■ Transaction costs |
| ■ Congestible | ■ Wealth effects |
| ■ Public goods | ■ Missing markets |
| ■ Open access regimes | ■ Intertemporal discounting |
| ■ Nonrival, excludable resources | and net present value |
| ■ Externalities | ■ Inalienable rights |
-



Market Failures and Abiotic Resources

Now that we have seen the circumstances under which the market is an effective allocation mechanism, we can examine how the market performs with respect to the goods and services provided by nature. There are eight classes of these, as discussed in Chapters 5 and 6:

1. Fossil fuels (nonrenewable stock)
2. Minerals (partially recyclable, nonrenewable stock)
3. Water (nonrenewable stock, and/or fund, depending on use, recyclable)
4. Solar energy (indestructible fund)
5. Ricardian land (indestructible fund)
6. Renewable resources (renewable stock)
7. Ecosystem services (renewable fund)
8. Waste absorption capacity (renewable fund)

If a resource is excludable, its market allocation is possible. If it is rival, we understand all the impacts of its use and production and consumption generate no externalities, then market allocation is also efficient within the current generation. If the well-being of future generations is not affected by the use of the resource, then market allocation may also be intergenerationally fair. As we will see, however, no good or service provided by nature meets all of these criteria. In this chapter we examine abiotic resources and lay some of the groundwork for subsequent discussion of policies that can improve the allocation of these resources.

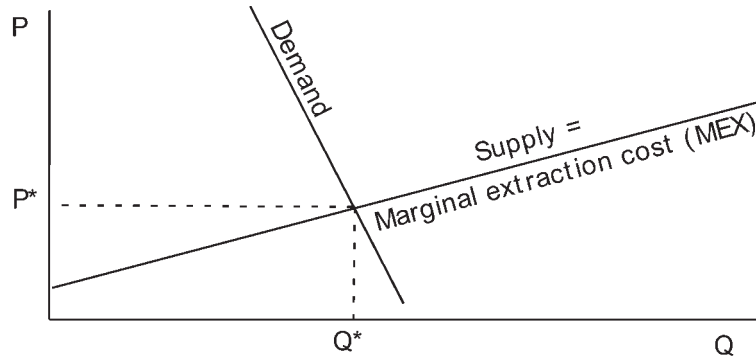


Figure 11.1 • Optimal extraction (Q^* , P^*) of fossil fuels in the absence of scarcity and market failures.

FOSSIL FUELS

Fossil fuels are both rival and excludable, and thus can be allocated by market forces. If we ignored resource scarcity (or the future) and market failures, the optimal allocation of fossil fuels would simply be the intersection of the demand curve with the supply curve, where the supply curve would equal **marginal extraction costs (MEX)**,¹ as depicted in Figure 11.1.

External Costs

However, problems arise. First of all, the production and consumption of fossil fuels generate serious externalities at the local, regional, and global levels. Most of these externalities are in the form of public bads. Examples of these externalities are categorized in Table 11.1 according to their spatial and temporal characteristics.

Many of these externalities have different impacts at different spatial levels. Optimal extraction of fossil fuels would need to include these marginal external costs, as shown in Figure 11.2, where MEC shows the cost of extraction including externalities. Because these externalities are so widespread, affecting not only virtually everyone in the world alive today but future generations as well, transaction costs for resolving these externalities through the market would be infinite. Given the inability of the unfettered market to address these externalities, extra-market institutions (e.g., government regulation) will be required.² However, these institu-

¹MEX include all costs, i.e., the costs of equipment and labor and the opportunity cost of money invested.

²This is not to say that government regulations cannot create market incentives for reducing externalities, as we will see in Chapter 17–18.

■ **Table 11.1****SPATIAL AND TEMPORAL CHARACTERISTICS OF SELECTED EXTERNALITIES ASSOCIATED WITH FOSSIL FUEL EXTRACTION AND CONSUMPTION. (A SMALL X DENOTES RELATIVELY MINOR IMPACTS.)**

Externality	Local	Regional	Global	Intergenerational
Global warming			X	X
Acid rain	x	X		X
Oil spills	X	X		X
Damage from extraction (see Table 11.2)	X			X
War ^a	X	X		X
Water pollution	X	X		X
Soil pollution	X			X
Air pollution (gaseous)	X	X	x	x
Air pollution (particulate)	X			

^aThe number of wars that have been fought and are currently being fought over the control of fossil fuels argues for the treatment of some wars, or at least some military expenditures, as an externality of fossil fuel production. See, for example, M. Renner, *WorldWatch* Paper 162: *The Anatomy of Resource Wars*. Washington, D.C.: WorldWatch (2002).

tions must be on the scale of the problem they address, as there is little incentive for governments (generally the most relevant institutions) to deal with externalities beyond their borders. At present, appropriate institutions for addressing international externalities either do not exist or are inadequate.

User Costs

Another problem is that fossil fuels are a nonrenewable resource upon which the well-being and even the survival of future generations is highly dependent. Even ignoring future generations, economists agree that the use of a nonrenewable resource now increases scarcity (decreases supply) in the future. As supply goes down, price should go up. Therefore, if the owner of a nonrenewable resource extracts that resource today, she loses the option of extracting it in the future when the price is higher.

The more of a resource we extract today, the greater the current supply and the lower the current price. Also, greater extraction now means greater scarcity in the future, and a higher future price. All else being equal, the **marginal user cost (MUC)**—the opportunity cost of producing *one more unit* of the resource today instead of in the future—should therefore be increasing with total production.³

The marginal user cost is a real cost of production, and it must be added to MEC and MEX to give the full cost per unit that represents all

User cost is the opportunity cost of nonavailability of a natural resource at a future date that results from using up the resource today rather than keeping it in its natural state.

Marginal user cost is the value of one more unit of the resource in its natural state. In a perfectly competitive economy, marginal user cost would in theory equal the price of a resource minus its marginal extraction cost.

³We caution that some theoretical studies suggest that rising marginal extraction costs and/or the presence of an inexhaustible substitute (a backstop technology) may lead to declining marginal user costs over time. Different sets of assumptions in mathematical models lead to different results.

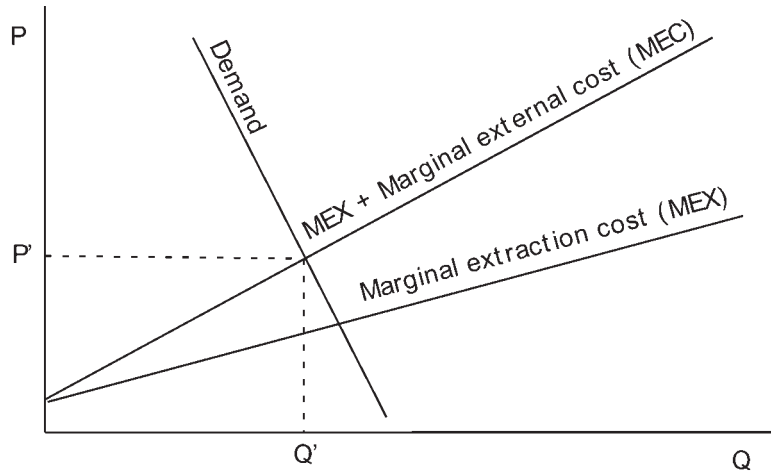


Figure 11.2 • Optimal extraction of fossil fuels (Q' , P') in the presence of negative externalities (global marginal external costs), without scarcity.

marginal opportunity costs. The individual producer takes prices as given, and therefore should produce up to the point where marginal benefit (price) equals marginal cost ($MEC + MEX + MUC$), as shown in Figure 11.3. Of course, when the producer does not have to pay marginal external costs, she is likely to ignore them.

User cost can also be thought of as the value of the resource in its natural state, the in-ground value before it has been extracted. Marginal user cost

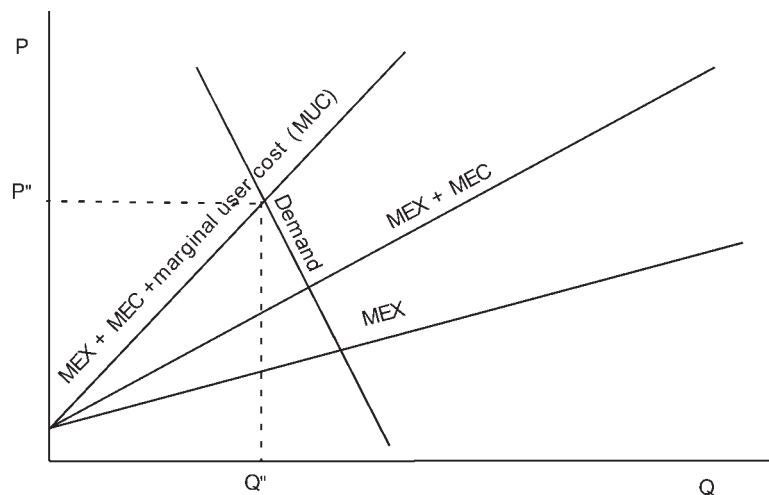


Figure 11.3 • Optimal extraction (Q'' , P'') of fossil fuels in the presence of scarcity and negative externalities.

is the value of one more unit of the resource in its natural state. Theoretically, in a competitive market, producers would pay resource owners a per-unit fee for extraction rights precisely equal to the marginal user cost.

THINK ABOUT IT!

Can you explain why, in a competitive market, producers would pay resource owners a per-unit fee equal to the MUC for the right to extract a resource?

This fee is known as a **royalty**; that is, the marginal user cost should be equal to the price of the resource minus the marginal extraction cost. Because no human effort is required to produce a natural resource in its natural state, user cost is unearned economic profit, also known as economic rent. As explained in Chapter 9, **rent**, or scarcity rent, is defined as unearned profits, or the payment to a firm for a product above and beyond what is required to bring that product to market.

We have explained that marginal user cost should be rising with increasing scarcity, but how fast will it increase? We must remember that user cost is the opportunity cost of extraction today, while the discount rate is the opportunity cost of leaving a resource in the ground instead of extracting it and investing the profits in the most profitable alternative productive activity. Conventional economists argue that, all else being equal, the optimal rate of resource extraction is one at which increasing in-ground scarcity drives the marginal user cost up at a rate equal to the usual rate of return on alternative (aboveground) investments.⁴ This is known as the **Hotelling rule**, after economist Harold Hotelling who first stated it. As a result, extraction rates should decrease through time, causing the price to increase.

If we imagine that MEC and MEX are zero, then the price under an optimal extraction regime will increase at the discount rate. This is an intuitive result—if the price increased more slowly than the discount rate, the resource owner would maximize profits by extracting the resource faster and investing the profits, which would then grow faster than the value of the resource in the ground. Alternatively, if the price of fossil fuels is growing faster than the discount rate, then leaving the resource in the ground to appreciate in value generates the greatest profits. In other words, if the opportunity cost of leaving the resource in the ground (the discount rate times the current value of the resource) is greater than that of extracting it (the user cost), we extract, and vice versa.

In theory, the market mechanism automatically (through the invisible hand) incorporates marginal user cost into the market price, and is equal to the market price minus the extraction cost. In reality, as natural resource

Royalty is the payment to the owner of a resource for the right to exploit that resource. Theoretically, in a competitive market, the per-unit royalty should be equal to the marginal user cost.

⁴H. Hotelling, The Economics of Exhaustible Resources, *Journal of Political Economy* 2: 137–175 (1931).

markets in general are highly imperfect because of cartels, an absence of competition, poorly defined property rights, and imperfect information, they will not reveal true user costs. As an alternative measure, economists can estimate the time of total depletion of the resource, the time when we'll have to turn to the best available substitute, assuming there is one. If there is a reasonable renewable substitute (e.g., solar power) or extremely abundant substitute (e.g., hydrogen fusion) for the resource in question (known as a backstop technology), the price of the resource can never rise above the price of the substitute. This reduces the opportunity cost of using the resource in the present—that is, it lowers the user cost, thereby leading to more rapid extraction and a lower price. To determine the user cost, the extra unit cost of the best substitute, over and above that of the depleted resource, is estimated. That amount is then discounted from the future date of exhaustion back to the present to tell us the marginal user cost.

In summary, the user cost will be low if (1) the discount rate is high, (2) the exhaustion is far in the future because either reserves are large or annual usage rates are low, and (3) good substitutes are expected to be available. And vice versa. We see once again the importance of the discount (interest) rate, expectations about substitutes, and uncertainty about stocks in the ground.

Box 11-1
**MARGINAL USER COST IN SUSTAINABLE
INCOME ACCOUNTING**

It is interesting that in discussing user cost, John Maynard Keynes, arguably the most influential economist of the twentieth century, was interested mainly in applying the concept to depreciation of the fund of manmade capital (in order to arrive at a proper measure of income). He made reference to the more “obvious” case of accounting for user cost in a copper mine (natural capital) as a way of clarifying his argument.^a Nowadays, if texts discuss user cost at all, they refer to the more “obvious” necessity of accounting for depreciation of manmade capital as an argument for applying the same logic to natural capital. Perhaps this reversal is a measure of how much we have recently come to neglect natural capital.

In any case, recalling our terminology of previous chapters, it is clear that user cost is a necessary charge for the depletion of stocks (natural or manmade inventories) and the depreciation of funds (natural and manmade productive equipment). In proper accounting usage, both inventories and machines are capital. Inventories are depleted; machines are depreciated. Both require the accounting of user cost. If user cost is not deducted in calculating income, then income will be overstated and will not be sustainable. In most national accounts around the world, user cost is erroneously counted as income.^b Keynes, as the major architect

of modern macroeconomics, was very interested in getting a correct measure of national income.

Few nonrenewable natural resources are fund-service in nature. We will examine the implications of user cost for the stock-flows and fund-services provided by renewable resources in Chapter 12.

^aJ. M. Keynes, *The General Theory of Employment, Interest, and Money*. Orlando, FL: Harcourt Brace (1991) p. 73.

^bSee S. El Serafy, *The Proper Calculation of Income from Depletable Natural Resources*. In Y. J. Ahmad, S. El Serafy, and E. Lutz, eds. *Environmental Accounting for Sustainable Development, A UNEP–World Bank Symposium, Washington, D.C.: World Bank, 1989*; and S. El Serafy, *Green Accounting and Economic Policy*, *Ecological Economics* 21: 217–229 (1997).

Flaws in the Analysis

From the perspective of ecological economics, however, this analysis of fossil fuels is inadequate. First, it looks only at the net present value of the resource for the existing generation, ignoring any ethical obligations to leave some of the resource for future generations; that is, it focuses on efficiency, ignoring scale and distribution. Second, neither producers nor consumers currently pay marginal external costs. Third, empirical evidence contradicts the conventional theory; as Figure 11.4 shows, oil

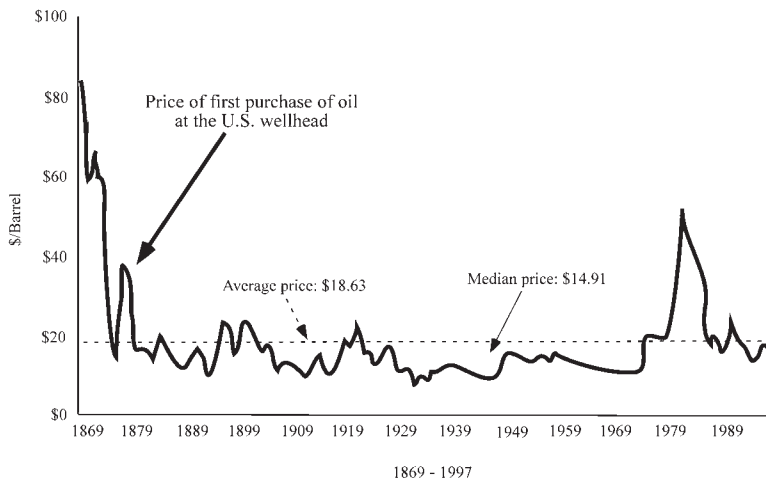


Figure 11.4 • Oil prices from 1869 to 1997, in 1996 dollars. These are the prices paid in the U.S. at the wellhead, net of transportation costs. The volatile nature of prices in the short run is obvious (note particularly the surge in prices in response to reduced OPEC production in the 1970s), but perhaps more noteworthy is the relative stability of prices over the last 120 years (Source: Adapted from WTRG Economics, 1998. Online: <http://www.wtrg.com/prices.htm>.)

prices have been fairly constant for the last 130 years. We'll return to this last point following our discussion of mineral resources.

■ MINERAL RESOURCES

Mineral resources are also rival and excludable and amenable to market allocation. As in the case of fossil fuels, their production and consumption generate serious externalities. In fact, mining accounts for nearly half of all toxic emissions from industry in some countries, such as the United States.⁵ As many of these negative externalities are less well known than those associated with fossil fuel use, we have summarized them in Table 11.2.

Although many of these externalities are fairly localized compared to problems from fossil fuel emissions, they can be very persistent and severe. For example, acid mine drainage still occurs on mine sites worked by Romans over 1500 years ago. Over 500,000 abandoned mine sites exist in the U.S. alone,⁶ with estimated cleanup costs of \$32–\$72 billion.⁷ Again, transaction costs for resolving these externalities will be extremely high to infinite, depending on our concern for future generations, and they cannot be resolved by unregulated markets.

It is worth noting here an interesting anomaly. Within a generation, for the market to efficiently allocate resources, they must be rival. However, future generations cannot participate in today's markets. Thus, if a good is rival between generations—that is, its use by one generation prohibits use by another—the market will still not allocate it efficiently because future generations cannot participate. Fossil fuels are rival between generations. Mineral resources, to the extent they can be recycled, are rival within a generation, but less so between generations. Thus, if mineral resources were efficiently recycled and had no negative externalities associated with their production and consumption, market allocation could be both intragenerationally efficient and intergenerationally fair. The more efficient the recycling process, the lower the marginal user cost, and the less the theoretically efficient price would need to rise over time. However, the record shows low levels of recycling, enormous increases in demand (from 93 million metric tons in 1900 to 2900 mil-

⁵P. Sampat, *From Rio to Johannesburg: Mining Less in a Sustainable World*. World Summit Policy Brief #9. World Watch Institute News Releases. Online: <http://www.worldwatch.org/press/news/2002/08/06/>.

⁶Center for Streamside Studies. Online: <http://depts.washington.edu/cssuw/Publications/FactSheets/minec.pdf>.

⁷Cleanup costs for some mines have reportedly been greater than the value of minerals extracted. Environmental Media Services, *Mining Companies Profit from Public Lands While Taxpayers Pay for Cleanup*, 2002. Online: http://www.ems.org/mining/profits_costs.html.

■ **Table 11.2****PRODUCTION EXTERNALITIES OF MINERAL RESOURCE EXTRACTION, SPECIFICALLY FROM HARD ROCK MINES**

Externality	What Is It?	What Does It Affect?
Acid mine drainage	Metal sulfides are common in mineral ores and associated rocks. When these rocks are mined and crushed, exposure to air and water oxidizes these sulfides, generating acids and toxic heavy metal cations.	Water required by oxidation also washes products into nearby surface water and aquifers; in addition to acidification effects, heavy metals build up in animal populations and humans.
Erosion and sedimentation	Heavy machinery, strip mines, and open pit mines destroy surface vegetation holding soils in place; water washes away small particles from erosion and waste materials, depositing it elsewhere.	Major impacts are on wetlands and other aquatic habitats; soil organisms, vegetation, and restoration efforts also affected.
Cyanide and other chemical releases	Cyanide and other toxic chemicals are regularly used to help extract minerals.	Cyanide released into ecosystem has adverse impacts on water, soil, aquatic, organisms, wildlife, waterfowl, and humans.
Dust emissions	Ore crushing, conveyance of crushed ore, loading bins, blasting, mine and motor vehicle traffic, use of hauling roads, waste rock piles, windblown tailings, and disturbed areas all generate dust.	Dust can be an air pollutant, and may also transport toxic heavy metals.
Habitat modification	Mining can have dramatic impacts on landscape and uses enormous amounts of water.	Ecosystem structure and function are affected.
Surface and groundwater pollution	Mining uses massive quantities of water, pumping water from mines affects water tables, and mine wastes pollute water.	Altered surface and groundwater flows, with accompanying impacts on wetlands and other water-dependent habitats.

Source: EPA Office of Waste Water Management, *Hardrock Mining: Environmental Impacts*, Online: <http://www.epa.gov/owm/permits/hrmining/env.htm>.

lion metric tons in 1998 in the U.S. alone), and substantial decreases in real prices.⁸

Do Prices Reflect Scarcity?

How do we explain the major anomaly between the empirical fact of falling prices of nonrenewable resources and the theoretical prediction of rising prices? Conventional economic theory generally assumes that prices

⁸D. E. Sullivan, J. L. Sznopce, and L. A. Wagner, *Twentieth Century U.S. Mineral Prices Decline in Constant Dollars*. Open File Report 00-389, Washington, DC: U.S. Geological Survey U.S. Department of the Interior, 2000.

increase as a function of scarcity, and it is an unalterable physical fact that extraction has reduced the quantity of in-ground stocks of nonrenewable resources. This does not necessarily mean that prices do not reflect scarcity, as long as we assume that scarcity is defined not only by the physical quantity of a resource remaining. Scarcity is also determined by new discoveries⁹ and by the availability of substitutes. Prices equilibrate supply and demand, and if supply increases from new discoveries or demand falls because substitutes are invented, scarcity is reduced, and prices fall as well. For example, fiber optic cables dramatically decreased the demand for copper in telephone lines, which might explain the fall in prices.

However, as we pointed out previously, oil discoveries peaked in 1962, production surpassed new discoveries in 1982, and consumption currently exceeds new discoveries by a factor of two to six. What's more, while we do have more potential substitutes for oil, relative to 100 years ago we have created far more technologies that depend on oil (complements) than technologies that substitute for oil. Just as substitutes reduce resource scarcity, complements increase it. Nonetheless, steady increases in the demand for oil have apparently not affected the price.

What is the explanation? First of all, we must recognize that if prices reflect in-ground scarcity, they do so very poorly, and for obvious reasons. There is considerable debate even among the experts about the precise amount of oil left in the ground (though less about ultimately recoverable reserves than about "proven" reserves), and estimates of "proven" reserves have changed dramatically over the years, often increasing substantially even in the absence of new discoveries.¹⁰ If the experts do not know how much remains underground, how can prices tell us?

While prices cannot effectively equilibrate unknown in-ground supply with demand, they can equilibrate the available aboveground supply with demand. Available aboveground supply is determined solely by the rate of extraction, which depends on known deposits, existing infrastructure and technology, as well as the resource owner's decision of how much to extract. Hotelling suggested that a rational producer will limit current production to take advantage of higher prices in future years. However, if real prices are not increasing, the owner has no incentive to leave the fossil fuels in the ground, and would rationally extract the resource as long as the marginal

⁹Remember, new discoveries do not increase the amount of resources in the ground; they just make it easier for us to get them! In the short-run market sense, new discoveries decrease scarcity, while in the long-run physical or geological sense, they increase the scarcity of resources remaining in the ground.

¹⁰In January 1988, Iran, Iraq, and Venezuela each reported a doubling of their reserves, presumably to earn higher quotas under OPEC. In spite of continuous extractions since then, their reported reserves have scarcely changed. C. J. Campbell, *Proving the Unprovable*, *Petroleum Economist*, May 1995.

extraction cost remains lower than the price. Essentially, the producer ignores MUC, as in the simplest static analysis depicted in Figure 11.1.

Even if the producer ignores MUC, we would still expect MEC to increase. As we suggested earlier, economic analysis typically assumes that nonrenewable resources will be mined from the purest, easiest-to-access sources first. As these are depleted, we then move on to sources that are more expensive to extract, again putting upward pressure on prices. However, there are two serious problems with this argument. First, as Norgaard has noted, when we begin to exploit a new resource, we typically know very little about where the best fields are. A great deal of chance is involved with the initial discoveries. Norgaard compared this to the Mayflower. If people always exploited the best resources first, the first pilgrims would have settled on the best land in America. However, prior to their arrival, the pilgrims knew virtually nothing about land resources in North America, and ended up where they did largely by chance.¹¹

Second, as we exploit a new resource, we diminish the total stock, but we gradually acquire more information about where to find it and how to extract it, and more of the resource becomes accessible. Thus, there are two effects at work. The scarcity effect decreases the total amount of resource available, but the information effect increases the amount that is accessible and reduces the costs of extracting it. Thus, as long as the information effect is dominant, the price of the resource should decrease. Eventually, however, the scarcity effect must come to dominate, and the price must then increase. Rather than predicting a gradual price increase in a resource, this model suggests the likelihood of decline followed by sudden, rapid increases. If we combine this analysis with the estimates of petroleum geologists, we would predict a sudden and dramatic increase in oil prices in the next 2–20 years.¹²

This result is particularly important if we are concerned with sustainability. As we pointed out earlier, economists assume that price increases will trigger innovation and generate substitutes for any given resource. If resource owners are optimists, they believe new discoveries will be made and substitutes invented. This means that their resource will not become scarce and its price will not go up (and may even go down). Under such circumstances, it makes sense to extract the resources as fast as possible and invest the returns. If the resources are being extracted quickly, above-ground supply is large, and the price is low. This reduces the incentives for exploration and the development of substitutes. The problem is that

¹¹R. Norgaard, Economic Indicators of Resource Scarcity: A Critical Essay, *Journal of Environmental Indicators and Management* 19(1): 19–25 (July 1990).

¹²D. B. Reynolds, The Mineral Economy: How Prices and Costs Can Falsely Signal Decreasing Scarcity, *Ecological Economics* 31(1): 155–166 (1999); C. J. Campbell and J. H. Laherrère, The End of Cheap Oil, *Scientific American*, March 1998.

developing substitutes requires technology, technological advance requires time, and the less warning we have of impending resource exhaustion, the less time there is to develop substitutes. Perversely, then, in a world of optimists, the pessimist is most likely to be correct, and vice versa.¹³ While these arguments are far from the only ones discrediting the belief that we can ignore resource exhaustion, they are important.

■ FRESHWATER

As the economically relevant characteristics of freshwater depend on the specific use to which it is put, and because it is used in most economic and ecological processes, the economics of water could fill a textbook on its own. Some relevant characteristics of specific water uses can be gleaned from discussions in other sections of this text. Specifically, water in fossil aquifers is a nonrenewable resource similar to fossil fuels with fewer externalities. Water as an ecological fund-service is similar to other ecosystem services discussed in Chapter 12. In this section, we'll limit our discussion to some unique attributes of water as a stock-flow resource. Specifically, we address the facts that water is 100% essential to human survival, that it has no substitutes,¹⁴ and that water distribution systems generally show substantial economies of scale.

The scarcity of clean and available water has traditionally been experienced as a local matter, but international disputes over access to water are increasing, indicating a global dimension to water scarcity. These characteristics have important implications for water markets. Though traditionally supplied by the public sector at least since early civilization's first large-scale irrigation projects, water is a rival good that can be made excludable under most circumstances and therefore technically amenable to market allocation. Indeed, in recent years, more and more cities, states, and even countries are turning their water supplies over to the private sector in the name of greater efficiency, and many neoclassical economists applaud this trend.¹⁵ However, the fact that water is nonsubstitutable and 100% essential means that there are serious ethical implications to the market allocation of water, and it therefore makes a good case study of why just distribution precedes efficient allocation in ecological economics.

In many places, water is very abundant and is used for fairly unimportant activities; it is even wasted. Higher prices for water would reduce this waste. However, because in its most important uses water has no substi-

¹³P. Victor, *Indicators of Sustainable Development: Some Lessons from Capital Theory*, *Ecological Economics* 4: 191–213 (1991).

¹⁴Though technology can, of course, increase the efficiency of water use in some applications.

¹⁵For example, the International Monetary Fund and the World Bank often make privatization of water supply a requirement for loans.

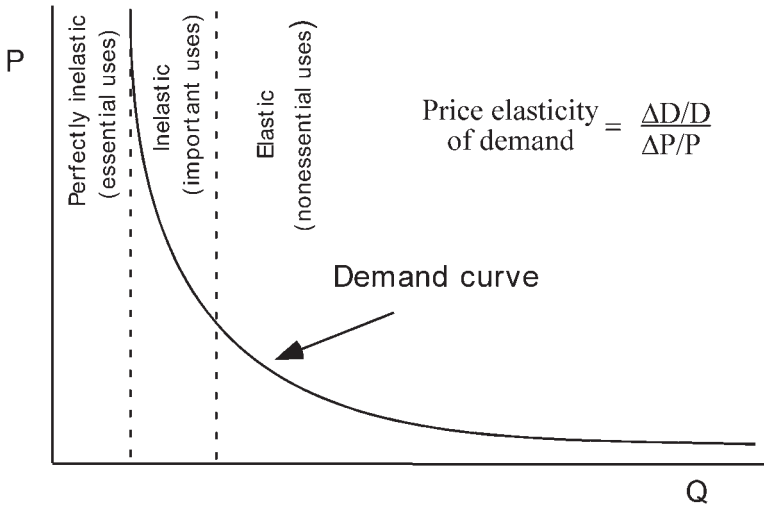


Figure 11.5 • The elasticity of water with respect to price (% change in quantity demanded with respect to % change in price), for different quantities of water.

tutes and is essential to our survival, as water supplies become scarce or prices increase, the demand for water becomes extremely inelastic with respect to price.¹⁶ A 1% increase in price will lead to less than a 1% decrease in demand (Figure 11.5). When water is abundant, we use it for nonessential activities, and demand is price elastic. As water becomes scarcer, we use it only for more important activities, and it becomes inelastic with respect to price. As water becomes still scarcer, it will be used exclusively for essential activities, such as raising food and drinking, and demand becomes perfectly inelastic.

This situation leads to two very serious problems. First is the distribution issue. In the market economy, the most “efficient” use is that use which creates the highest value, and value is measured by willingness to pay. In a world with grossly unequal income distribution and a growing relative scarcity of water, many people have very limited means to pay. “Perfect” market allocation of water could easily lead to circumstances in which a rich person could pay more to water a lawn than a poor family could pay to water the crops it needs to survive. While economically a green lawn might be more efficient, ethically most people would probably agree that survival of the poor should take precedence.

The second issue is efficiency. Markets are rarely perfect, and, in the case of water, they are likely to be less perfect than most. Providing water

¹⁶This issue has played an important role in the history of economic thought, usually discussed under the heading of the “diamonds-water paradox.” See Box 14.1, footnote 1 for an explanation in the context of distinguishing use value from exchange value.

requires substantial infrastructure that would be very costly to duplicate. For this reason it makes sense to have only one provider, so even where water is privatized, there is typically not a competitive market but rather a natural monopoly. A natural monopoly occurs when the marginal cost of production is decreasing, which is the case for many public utilities. Dealing with inelastic demand, the monopoly provider knows that a 10% increase in price will lead to less than a 10% decrease in quantity demanded, leading to higher revenue and lower costs. Moreover, everyone needs water and cannot exit the market no matter how inefficient and expensive the monopoly supplier is. With no threat to their market share, firms bent on maximizing short-term profit may delay needed improvements in infrastructure. Only extensive regulation will deter the private supplier from increasing prices and decreasing quality. With no competition to drive down prices, nor regulation to control costs, private sector provision of water is likely to be less efficient than public sector provision—as well as less just.

■ RICARDIAN LAND

As we explained in Chapters 4 and 5, by Ricardian land we mean land simply as a physical space capable of capturing sunshine and rainfall, and not the various productive qualities inherent to the land itself. The latter qualities, such as soil fertility, we class as ecosystem services. Within a generation, Ricardian land is both rival and excludable, and hence can be allocated by markets. Between generations, it is nonrival, which suggests the market allocation of land might meet the criterion for both efficiency within a generation and fairness between generations.

Before we reach this conclusion, however, we must ask: What is it that makes Ricardian land valuable? Certainly in market terms, the most valuable land in the world is found within the borders of big cities, where prices may pass \$100,000 per square meter, and the least valuable land is generally found in the most deserted areas. What makes land valuable would appear to be proximity to other humans. Some might reply that the low value of land in uninhabited areas is due to other factors, such as extreme cold or extreme heat, and those same factors prevented people from settling there in the first place. But if we look at some of our planet's less-inviting habitats, we find that where they are inhabited, land prices are highest at the sites of densest habitation and lowest where population is thinnest, even if the sites are otherwise virtually identical.

Why would the presence of other humans make land more valuable? Humans are social animals, dependent on each other for both psychic and physical needs. Living near others allows individuals to specialize, and the economic benefits of specialization are common knowledge. Empirically, in a growing economy with growing populations, land appreciates in

value even in the absence of improvements by the landowner. As cities grow, the land on their peripheries becomes more valuable. If the government builds a subway system or road, the value of adjacent land can skyrocket. Proximity to new infrastructure, such as sewage systems, electric grids, highways, and subways, can similarly increase land value.

The truth is that land attains value as a positive externality of the decisions of others. Land values thus result from a market failure, and we cannot simply assume that markets are the best means for allocating even Ricardian land. These insights into land value were first popularized by the nineteenth-century economist Henry George.¹⁷

The origin of land value is not just an academic argument; it is directly related to important policy debates. For example, some time in the early 1990s, a case was widely discussed in which an elderly woman on the outskirts of Chicago owned land worth some \$30 million. An endangered butterfly was found on her land, imposing serious restrictions on development and causing the value of the land to plummet. Many people argued that this was entirely unfair, and the government should be forced to compensate the woman for “taking” the value of her land. However, what was it that caused her land to be worth so much in the first place? The woman had owned the land for decades, during which time Chicago had grown considerably. Government-built roads, sewage, and electric grids had gradually expanded, making her formerly remote piece of land highly valuable. Government action created the value in her land, and a different government action subsequently reduced its private value through an effort to meet important public needs. This raises an important question: Are individuals entitled to wealth created by society or by nature, rather than through individual effort, or should this wealth belong to society as a whole?

In addition to the market failure associated with land values, there is another reason that market magic does not work with land. Land is present in a fixed amount, and supply is perfectly inelastic—it does not respond to changes in price. With fixed supply and increasing demand (as populations and wealth increase), land prices trend upward. Thus, whoever manages to acquire land will in general see the value of that land grow through no effort of her own. This makes land the subject of speculative investment. Land purchased for speculation is often left idle, but the demand for land for speculative purposes must be added to demand for land for productive purposes, driving up the price even further, and reducing the ability of people to buy land for production. In other words, under certain circumstances, speculative markets in land can reduce the production from land. None of this means that land ownership and land

¹⁷H. George, Significant Paragraphs from Henry George's *Progress and Poverty*, with an introduction by John Dewey. Garden City, NY: Doubleday, Doran, 1928.

markets are necessarily bad—it simply means that we should not automatically attribute all the theoretical virtues of markets to markets in land.

■ **SOLAR ENERGY**

Primarily for completeness, the final abiotic resource we will consider is the flux of a solar energy that continuously warms the Earth and turns its biogeochemical cycles. Obviously, neither human institution nor human invention (short of giant mirrors or umbrellas in space) can directly change the allocation of sunlight on the Earth, nor the supply. Indirectly, however, market forces can have significant impacts on the scale, distribution, and allocation of solar energy.

In terms of scale, human impacts on ecosystems appear to be degrading their capacity to capture solar energy—for example, forests sickened by acid rain capture less solar energy than healthy ones. In terms of distribution, land is an essential substrate for the capture of solar energy in most forms other than heat, and markets thus determine indirectly who can utilize the solar energy striking the Earth. Allocation—to what uses sunlight is put—is also determined in part by who owns the land it strikes. In this sense, solar energy is both rival and excludable. Also something of an allocation issue, land uses affect the spatial distribution of solar energy—recall how the evapotranspiration from the Amazon transports solar energy in the form of heat to the temperate zones. In terms of policy considerations, however, these issues can be treated as attributes or externalities of the use of other resources.

BIG IDEAS to remember

- | | |
|---|--|
| ■ Eightfold classification of natural goods and services:
fossil fuels (nonrenewable stock); minerals (partially recyclable, nonrenewable stock); water (nonrenewable stock, recyclable); solar energy (indestructible fund); Ricardian land (indestructible fund); renewable resources (renewable stock); ecosystem services (renewable fund); waste absorption capacity (renewable fund) | ■ Stocks (depletion), funds (depreciation)
■ Renewable, nonrenewable
■ Rival, nonrival, excludable, nonexcludable
■ Extraction cost, marginal extraction cost
■ External costs, marginal external costs
■ User cost, marginal user cost
■ Royalty
■ Hotelling rule
■ Price as problematic measure of resource scarcity
■ Henry George |
|---|--|



Market Failures and Biotic Resources

With the exception of solar energy and freshwater in the solar-powered hydrologic cycle, abiotic resources are nonrenewable on human time scales. Their value to humans consists solely in their extraction and use. Biotic resources, on the other hand, are renewable, if degradable, and are valuable as much for the services they provide as for any goods that can be derived from them. In this chapter, we examine whether or not specific natural resources meet the criteria for market allocation, turning our attention to biotic stock-flow resources (ecosystem structure) and ecological fund-service resources (ecosystem function), paying special attention to waste absorption capacity.

■ RENEWABLE RESOURCE STOCKS AND FLOWS

Renewable resource stocks and flows are rival and potentially excludable, depending on whether or not institutions exist that can regulate access to them. If depleted at a rate no faster than they regenerate, they are nonrival between generations. Unfortunately, unless we explicitly take future generations into account, economic incentives quite likely will lead us to deplete many of these resources faster than they can regenerate, and may eventually threaten them with extinction. What's more, as we have pointed out repeatedly, the use of renewable resource stocks and flows unavoidably deplete ecosystem funds and services as an “externality” of their production. This dramatically complicates economic analysis of these resources.

In our earlier discussion of renewable resource stocks and flows, we

looked at their physical properties with little if any discussion of the economics involved. Recall the sustainable yield curve from Chapter 6, here reproduced as Figure 12.1.¹ At first glance, it would appear that the goal of economists would simply be to make the resource as productive as possible. If this were the case, we should strive to maintain a population that produces the maximum sustainable yield, or MSY. However, this ignores two major issues. First, there are costs to harvesting, which we will call $P_E E$ (price of effort times effort, where effort includes all the resources required to harvest a stock), and these costs are likely to increase per unit harvested as the population in question grows smaller. Obviously, the smaller the population of fish that remains, the harder they are to catch. Even for forests, the most accessible timber will be harvested first, and as forest stocks decrease, it will cost more to bring the less accessible stocks to market.

Second, and even more important, if we were to consider all resources substitutable, as many economists do, and money is the perfect substitute for any resource, then the economic goal would not be to maximize the sustainable harvest of any specific resource, but rather the monetary sum of annual profits yielded by the resource. But even that is incomplete, as we will discuss shortly, for the market goal is in fact to maximize present value—the monetary sum of discounted future profits.

To simplify analysis, we can assume a linear relationship between effort, stock, and harvest, known as the **catch-per-unit-effort hypothesis**. For any given effort, more stock leads to a larger harvest in a linear fashion, and for any given stock, more effort yields a larger harvest. However, unless harvests are less than the annual increase in stock (i.e., below the sustainable yield curve), an increase in effort in any given year, all else being equal, implies a smaller stock and thus a smaller harvest from the same level of increased effort in the following years.²

Using simple algebra, we could assume that $Y = qXE$, where Y is harvest, X is stock, E is effort, and q is a constant we can think of as the “catchability coefficient.”³ As those of you at all familiar with math will recognize, this is the equation for a line starting at the origin with slope equal to effort, E . In Figure 12.1, we have drawn in lines $Y = qXE$, $Y = qXE'$ and $Y = qXE''$, where $E'' > E' > E$. Say the stock is Q'' in year 0, and effort is E . Our harvest then is Q . Of this harvest, $Q''Q'$ corresponds to the

¹We changed the scale of the figure to make it clearer. Specifically, the Y-axis has been stretched, so that the 45-degree line appears to be steeper than 45 degrees.

²When population stocks are large, a large harvest can lead to a larger sustainable yield in the following year, but this sustainable yield will still be smaller than the harvest required to reduce the population to that stock that provides that sustainable yield.

³C. Clark, *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*, New York: Wiley, 1990.

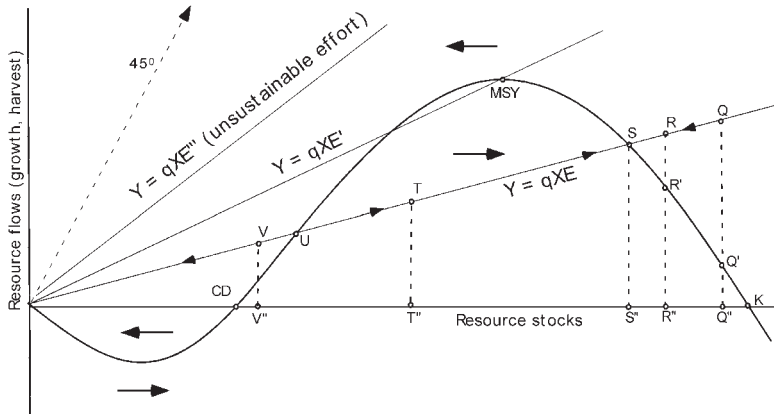


Figure 12.1 • The sustainable yield curve and catch-per-unit-effort curves. Recall from Chapter 6 that the curved line represents the rate of growth of stock for any level of population, which is the same as sustainable yield. The straight lines $Y = qXE$ represent the harvest for a given effort at any level of stock. A steeper slope indicates greater effort.

annual growth for that year, and $Q'Q$ must therefore reduce the stock, to R'' . (Note that this figure is not drawn with the same scale on both axes; i.e., resource flows are in smaller units than resource stocks.) At R'' and effort E , the harvest will be at point R , reducing the stock by $R'R$. As long as effort remains constant, this process continues until we reach stock S'' and point S on the sustainable yield curve.

Now imagine an external shock, such as El Niño, pushes populations down to T'' in a given year. As long as effort stays the same, the annual harvest will now be less than the growth increment or net recruitment, and the stock will gradually recover, until we again reach S'' . S is therefore a *stable equilibrium point*. However, if another El Niño year occurs before the fish population has recovered, it could push the population down to V'' . At V'' , the same effort will lead to a harvest greater than the growth increment, and the population will not recover. Recall from Chapter 6 that any harvest below the sustainable yield curve leads to a higher stock in the following year, and any harvest above the curve leads to a lower stock. The arrows on the $Y = qXE$ curve illustrate this dynamic. Thus, point U , where the catch-per-unit-effort curve intersects the sustainable yield curve, is an *unstable equilibrium*.

Maximizing Annual Profits

Suppose the goal is to maximize sustainable annual profits (π) from exploiting the fishery. This requires that we figure out where on the sustainable yield curve profits are maximized. Unfortunately our graphical analysis using the effort curve does not directly show profit. To analyze

the question from the perspective of annual profit is worthwhile and will require a somewhat different graph (Figure 12.2). The axes and the yield curve remain the same, except that we multiply the vertical (flow) axis by an assumed constant, P_F , the price of fish. This converts the yield curve into a total revenue (TR) curve without changing its shape, since we are multiplying by a constant. Since profit, π , is equal to $TR - TC$ (total cost), we need to add a TC curve. If we define effort as all the equipment, labor, and other resources that go into fishing, then TC is equal to the amount of effort times the price of effort, and therefore can be derived from a series of catch-per-unit-effort curves.⁴ We may think of TC as a curve starting at maximum population and rising to the left as more fish are caught. TC will increase as more fish are caught both as a result of stock depletion and as a result of harvesting a larger sustainable yield (at least up to MSY). But even beyond MSY the TC will probably still rise because stocks have become so sparse that the fish have become hard to find.⁵

Focusing on the flow dimension only, $\pi = TR - TC$, and maximum π occurs at π^* where $MR = MC$, where the slope of the tangent to the TR curve (MR) equals the slope of the tangent to the TC curve (MC). Remember that we have left the stock dimension out of our concept of total revenue—we are analyzing profit as a sustainable flow, not as the result of unsustainable stock reduction. Some stock reduction is necessary to arrive at the profit-maximizing stock, but ignore that for now; just assume for now that the stock-reducing, one-time catch of fish are thrown away. We'll deal with them later.

⁴Technical note on the derivation of Figure 12.2 from Figure 12.1: The catch-per-unit-effort curve ($Y = qXE$ in Figure 12.1) tells us the amount we can harvest from a given level of effort (E) at any level of stock. The TC curve tells us how much it costs to achieve any sustainable harvest. Theoretically there are two levels of stock (one level of stock when there is no critical depensation) where the harvest from a given constant effort is sustainable, and they are given by the intersection of the $Y = qXE$ curve with the sustainable yield curve. But in a world full of stochastic shocks and constant change, the unstable equilibrium will not be sustainable and can be ignored. The Y coordinate of the TC curve is therefore E^*E_P (where E_P = the constant price of effort), and the X coordinate is given by the stock at which the yield from a given $Y = qXE$ curve is sustainable, i.e., the stock at which $Y = qXE$ intercepts the sustainable yield curve. By letting E increase from zero to maximum sustainable effort, one can map out the TC curve. Total cost will be zero at K but rises until it reaches the maximum sustainable level of effort. Greater effort increases the sustainable yield until MSY is reached, at which point additional increases in effort lead to smaller and smaller sustainable yields. Eventually, increases in effort become unsustainable. At this point, the TC curve is truncated, and the dashed line in Figure 12.2 indicates unsustainable levels of effort. With the sustainable yield and constant effort curves of the type used in Figure 12.1, the TC curve in Figure 12.2 would not be a straight line; but a straight line facilitates the analysis and is no greater an abstraction from reality than any of the other curves depicted.

⁵Note that Figure 12.2 is not the same shape as Figure 12.1. Figure 12.2 depicts a more resilient population that can sustain higher levels of effort without causing extinction, to better illustrate the points we are making here.

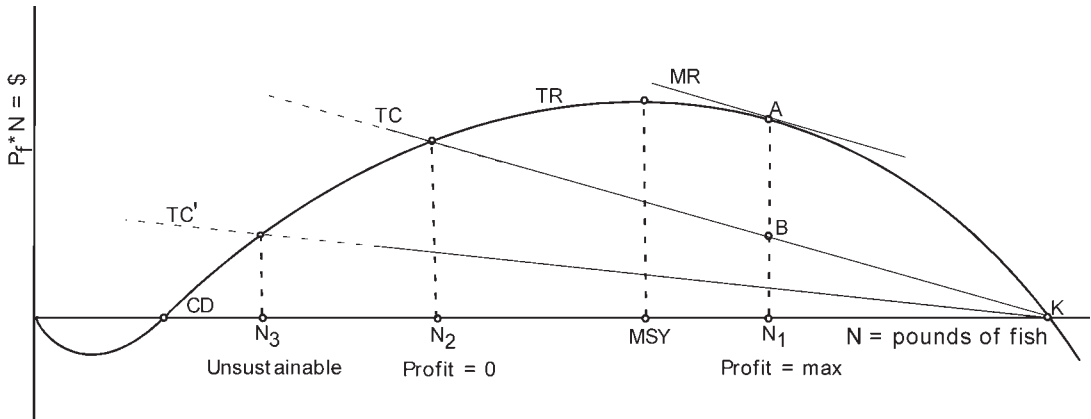


Figure 12.2 • Maximizing annual profit from renewable resources.

In Figure 12.2, we can see that the maximization of annual profit (AB) occurs at N_1 , which is larger than the stock corresponding to MSY. In other words, in this analysis the profit-maximizing capitalist will not even reach MSY, much less drive the population to extinction! If total costs were zero (or constant), marginal costs would be zero, and profit maximization would occur where marginal revenue was also zero, namely at MSY where the tangent to the TR curve is horizontal (slope equals zero). So even zero harvest cost would not lead the capitalist to exploit beyond MSY!

THINK ABOUT IT!

Can you find the tacit assumption responsible for the happy result that profit-maximizing exploitation does not require much stock depletion?

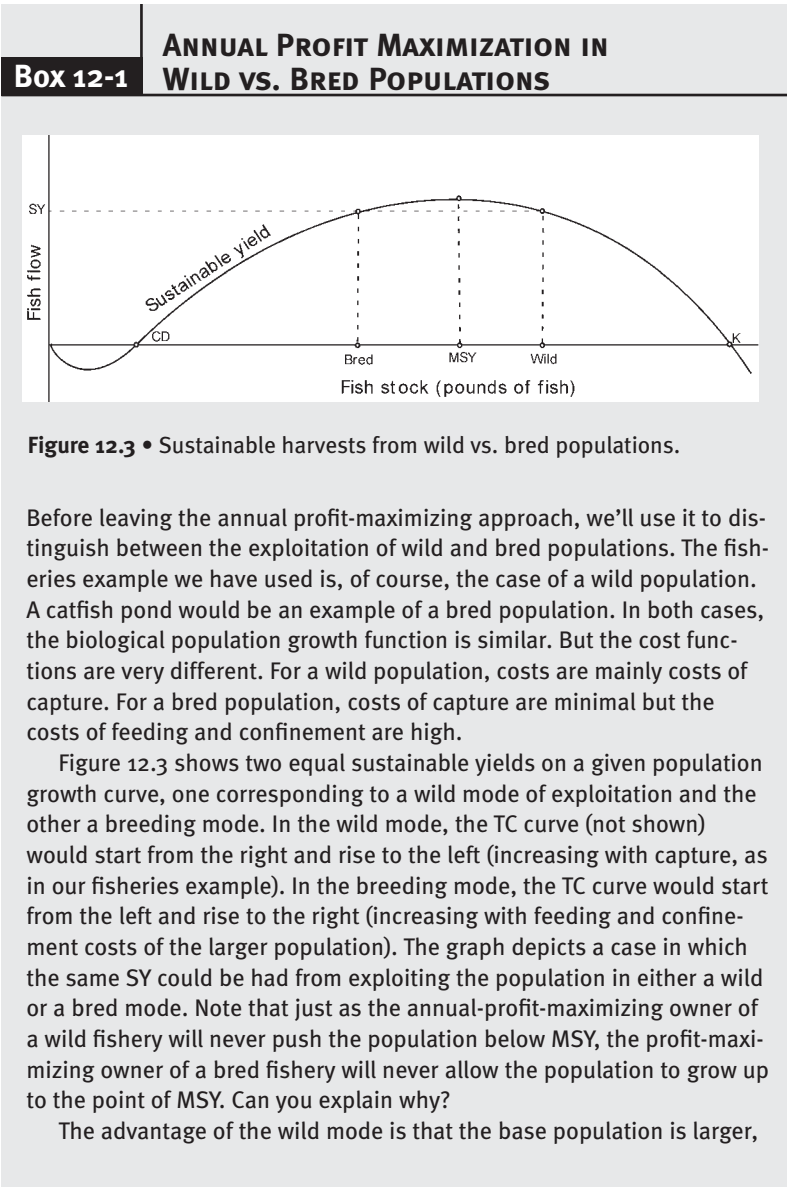
We have assumed a single capitalist exploiting the fishery, a single owner or decision maker. Instead of private property with excludability, suppose the fishery were open access, as most are. Under an open access regime (in which you'll recall the resource is nonexcludable), new fishermen will enter as long as there are profits to be made.⁶ New entrants will push the stock down to N_2 at which $\pi = \text{zero}$, or ($TR = TC$). At N_2 many more resources are going into fishing, but the sustainable catch is less than at N_1 , and no one is making a profit.⁷

Curve TC' depicts lower harvest costs, which might come about from a

⁶Profits, in economic parlance, are returns above and beyond the cost of production, where the cost of production obviously includes wages. For a small fishing crew sharing returns, profits would mean higher wages than they could find elsewhere.

⁷Note that if total costs are high enough, it is possible that the open access equilibrium might be at a greater yield than the profit maximizing equilibrium, though still with zero profit.

technological advance such as sonar devices for locating fish schools. At these lower costs, in an open access fishery it would be profitable for new fishermen to keep entering the fishery even after harvests become unsustainable. This is the case where the tragedy of open access resources may well lead to extinction, and may be a realistic depiction of what was happening with North Atlantic cod and many whale populations before regulation began.



containing more biodiversity, providing more ecological fund services, and feeding is free. The advantage of the bred mode is that the smaller base population takes up less ecological space, leaving more room for other life, wild or human. Of course the food for the captive population requires some ecological space as well as collection. Controlled breeding and genetic engineering would seem to require the breeding mode. Buregeoning human populations and the technical thrust toward genetic engineering push the shift toward bred genetically engineered populations (e.g., the famous 200-pound salmon). Also, the quest for higher biological growth rates to keep up with the interest rate favors the bred stock, since a smaller base population yielding the same annual increment obviously implies a higher rate of growth. But smaller populations growing more rapidly, and ever more dependent on human management for feeding, reproduction, and disease control, surely will increase the instability, brittleness, and vulnerability to uncertainty of the whole biotic system. Do you think the Federal Reserve Board considers these factors when setting interest rates? Should this risk be factored into interest rates, and by extension, discount rates?

Profit-Maximizing Harvest When Profits Can be Invested: Net Present Value

Returning to the higher-cost scenario (TC), suppose we make sure that there is a single owner, not open access. Are we then sure that we will end up at N_1 ? Unfortunately not, because of the troublesome issue we swept under the rug initially: What happens to those fish that are part of the stock reduction rather than the annual recruitment? They are not thrown away, and their number is large relative to the annual growth. Those fish are sold. Stock reduction fish have the advantage of being available now—you don't have to wait for them to be hatched and grow. But the more you reduce the stock of fish today, the fewer fish you will have tomorrow, and the more difficult it will be to catch those fish. The population of fish is like the proverbial goose that lays golden eggs in perpetuity. Surely no rational capitalist would kill such a productive goose.

Or would she?

If the capitalist wanted to maximize the sum of golden eggs from now till the end of time, then obviously she would not kill the goose. But the goose also has a liquidation value as a cooked goose. Suppose the capitalist could kill the goose, cook it and sell it for a sum of money, which when put in the bank at the going interest rate would yield an annual sum greater than the value of the golden eggs? Then it's goodbye goose, hello bank! The population growth rate of the goose (its egg-producing fecundity) is in direct competition with the interest rate, the "fecundity" of the money. Neoclassical economists argue that money itself may have no

reproductive organs, but it is a surrogate for many other things that can reproduce, and on average those other things can reproduce faster than the goose. So the goose-killing, reinvesting capitalist has converted a slow-growing asset into a fast-growing one, and we are all therefore better off. According to economists, cooking the goose in this case maximizes **net present value (NPV)**—the value to us today of all cost and benefit streams from now into the future. Economists calculate NPV by using a discount factor to give less weight to costs and benefits the farther in the future they occur (see Chapter 10).

Let's take the story a bit further in a thought experiment. Suppose an economy consists only of renewable resources. The interest rate is equal to some weighted average of the growth rates of all renewable resource populations. Everything that grows more slowly than the average (the interest rate) is a candidate for extinction (unless at some stock its growth rate rises above the interest rate). But something is always below average. When the below average is eliminated, what happens to the average in the next period? It goes up, of course. The tendency, it seems, would be to end up with only the fastest-growing species. Biodiversity would entirely disappear. In a world in which everything is fungible,⁸ that would not matter. We could all eat algae, if that were the fastest-growing species.

But we have forgotten prices. Surely prices would rise as particular slow-growing species became scarce and the rising price would compensate for low biological growth rate, so that the value of the species would grow at a rate equal to the rate of interest before it became extinct. Yes, but remember that when the price goes up, the price of the existing stock rises as well as the price of the flow of recruits. As the price increases, the incentive to liquidate the now more valuable remaining stock rises, along with the incentive to reduce current offtake to allow an increase in the more valuable new recruits.

The bluefin tuna is an excellent example of this argument. In 2001, a single 444-pound bluefin tuna sold in Japan for nearly \$175,000, or about \$395/lb. Although this was an anomaly, restaurants in Japan regularly pay up to \$110 per pound for bluefin tuna.⁹ Admittedly this occurs

⁸Something is fungible if one unit of it substitutes indifferently for another unit. For example, two buckets of water from the same well are fungible (you can't tell any difference between them). But two buckets of water from two different wells may not be fungible because of qualitative differences such as hardness, taste, etc. Money is fungible; we cannot tell if the money the government spends on foreign aid came from my tax dollars or from yours. Things convertible into money—goods, services, even biological species—acquire a kind of artificial or abstract fungibility, even though physically they are not at all fungible. This makes it easy to commit the fallacy of misplaced concreteness (see Chapter 2).

⁹G. Schaeffer, Tuna Sells for Record \$175,000, Associated Press International, January 5, 2001.

under a regime of imperfect property rights,¹⁰ but how confident are you that private ownership of the bluefin would solve the problem? The higher price means higher liquidation value, as well as higher future revenue, from new tuna.

How do we decide whether or not to harvest a (marginal) ton of fish that if left in the water reproduces (giving us a flow of golden eggs) and if harvested (giving us a cooked goose) yields interest on the profit?

The neoclassical approach is to ask: What are all the opportunity costs of harvesting the fish today? Obviously, if we harvest the fish today, that same ton of fish is not there to harvest tomorrow. Unlike oil or iron, however, we lose not only the opportunity to harvest that ton of fish, but also the offspring those fish would have had, as well as the increase in biomass they would have experienced, if left in the water. In addition, as the economy grows, people demand more fish, and growing human populations further increase the demand for fish. Greater demand means a higher price, and we lose that additional profit if we harvest today. Moreover, if we leave the fish in the water, they will reproduce, and the greater population means that it will be cheaper and hence more profitable to catch a ton of fish next period than this period.

In contrast, the benefit of catching the fish this period rather than next is that the profits from their harvest can be invested; that is, the opportunity cost of not catching the fish is the money forgone from not being able to invest the profits from that ton of fish between this period and the next. As we harvest more and more fish, the growth rate (the annual increase in biomass/total biomass) increases, so the marginal costs of harvesting now are increasing.

The economist will therefore favor harvesting as long as the marginal benefits of catching the next ton of fish are greater than the marginal costs, and will stop when they are equal. The tricky thing about the decision is that when we consume more stock today in exchange for less stock and less yield tomorrow, it is not just one tomorrow, but tomorrow and tomorrow and tomorrow in perpetuity. We must compare a one-time benefit with a perpetual loss. As we mentioned above, economists address this problem rather unsatisfactorily by the financial convention of discounting and present value maximization. What does all of this mean in practical terms?

Compared to our static analysis when we ignored the stock reduction

¹⁰Fishing of bluefin tuna in the Atlantic at least is regulated, with quotas for the Eastern Atlantic held by European countries, and quotas in the Western Atlantic held by the U.S., Canada, and Japan. But evidence indicates that these quotas are too high, and that the two populations are not even distinct. Thus, regulations designed to assign property rights to a formerly open access resource may fail in their objective of preserving the species. See T. Bestor, *How Sushi Went Global*, *Foreign Policy* 121 (November/December 2000).

required to reach the annual profit-maximizing equilibrium, the opportunity to invest profits from stock reduction will lead to a lower stock of fish (or any other renewable resource). In terms of Figure 12.2, the profit-maximizing harvest will be to the left of N_1 , and if the interest rate is high enough relative to the growth rate of the species, it may even be to the left of MSY.

Take the case of bluefin tuna, where the cost of capture of one fish may be a negligible portion of its sales price. Imagine current harvests were sustainable and in the vicinity of MSY, and there was a single resource owner intent on maximizing profits. The reduction in sustainable yield from MSY to a somewhat lower stock may be small, while the stock liquidation required to get there is still quite large. If the interest payments on the profits from the sale of that liquidated stock are greater than the value of the lost annual yield, then profit maximization favors sustainable harvest at a stock lower than MSY, and closer to the point of critical depensation.

What are the implications of this scenario in an extreme but not at all unrealistic case where harvest costs ($P_E E$) are negligible compared to harvest revenue ($P_Y Y$), even for very low-resource stocks? Timber is a good example. Say that a forest of redwoods not yet biologically mature¹¹ increases in size and thus value by 3.5% per year.¹² In contrast, the average real growth rate of money on the U.S. stock market over the last 70 years was about 7%.¹³ Clearly, it makes economic sense to harvest the resource now and invest the profits in the bank. In fact, for any species that is relatively inexpensive to harvest and grows more slowly than alternative investments, it makes economic sense to harvest the species to extinction. In general, averaged over the time it takes to reach harvest size, many valuable species grow quite slowly relative to alternative investments, and technology tends to reduce unit harvest costs over time. And for such resources it is, once again, goodbye golden eggs, hello bank.

In summary, the advantage of the catch-per-unit-effort curve analysis is that it builds in from the beginning the stock reduction effect and in that way is more realistic. The advantage of the TR–TC diagram is that it shows

¹¹Biological maturity occurs when growth rates for the forest taper off toward zero; i.e., new growth is just matched by rates of decay.

¹²This is actually an unrealistically high rate of growth. Data exists on a 1-acre plot of redwoods that has been monitored for over 70 years. Though the rate of growth on this plot is so high that it is widely known as the “wonder plot,” the most rapid 10-year mean annual increment in total stand volume was 3.5%. The mean annual increment in growth from 1923 to 1995 is well under 1%. These figures were calculated by the authors from data provided in G. Allen, J. Lindquist, J. Melo, and J. Stuart, *Seventy-Two Years Growth on a Redwood Sample Plot: The Wonder Plot Revisited* (no date); Online: <http://www.cnr.berkeley.edu/~jleblanc/WWW/Redwood/redwd-Seventy-.html>.

¹³S. Johnson, *Are Seven Percent Returns Realistic?* Online: <http://www.sscommonsense.org/page04.html>. Common Sense on Social Security.

that annual profit maximization can be sustainable and efficient in the absence of open access, and with some limits on the biologically blind financial logic of discounting and present value maximization.

RENEWABLE RESOURCE FUNDS AND SERVICES

The analysis of optimal harvests of renewable resources so far has only treated them as stocks and flows of raw materials. But, as we discussed in Chapter 6, renewable resources are also funds that provide ecosystem services, and we cannot ignore one when deciding how best to allocate the other. While natural resource stocks and flows have some characteristics of market goods, the services generated by funds typically do not. Such services are generally nonexcludable and for many, no feasible institutions or technologies could make them excludable. Thus, free markets will not produce them. They are also nonrival and noncongestible, and selling them in the market would not equate marginal costs with marginal benefits.

Treating the destruction of ecosystem services as a negative externality of aggregate economic production offers some insights into optimal harvest levels. Returning to the analysis of renewable resource stock-flows, we would need to add all external costs to total private harvest costs. If marginal external costs increased linearly as the fund-stock is reduced in size, then total external costs would increase exponentially. In reality, marginal external costs are likely to increase at a greater than linear rate as we near an ecological threshold, such as the critical depensation point, which will inevitably be reached when harvest effort is too great to be sustained. Figure 12.4 is similar to Figure 12.2 but relabels the TC curve as the total private cost (TPC) curve and includes a total social cost (TSC) curve that adds external costs to the TPC. The TSC curve approaches the vertical

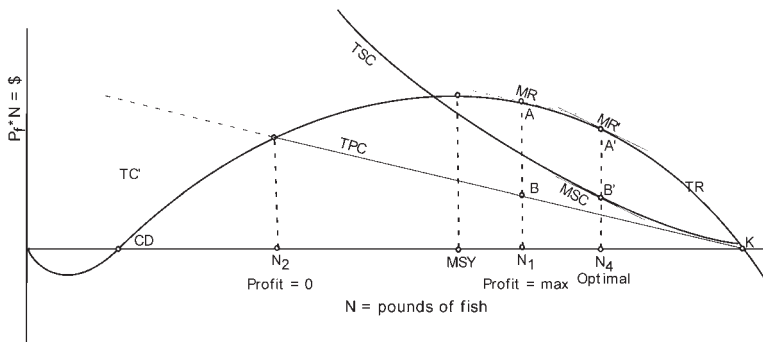


Figure 12.4 • Optimal harvest of renewable resources when accounting for ecosystem services.

(i.e., unacceptably high marginal social costs) as we approach the point of critical depensation. The optimal harvest is where marginal social costs equal marginal revenue, labeled N_4 on the graph. Whenever the renewable resource contributes to the provision of ecosystem services, the optimal harvest from an ecological economic perspective will always be at a higher stock with lower private costs than in the annual profit-maximizing equilibrium.

Of course, optimality would require the micro-level internalization into prices of all ecosystem services. Yet human impacts on these services are characterized at best by uncertainty (we know the possible outcomes of damage to ecosystem funds on ecosystem services, but don't know the probabilities) and more often than not by ignorance (we don't even know the range of possible outcomes). In fact, we almost certainly do not know the full extent of the ecosystem services from which we benefit. In addition, the value of all externalities would need to be worked out by economists, ecologists, and others, and incorporated into the prices of the goods that generate the externalities. And, of course, the marginal value of an ecosystem service changes along with the quantity of the ecosystem service supplied, so the value of externalities would be constantly changing. As we have pointed out, all economic production incurs externalities. The notion of calculating the constantly changing values of all externalities for all goods would be a Promethean task. Once achieved, it would still require some institution to incorporate the fees into market prices. And we must remember that the magic of the market is precisely its unplanned, decentralized nature, and its ability to utilize "knowledge not given to anyone in its totality."

Effectively internalizing externalities, in contrast, requires precisely the opposite—centralized planning by individuals provided with knowledge in its totality. While an optimal allocation of everything is not a feasible goal, in Part VI on policy (Chapters 20–23), we will explore approaches to achieving a satisfactory allocation.

Note also that we have again left out the potential for investing the profits from reducing the stock. This is intentional. No potential return could substitute for either life-sustaining ecosystem services or the raw materials essential for all economic activities. In addition, many investments are profitable precisely because they do not account for the opportunity costs of resource depletion (MUC) or the social costs of the ecosystem services inevitably degraded or destroyed through resource extraction.

Nonmarket ecosystem fund-services simply cannot be converted to money and invested, like we can do with a cooked goose. Also, natural resources are growing physically scarcer. Technology seems to be developing new uses for most natural resources faster than it develops substitutes, which increases future demand. Increasing demand and decreasing sup-

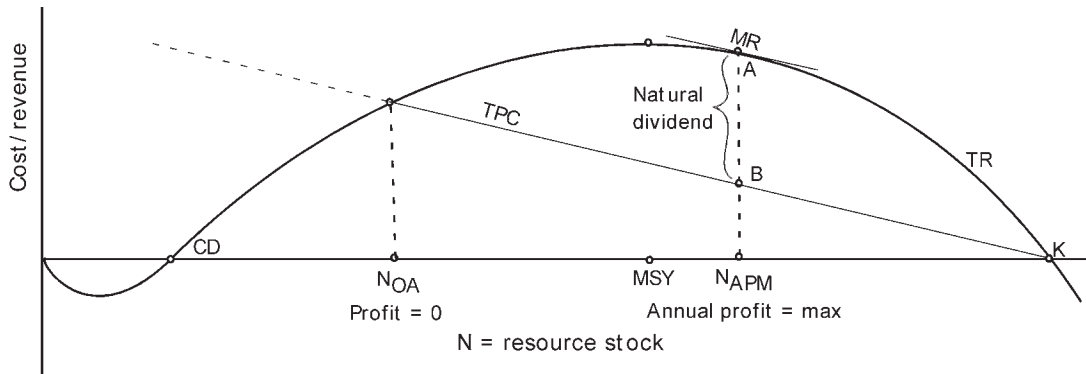


Figure 12.6 • The natural dividend from renewable resource harvests.

The *natural dividend* is the unearned income from the harvest of renewable resources. As nature and not human industry produces renewable resources, all profits above “normal” profit (included in TC) are unearned, and the natural dividend is equivalent to $TR - TC$.

The Natural Dividend From Renewable Resources

Unearned income is the amount above and beyond what is required to bring a resource to market. In the case of nonrenewables, the unearned income was called scarcity rent. In the case of renewables, there can also be an unearned income deriving from nature’s reproductive capacity, which we propose to call the **natural dividend**. Extraction and harvesting impose real costs, and those engaged in these activities are earning a legitimate income. This is included in the TC curve as “normal profit,” which is the opportunity cost of the owner’s labor and capital. Profit above normal profit (e.g., AB) is called “pure economic profit.” Since it is beyond the opportunity cost of the owner, we think of it as an unearned growth dividend from the reproductive power of nature. We have already seen that in an open access equilibrium, total costs are just equal to total revenues, and profits are zero (at stock N_{OA} in Figure 12.6). However, if there is a single owner of the resource, say, a government agency, the owner can limit harvest to the profit maximizing level, stock N_1 . The pure profits in this case (AB in Figure 12.6) are the natural dividend and arise not from special abilities of the particular owner, but from the reproductive powers of nature. The natural dividend can also be thought of as the value of the resource in the ground (or sea). While the natural dividend typically accrues to the owner of the resource, it is a purely unearned profit from production. To whom should the dividend belong? That is a political decision, more in the realm of fair distribution than efficient allocation.

■ WASTE ABSORPTION CAPACITY

Waste absorption capacity is really just another ecosystem service. We treat it separately here because it is extremely important and because it has

different characteristics from most other ecosystem services. Waste absorption capacity is the ability of the ecosystem to absorb and process pollution, and the economics of pollution is the predominant focus of neoclassical environmental economics. As we pointed out previously, waste absorption is a rival good. If I dispose of my sewage in a wetland, there is less capacity subsequently for that wetland to process someone else's wastes. As we also pointed out, many countries are trying to create institutions that make waste absorption capacity an excludable good. These can range from regulations that directly limit industrial emissions to mandatory catalytic converters in cars to tradable emissions permits for sulfur oxides. Tradable permits and quotas for pollution essentially make waste absorption capacity a private good. These mechanisms will be discussed at length in Chapter 21.

However, we must bear in mind that pollution is pure externality and a **public bad**, which is something that is nonrival, nonexcludable, and undesirable. Therefore, even when a market exists in air pollution, for example, this is not the same as saying that a market exists for clean air. Nonetheless, pollution permits are one of several mechanisms that can help achieve a socially optimal level of pollution (see Chapters 20–22).

There is, of course, no direct social benefit to pollution *per se*, but as we have repeatedly stated, all productive processes generate some pollution, and if we prohibit all pollution, we virtually prohibit production. This is why economists use the apparent oxymoron of “optimal pollution.” By optimal, economists simply mean potentially Pareto efficient. A reasonable estimate of the benefits of pollution is therefore the marginal net private benefits of production (MNPB) associated with a unit of pollution. The problem is, of course, that our knowledge of external costs of pollution is characterized predominantly by ignorance and uncertainty, to a lesser extent by risk, and to a minimal extent by certainty. Since we do not know the full social costs of pollution, it is exceedingly difficult to balance costs with benefits. Policy makers are also not well informed concerning the MNPB of pollution to polluters.

We have to recognize that waste absorption capacity is a dynamic process, and we must define carefully what we mean by it. We define **waste absorption capacity** as the ability of an ecosystem to assimilate a given flow of waste. If the waste flow exceeds the waste absorption capacity, then waste will accumulate. As waste accumulates, the ecosystem becomes less able to assimilate it, leading to even more rapid accumulation. Thus, in a dynamic analysis, if the flow of waste into an ecosystem exceeds the ability of the ecosystem to assimilate that waste, then waste will build up indefinitely until virtually all services generated by the biotic processes in that ecosystem collapse. The system will thereafter be able to process only negligible amounts of waste.

What is the marginal external cost of the increase in flow rate that takes the system beyond the point of no return? It is the value of the lost services from that ecosystem for all time. If the ecosystem provides vital functions, the marginal external cost is basically infinite. However, if the waste flow is halted before collapse of the ecosystem, the system can slowly process the waste and restore itself. This is basically what is happening in Lake Erie on the U.S.–Canada border. If the ecosystem in question is simply a local ecosystem with similar ecosystems nearby, even after collapse, stopping the waste flow can lead to recovery. Wastes will be absorbed, dissipate, and settle out of the system, new organisms will colonize the system, and the restoration process will begin.

It is worth illustrating these points graphically. Figure 12.7 is an appropriately modified version of the analysis of externalities from Figure 10.2. Note that economic output and waste output are measured on the same axis in recognition of the laws of thermodynamics. In reality, the relationship is not as fixed as Figure 12.7 indicates; many technologies are available for producing different goods, some of which generate less pollution than others, though for any given technology, the relationship will be fixed.

We assume here that the economic output under consideration is not essential—either substitute products and processes are readily available, or the good itself is simply not that important to human quality of life. Chlorine-bleached paper from wood pulp is a good example. Unbleached paper from kenaf or hemp are excellent substitutes, and paper itself is important, but not essential to life. Among the many wastes emitted by paper mills are organochlorines, and paper mills are the largest emitter of organochlorines into the water supply in the U.S. and numerous other countries as well.

Organochlorines resist biodegradation, and hence the waste absorption capacity for these substances is quite small; it is indicated by the perpendicular at point Q_A, W_A in Figure 12.7. Organochlorines include some of the most toxic substances known, such as dioxins, which readily accumulate in the environment and in animals, including humans. Health problems associated with dioxins include cancer, immune system disorders, and developmental problems in children and fetuses. We assume without too much exaggeration that surpassing the waste absorption capacity for organochlorines for extended periods could make the affected areas essentially uninhabitable for humans, and on our crowded planet, this is an unacceptably high cost. This is indicated on the graph by the marginal external cost (MEC) curve, which approaches vertical as we near the waste absorption capacity of the environment ($Q_A W_A$). These nearly infinite marginal costs occur only if the accumulation of pulp mill waste continues unabated for some time.

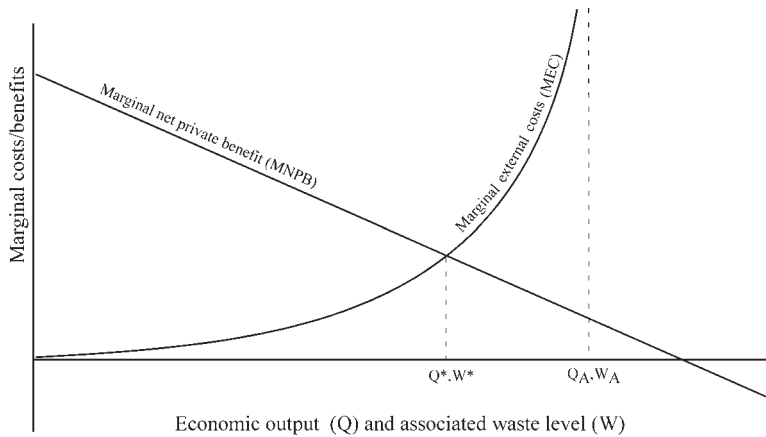


Figure 12.7 • Waste absorption capacity: marginal costs and benefits of pollution.

On the other hand, the extreme levels of uncertainty and ignorance concerning waste absorption capacity and the impacts of paper mill pollutants are not reflected in this graph. In reality, both the MEC curve and the line depicting waste absorption capacity should be thick smears instead of the fine lines depicted. We have labeled the optimal level of output/pollution as Q^*, W^* , but again caution that with our given state of knowledge, this is a broad range, not a precise point. The reasoning behind the location of Q^*, W^* was described in Chapter 10.

We cannot stress enough the importance of looking at pollution flows and waste absorption capacities as dynamic. Some economists have argued that pollution causes zero damage before reaching point Q_A, W_A , because the ecosystem is capable of assimilating the waste. But there clearly are substantial costs to pollution even when the ecosystem is capable of assimilating them. MEC may approach infinity at Q_A, W_A .

Depending on the population level and the level of economic activity (i.e., based on scale), MNPB may become zero before or after reaching the waste absorption capacity of the environment. The more full our planet becomes—the larger our scale—the more likely that MNPB will still be positive when we reach the waste absorption capacity and MEC approaches the vertical. Why? Goods and services are characterized by diminishing marginal utility. This means that for a given amount of goods, more people imply lower per-capita consumption, and hence higher utility from each unit consumed. This shifts the MNPB curve upward.

Finally, we must stress once again here that even if policy makers could measure the full marginal costs and benefits of pollution and set the number of permits accordingly, pollution markets would still fail to generate all the wonderful properties associated with the free market. Different individuals obviously have different preferences (utilities) with respect to polluted environments. Markets are widely extolled because they allow the individual to choose what she produces and consumes so that her marginal benefits from either are exactly equal to her marginal costs. Pollution, however, affects public goods, and all individuals must consume the same amount. It would be impossibly complex to create a system in which each individual was paid by the polluter according to his or her own dislike of pollution. This by no means implies we are opposed to markets in waste absorption capacity, but it does mean that we should not associate with them all the market virtues associated with the buying and selling of market goods.

■ BIOTIC AND ABIOTIC RESOURCES: THE WHOLE SYSTEM

To achieve a sustainable, just, and efficient economic system, we clearly must understand the nature of the resources upon which that system depends. We need to understand the role these resources play in meeting the needs of humans and other species on this planet, and the characteristics that affect their allocation within and between generations via market and nonmarket mechanisms. It would, of course, be impossible to analyze every individual resource. Instead, we introduced the important concepts of rivalness, excludability, externalities, ignorance and uncertainty, and stock-flow and fund-services, and applied these concepts to specific categories of natural resources. To facilitate this, we created a rough taxonomy of biotic and abiotic resources subdivided into eight categories, and applied the above concepts to each of these categories.

Our first goal with this approach was to help you understand precisely why markets fail to efficiently allocate each individual resource and start you thinking about what types of institutions and mechanisms might work better. We began with abiotic resources, which are fairly simple to understand. We then moved on to the stock-flows provided by nature, the raw materials on which the economy depends. We began to see the emergence of complexity: unpredictable ecological thresholds beyond which a population will collapse, impacts from outside variables such as climate change and habitat degradation. The analysis grew a bit more complicated. Once we turned to ecosystem services and waste absorption capacity, it became obvious that these were elements of a whole system and could not be understood apart from that system. Ecosystem fund-services, including waste absorption capacity, are provided by the complex interaction of ecosystem stock-flows, and are necessary to sustain those stock-flows. We can't think

of allocating the stock-flow independently of the fund-service. Both fund-service and stock-flow are seriously affected by the waste flows from non-renewable abiotic resources. So what does this mean?

Our second goal with this approach was, paradoxically, to guide you toward the conclusion that the first goal is insufficient. While it helps to understand the particular characteristics of each individual natural resource, it is more important by far to recognize that these resources are so intimately intertwined that we cannot allocate any one resource without considering the impact it will have on others. The reductionist approach (breaking down low-entropy resources into relatively narrow categories) is important, but inadequate on a complex, living planet. Seemingly efficient allocation of each resource individually will not necessarily lead to the efficient allocation of all resources together. Ecological economics is concerned with integrated systems, not individual commodities, and with complex societies, not atomistic individuals. Breaking a system down to better understand its individual components is a useful analytic tool, but it can seriously mislead us unless we subsequently synthesize these components into an integrated understanding of the whole.

BIG IDEAS to remember

-
- | | |
|--|--|
| ■ Stock-flow versus fund-service resources | ■ Maximizing annual profit versus maximizing net present value |
| ■ Sustainable yield, maximum sustainable yield | ■ Exploitation of wild versus bred populations |
| ■ Absorptive capacity | ■ Natural dividend |
| ■ Per-unit effort curve | |
| ■ Stable and unstable equilibrium | |
-



Conclusions to Part III

In Chapters 8 and 9, we provided a brief introduction to market microeconomic theory, showing how spontaneous order can arise naturally from the rational, uncoordinated actions of millions of individuals. We then described how this spontaneous order can lead producers to efficiently allocate raw materials, labor, and capital toward their most profitable uses, while simultaneously allowing consumers to efficiently allocate their resources toward consumption in a manner that maximizes well-being. However, in Chapter 10 we extended this analysis to show how the occurrence of these optimal outcomes depends entirely on the specific characteristics of the goods and services we wish to allocate. In Chapters 11 and 12, we examined the natural resources upon which all economic production depends to see if they met the criteria for market allocation. Unfortunately, none of the goods and services provided by nature had all of the characteristics required for efficient market allocation. In subsequent chapters, we will build on this analysis to propose policies and institutions better suited to allocating the economic means provided by nature toward the desirable end of a high quality of life for this and future generations. Before we turn to this task, however, we will first examine the macroeconomy, Part IV of our text.

PART IV

Macroeconomics





Macroeconomic Concepts: GNP and Welfare

Microeconomics focuses on how markets function. It is useful for analyzing markets, with the aim of ensuring that they operate efficiently. Beyond that, however, it steers clear of policy recommendations. In essence, it assumes that the best policy is to let the market do its thing without interference. Macroeconomics looks at the economy as a whole, at the national or global level. In contrast to microeconomics, macroeconomics more often recognizes the importance of policy interventions, especially fiscal policy (government spending and taxation) and monetary policy (money supply and interest rates).¹ These policy interventions are important. However, the making of policy implies a goal. The traditional goal for macroeconomic policy is stable market-driven economic growth without limit, and to a lesser extent full employment. But unlimited economic growth is impossible. Many of the scarcest resources are nonmarket goods and services, and many of the most serious problems we now face extend beyond the borders of the nation-state.

In ecological economics, optimal scale replaces growth as a goal, followed by fair distribution. Traditional macroeconomics generally leaves allocation to market forces at the microeconomic level. Ecological economics more often recognizes that markets are inadequate for allocating

¹There are many schools of macroeconomics, some of which do not call for policy interventions. For example, new classical macroeconomics, also known as rational expectations theory, argues that policy interventions are ineffective. The monetarists argue that policy interventions can be counterproductive. Both schools are quite conservative, favoring small, weak government. In practice, however, policy makers do use macroeconomic policy in efforts to attain policy objectives.

many scarce resources, and policy interventions are necessary to supply adequate quantities of these nonmarket goods. These different goals of ecological economics will favor different uses of traditional policies, and also suggest an array of alternative policy interventions.

In Chapters 13–16, we provide a brief introduction to some of the concepts, issues, and policy tools of mainstream macroeconomics and apply them to the policy goals of ecological economics. Remembering the circular flow diagram, we recall that macroeconomics deals with the aggregate flows of national product and income (the real sector). It also deals with the aggregate money supply, and with interest rates (the monetary sector). Following a discussion of the relationship of macroeconomics to microeconomics, this chapter will first look at the aggregate measure of the real sector, gross national product. Then we will have a look at money and the aggregate monetary sector, followed by a consideration of welfare indices other than GNP.

In Chapter 14, we will look at the medium by which wealth is measured in conventional economics: money. We then focus on distribution in Chapter 15. In Chapter 16, we will present the basic macroeconomic model for combining the two sectors into a simple general equilibrium model of the economy. This model (the IS-LM model) shows how the behavior of savers and investors in the real sector interacts with the behavior of the monetary authority (usually a national central bank, such as the Federal Reserve bank in the U.S.) and the money-holding public to determine the interest rate and the level of national income and employment. We will show how the goals of ecological economics lead to different policy recommendations than those supported by the mainstream. We will then discuss the possibilities for extending the IS-LM model to incorporate ecological constraints.

■ A TROUBLED MARRIAGE

Microeconomics developed historically prior to macroeconomics. Indeed, if we understand the behavior of the decision-making units—firms and households—and how competitive markets work, then we will come to accept Adam Smith’s “invisible hand”—that individuals seeking only their own benefit will automatically serve the common good. The microeconomic search for maximum private benefit will automatically result in the large-scale consequence of the greatest public welfare of all, or so it was thought. There was no need for special consideration of the macroeconomic picture, since the invisible hand would guarantee that if the microeconomics are right, then the macroeconomic picture will be right. Monopoly can ruin this nice result (as we showed in Chapter 8), so mar-

kets must be kept competitive, but that is about the only collective action needed.

In Chapter 2, we met Say's Law (supply creates its own demand), which guarantees that production always generates sufficient aggregate income to purchase aggregate production. Therefore, there can be no general glut of all products—at worst an imbalance in the mix of products, a misallocation, too much of something, too little of something else. That misallocation will soon be corrected by relative price changes in competitive markets. The same applies to the labor market—if there is unemployment (a surplus of labor), it simply means real wages are too high, not that there is a problem on a larger scale. If unemployment persists, you just need to let wages fall some more.

This view lasted well into the Great Depression of the 1930s and still has its adherents today. But under the leadership of John Maynard Keynes, economists began to think that prolonged unemployment, though theoretically impossible, was sufficiently real to warrant rethinking the theory. This rethinking led to the discovery of the leakages and injections from the circular flow, and the problem of making sure that total injections equal total leakages. We considered this in Chapter 2.

In addition, economists remembered the **fallacy of composition**, the false belief that whatever is true for the part must be true for the whole, or vice versa. For example, one spectator in a football stadium can get a better view by standing up. But all spectators cannot. If all stand on tiptoe, then no one has a better view than when everyone was comfortably seated. Similarly, one country can have a surplus or deficit in its balance of payments. But for the world as a whole, neither surplus nor deficit is possible because the sum of all exports must identically equal the sum of all imports. One worker may gain employment by being willing to work for a lower wage, but all workers probably could not, because lower wages for everyone means less income for the majority of the people, which means less spending on goods and services, and less demand for labor even at the lower wage. Reduced spending leads to reduced investment, which further lowers aggregate demand. In addition, any individual can easily convert his money holdings into real assets, but the community as a whole cannot, because when everyone tries to exchange money for real assets, someone has to end up holding the money.

There were, in sum, ample reasons to begin to develop a “macroeconomics” to deal with aggregate phenomena, especially unemployment and inflation. But contrary to what one might expect, the new macroeconomics did not build on the foundations of microeconomics. The macroeconomy is the aggregate of all the micro units, but macroeconomics is not just microeconomics aggregated. If it were, then we would be back to the invisible hand and the conclusion that macroeconomics was not necessary.

The entire economy described in microeconomic terms is the general equilibrium model. In it, all supply-and-demand relations in all markets are presented as one great interdependent system of simultaneous equations—say, a million equations in a million unknowns. This system of equations is solved by the market—a giant social computer that works by trial-and-error iteration. Economists have devoted much effort to counting equations and unknowns and making sure they were equal so that the general system, at least theoretically, could be solved. While the general equilibrium model is enlightening and conceptually satisfying, it is not very helpful from a policy perspective simply to know that everything depends on everything else. Policy needs a few leverage points at which to influence the gross behavior of the big system in its most important aspects. That is what macroeconomics has sought—simple models of the economy in terms of key aggregate variables, such as the money supply, aggregate price level, the interest rate, aggregate consumption and investment, exports and imports. And, of course, the biggest goal and leverage point of all—the rate of growth of GNP.

Ecological economics challenges today's standard emphasis on growth. Growth, yesterday's panacea, is rapidly becoming today's pandemic. Growth was a panacea because it was thought to be the solution to the macroeconomic problems of overpopulation, inequitable distribution, and involuntary unemployment. Microeconomists do not have much to say about growth, although not many would oppose it. Microeconomics is dominated by the concept of optimum and its associated "when to stop rule." As we argued in Chapter 2, if the macroeconomy is a part rather than the whole, then the logic of microeconomic optimizing applies, and at some point people trained in microeconomics will have to ask the macroeconomist, What is the optimal scale beyond which this economic subsystem should not grow? And when growth becomes uneconomic, as it must once we are at the optimum, then how are we going to deal with overpopulation, inequitable distribution, and involuntary unemployment?

It is the job of ecological economists to think about that: What happens after we reach the optimal scale, and how do we return there if we accidentally surpass it?

The whole is something more than the sum of its parts. In recognizing this, ecological economics bridges micro- and macroeconomics, though the exact relation between the two remains a bit mysterious. For our purposes, the relation between macro and micro is that shown in the circular flow diagram, Figure 2.4, repeated here as Figure 13.1.

The firms and households are our focus of attention in microeconomics. The firm as producing unit decides a supply plan for goods and a de-

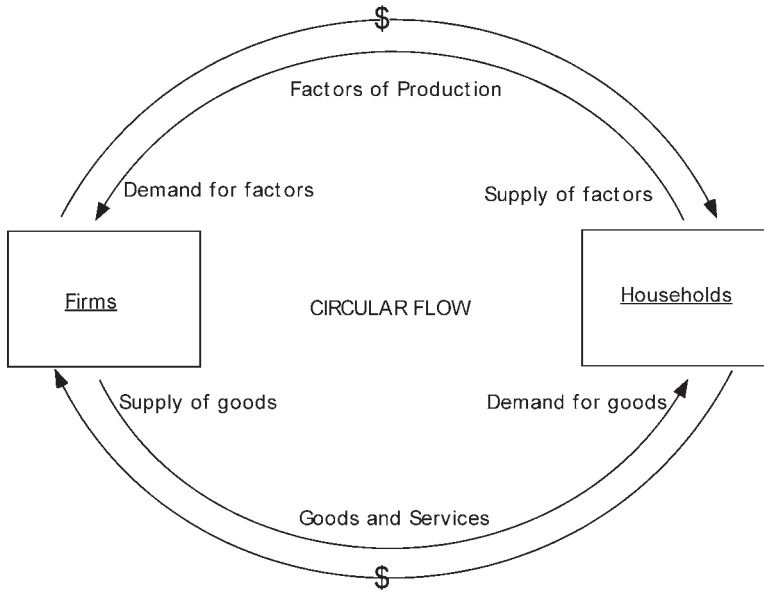


Figure 13.1 • The circular flow of the economy.

demand plan for factors. The household as consuming unit decides a demand plan for goods and a supply plan for factors. Microeconomics deals with these supply-and-demand decisions and their interactions in markets to determine the price and quantity of goods and factors exchanged in the markets. Because of its focus on prices, microeconomics is often called “price theory.” Macroeconomics deals with the total volume of aggregate goods and services flowing through the goods market (national product), and the total volume of factors flowing through the factors market (national income). Because of its focus on aggregate income, macroeconomics is sometimes called “income theory.”

Although we will speak of production and consumption because these terms are well established, it is important to remember (from earlier chapters), that in a *physical* sense, there is neither production nor consumption—only transformation. Raw materials are transformed into useful things (and waste) by “production.” Useful things are transformed into waste by “consumption.” What we are producing and consuming are “utilities”—useful temporary arrangements of matter and energy that serve our purposes. The throughput remains fundamental in both micro- and macroeconomics, even though it is not explicit in the accounts of firms and households, or in the aggregate accounts of nations. And the throughput is governed by the First and Second Laws of Thermodynamics, not by circular flow accounting conventions.

■ GROSS NATIONAL PRODUCT

Because economic growth is the paramount goal of nations, it is important to know just how it is measured. Growth in *what*, exactly? Economic growth is measured as growth in GNP, gross national product (or GDP, gross domestic product).²

As previously discussed in terms of the circular flow diagram, we have two measures of the aggregate circular flow that give the same number—national product and national income. Sometimes they are called national product at consumer goods prices (lower loop in Figure 13.1) and national product at factor prices (upper loop in Figure 13.1), or national income. Let's focus first on the lower loop, national product proper at consumer goods prices.

Gross national product (GNP) is the market value of final goods and services purchased by households, by government, and by foreigners (net of what we purchase from them), in the current year.

In this measure, **gross national product (GNP)** is the market value of final goods and services purchased by households, by government, and by foreigners (net of what we purchase from them), in the current year. With a few exceptions, anything not purchased this year is not counted.³ Household production for the household itself is not sold and thus not counted—cooking, cleaning, childcare, and so on are omitted, unless done by a paid domestic helper. Intermediate transactions among firms are not counted. Only the sale of the final product to the household is counted. The wheat sold by the farmer to the miller is not counted, the flour sold by the miller to the baker is not counted; only the bread sold by the baker to the household for final consumption is counted. The value of the bread is the sum of the values added by the farmer, by the miller, and by the baker. Values added to what? To the basic natural resource—the wheat seed, the soil, the rain, the sunlight, and so on. The basic natural resources in most cases are considered to be free. Therefore, GNP is the sum of value added. It does not include any attribution of value to that to which the value was added. What is it that adds value to free natural resources? The transforming services of labor and capital funds.

Note that these accounting conventions are consistent with the neo-classical production function discussed in Chapter 9—namely, that production is a function of labor and capital only.⁴ The exchange of existing

²The difference, not significant for our purposes, is that GNP counts production by all U.S. citizens whether at home or abroad. GDP counts all production within the geographic borders of the U.S., whether by citizens or by foreigners.

³E.g., annual rent is imputed to measure the current service of owner-occupied houses. The owner is thought of as renting his house from himself in the current year. Yet the owners of automobiles are not thought of as renting their cars to themselves.

⁴One might object that natural resources are not really free. A ton of coal does cost money on the market, but the money price is equal to the labor and capital cost of finding and extracting the coal. Coal in the ground, or *in situ*, as the resource economists say, is considered a free gift of nature. A particularly rich and accessible coal mine will require less labor and capital per ton of coal

assets is not counted because it is not current-year production. The value of a used car bought this year is not counted because it is a transfer of an existing asset. But the commission of the used car salesman will be counted as a service rendered this year. And of course the total value of a new car will be counted this year. The same holds for trading stocks on the stock market.

Total GNP is often divided by the population and stated as per-capita GNP. This is a simple mean and tells us nothing about the distribution of per-capita GNP of individuals about the mean. The mean may or may not reflect a representative central tendency in the distribution. Often modal or median per-capita income is a better measure of central tendency.⁵

GNP is measured in units of “dollar’s worth”. Dollar’s worth of what? Of final goods and services traded in the market in the current year. It is the quantity of all such goods and services, times their price, all summed up. Changes in GNP over time can reflect price changes or quantity changes. To eliminate the effect of price level changes (inflation or deflation), economists correct the dollar figure by converting current dollars into dollars of constant purchasing power. This conversion is done by dividing nominal GNP by a price index that measures the rate of inflation. Suppose that there has been 20% inflation between 1990 and 2000. To convert year 2000 nominal GNP into real GNP, measured in dollars of 1990 purchasing power, we divide GNP in 2000 by 1.20; this is the price index that in the base year of 1990 would have been 1.00, but because of 20% inflation rose to 1.20 in 2000. This gives “real GNP” or rather GNP measured in dollars of constant purchasing power as of a base year.

Changes in real GNP are due to changes in quantities, not price levels. So real GNP, although measured in value units, is an index of quantities of something physical, and is therefore considered a better measure of economic growth than nominal GNP. Just as a dollar’s worth of gasoline corresponds to a definite physical quantity of gasoline, so a dollar’s worth of real GNP corresponds to some aggregate of physical goods and services. But because different goods and services have differing material and energy intensities, there is not a tight one-to-one relationship between real

than a marginal mine. Will its coal sell for less than that of the marginal mine? No, and this gives rise to producer surplus or differential rent. The more accessible mine earns a rent, which results from saved labor and capital relative to the marginal mine. Coal *in situ* is still a free gift of nature, but some free gifts are nicer than others, and differential rent takes that into account. The rent is attributed to the value of labor and capital saved in extraction, not to any original value of the coal in the ground.

⁵The mode is the income category that has the most members. The median is the per-capita income number for which there are as many members above as below. As students of statistics will know, for a normal distribution, the mean, median, and mode will coincide, all giving the same measure of central tendency.

GNP and physical throughput, as there is in the case of dollar's worth of gasoline and the throughput it represents.⁶

The point to emphasize is that although GNP is measured in value terms and cannot be reduced to a simple physical magnitude, it is nevertheless an index of an aggregate of things that all have irreducible physical dimensions. The relationship between real GNP and throughput is not fixed, but neither is its variability unlimited. And, to the extent that one believes that GNP growth can be uncoupled from throughput growth, then all the more reason to be willing to accept limits on throughput growth. If the environmental protection achieved by limiting throughput costs little or nothing in terms of reduced GNP growth, then no one should oppose it. If GNP could grow forever with a constant throughput, then ecological economists would have no objection.

GNP and Total Welfare

GNP is a measure of economic activity, not a measure of welfare. It tells us how fast the wheels are turning, not where the car is going. Economists all say that. Yet in the absence of a true measure of welfare, most policy makers look to the GNP as a trustworthy index of the general direction of change of welfare, based on the following:

$$\text{Total welfare} = \text{economic welfare} + \text{noneconomic welfare}$$

The faith-based assumption is that economic welfare and total welfare move in the same direction. But the increase in economic welfare could induce a more than offsetting decline in noneconomic welfare. For example, GNP goes up as labor becomes more mobile. But the welfare of being close to family and friends gets sacrificed as people have to move. Also, the extra income and job satisfaction of two-earner households raise economic welfare, but the stress of lost leisure and the extra financial burden and lost satisfactions resulting from external childcare reduce noneconomic welfare. Pollution-induced illnesses constitute an enormous loss of noneconomic welfare. Because the category "noneconomic welfare" is unmeasured while economic welfare has a numerical measure, we tend to overestimate the importance of the latter and underestimate the importance of the former. In Figure 13.2, the MDU curve, traditionally missing in economic analysis, represents the loss of "noneconomic welfare."

⁶But even here, economists try to keep the aggregate mix constant in calculating the price index. They assume a given basket of goods and given relative prices of goods in the basket in order to calculate a weighted average price of the basket and its change over time. This average price is not supposed to reflect either changes in relative composition of the basket of goods or changes in *relative* prices of the goods in the basket. Since relative prices inexorably do change over time, as does the composition of the representative basket of goods consumed, price level indexes inevitably "wear out" over time and have to be recalculated. Therefore, real GNP figures lose comparability over longer time periods.

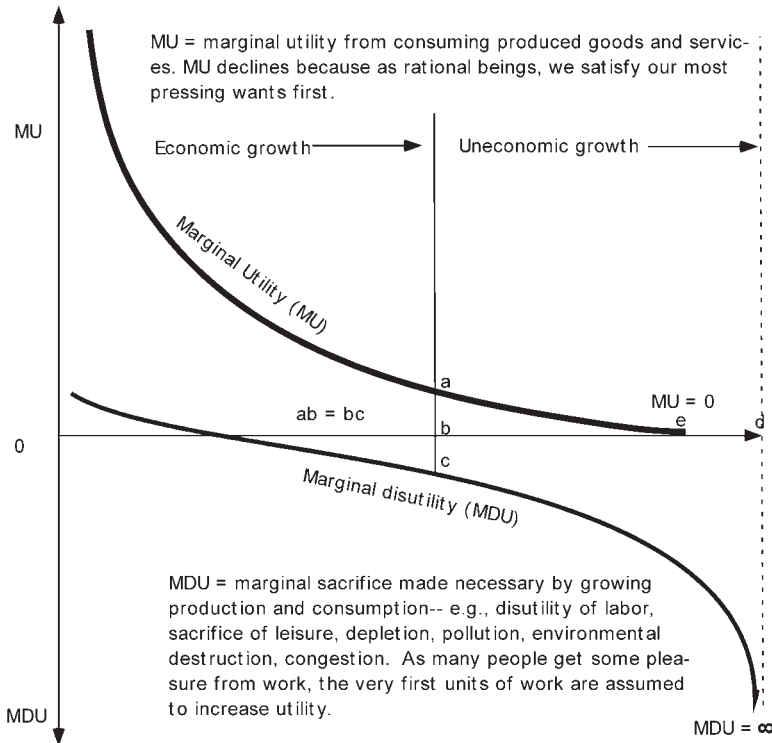


Figure 13.2 • Limits to growth of the macroeconomy. Point b = economic limit or optimal scale, where marginal utility (MU) = marginal disutility (MDU) (maximum net positive utility); e = futility limit, where MU = zero (consumer satiation), d = catastrophe limit, where MDU = infinity. At point d, we have gone beyond sustainable scale.

Defensive Expenditures and the Depletion of Natural Capital

Two other categories are problematic in national income and product accounts: regrettably necessary defensive expenditures and the depletion of natural capital. Let's have a look at each.

Regrettably necessary defensive expenditures, or "defensive expenditures" for short, are those expenditures that we have to make to protect ourselves from the unwanted consequences of the production and consumption of other goods by other people—for example, extra thick walls and windows to block out the sound of living near an airport or busy street or medical services resulting from pollution-induced asthma. In the sense of just measuring activity, these are freely chosen expenditures that people make in order to be better off in their concrete circumstances, and therefore should be counted—they are if not "goods," at least "anti-bads." In another sense, they are really involuntary intermediate costs of

Regrettably necessary defensive expenditures, or "defensive expenditures," are those expenditures that we have to make to protect ourselves from the unwanted consequences of the production and consumption of other goods by other people.

production that should not count as welfare to the final consumer, or as final consumption. This category could be broadly or narrowly defined. The examples just given reflect a narrow definition. Some would include all costs of global warming, and extra legal and law-enforcement costs resulting from a general breakdown in trust and increases in complexity attributed to economic growth. Exactly where to draw the line is a matter of judgment.

The depletion of natural capital is a more clear-cut category. GNP is gross national product. It is gross of depreciation of capital. If we deduct depreciation of manmade capital, we get net national product (NNP), which is a closer approximation to what we can consume without eventual impoverishment. But even in calculating NNP, there is no deduction for the depreciation and depletion of natural capital. Even NNP is gross of natural capital consumption (as well as gross of defensive expenditures). What's more, manmade capital is not a perfect substitute for natural capital for the simple reason that the former cannot exist without the latter. The two are complements. Putting a dollar value on the depreciation of both manmade capital and natural capital implicitly assumes that both types of capital are perfect substitutes, and that we can accept the loss of natural capital as long as manmade capital grows by a compensating amount. In reality, less natural capital makes our manmade capital less valuable as well. Of what use is a car if there is no gas to put in it?

■ SUSTAINABLE INCOME

The true definition of **income**, implicitly stated above, is the maximum that a community can consume in a given time period without causing itself to have to consume less in future time periods.⁷ In other words, income is the maximum you can consume this year without reducing your capacity to produce and consume the same amount next year, and the year after—without reducing future productive capacity, that is, without consuming capital. Strictly speaking it is redundant to say “sustainable income” because income by definition is sustainable. Yet this feature of income has been so overlooked that a bit of redundancy for the sake of emphasis seems useful. If it's not sustainable it is, at least in part, capital consumption, not income.

The whole idea of income accounting is the prudent concern to avoid inadvertent impoverishment by consuming capital. Of course, there are times when we may choose to consume capital—for example, using a nest egg during retirement or liquidating the inventory of a store going out of business. Most of us, however, prefer not to run our national economy

⁷J. Hicks, *Value and Capital*, 2nd ed. Oxford, England: Clarendon, 1948.

and ecosystem as if it were a business in liquidation. Certainly you may choose to consume capital and voluntarily become impoverished. The income accountant's job is to make sure you know what you're doing, not to tell you what to do. But if the accountant does not deduct the consumption of natural capital in calculating income, then she has failed at her professional duty.

To be concrete, if you cut only this year's net growth of a forest, that's income because you can do the same thing again next year. If you cut down the whole forest, you cannot do it again next year, and the value of the cut forest is mostly capital consumption, not income. Yet in GNP, we count the whole amount as this year's income. The same is true for over-exploited fisheries and croplands, and for depleted mines, wells, and aquifers.⁸ Some neoclassical economists have come to realize that nature's services are a huge infrastructure to the economy, and we are failing to maintain that infrastructure.

Why do our national accountants fail to subtract natural capital consumption in calculating income? Neoclassical economics does not count natural capital consumption as a cost because in its preanalytic vision of the world, nature is not scarce. The reason natural funds and resource flows are absent from the usual neoclassical production function is also the reason there is no deduction for natural capital consumption in national income accounting.

■ ALTERNATIVE MEASURES OF WELFARE: MEW AND ISEW

In the early 1970s, there was considerable criticism of GNP growth as an adequate national goal—so much so that economists felt obliged to reply. The best reply came from William Nordhaus and James Tobin.⁹ They questioned whether growth was obsolete as a measure of welfare, and thus as a proper guiding objective of policy. To answer their question, they developed a direct index of welfare, called Measured Economic Welfare (MEW), and tested its correlation with GNP over the period 1929–1965. They found that, for the period as a whole, GNP and MEW were indeed positively correlated; for every six units of increase in GNP, there was, on average, a four-unit increase in MEW. Economists breathed a sigh of relief, forgot about MEW, and concentrated again on GNP. Although GNP was not designed as a measure of welfare, it was, and still is, thought to be

⁸The running down of renewable stocks or funds of natural capital is *depreciation*, analogous to the depreciation of a machine. The running down of nonrenewable natural capital is *liquidation*, analogous to the liquidation of an inventory. Both represent capital consumption.

⁹W. Nordhaus and J. Tobin, "Is Growth Obsolete?" In *Economic Growth*, National Bureau of Economic Research, New York: Columbia University Press, 1972.

sufficiently well correlated with welfare to serve as a practical guide for policy.

Some 20 years later, Daly and Cobb revisited the issue and began to develop an Index of Sustainable Economic Welfare (ISEW) with a review of the Nordhaus and Tobin MEW. They discovered that if one takes only the latter half of the Nordhaus-Tobin time series (i.e., the 18 years from 1947 to 1965), the positive correlation between GNP and MEW *falls* dramatically. In this most recent half of the total period—surely the more relevant half for projections into the future—a six-unit increase in GNP yielded on average only a one-unit increase in MEW. This suggests that GNP growth at this stage in U.S. history may be quite an inefficient way of improving economic welfare—certainly less efficient than in the past.

The ISEW was then developed to replace MEW, since the latter omitted any correction for environmental costs, did not correct for distributional changes, and included leisure, which both dominated the MEW and introduced many arbitrary valuations.¹⁰ The ISEW, like the MEW, though less so, was positively correlated with GNP up to a point (around 1980), beyond which the correlation turned slightly negative.¹¹ Figure 13.3 shows estimates of GNP and ISEW for seven different countries.

Measures of welfare are difficult and subject to many arbitrary judgments, so sweeping conclusions should be resisted. However, it seems fair to say that for the United States since 1947, the empirical evidence that GNP growth has increased welfare is weak, and since 1980 probably nonexistent. Consequently, any impact on welfare via policies that increase GNP growth would also be weak or nonexistent. In other words, the “great benefit,” habitually used to justify sacrifices of the environment, community standards, and industrial peace, appears, on closer inspection, likely not even to exist.¹² Certainly if economic growth is to be the

¹⁰The concept of leisure is an important part of welfare, but the problems of valuing leisure are difficult. Is the leisure chosen or unchosen? Should sleep time count as leisure? Is commuting time leisure or “time cost of working”? Should we use the wage rate? The minimum wage? Should the “leisure” of mom taking care of children be valued at her opportunity cost if she’s a doctor, or at the cost of avoided daycare? Such difficult choices have a big effect on the index.

¹¹Neither the MEW nor the ISEW considered the effect of individual country GNP growth on the *global* environment, and consequently on welfare at geographic levels other than the nation. Nor was there any deduction for legal harmful products, such as tobacco or alcohol, or illegal harmful products, such as drugs. No deduction was made for overall diminishing marginal utility of income resulting from GNP growth over time (although a distributional correction for lower marginal utility of extra income to the rich was included). Such considerations would further weaken the correlation between GNP and welfare. Also, GNP, MEW, and ISEW all begin with personal consumption. Since all three measures have in common their largest single category, there is a significant autocorrelation bias, which makes the poor correlations between GNP and the two welfare measures all the more surprising.

¹²For further evidence from other countries, see M. Max-Neef, *Economic Growth and Quality of Life: A Threshold Hypothesis*, *Ecological Economics* 15: 115–118 (1995).

Image Not Available

Figure 13.3 • Indices of GNP (solid) and ISEW (dashed) for seven countries. 1970 = 100 in all cases. (Source: R. Costanza, J. Farley, and P. Templet, “Quality of Life and the Distribution of Wealth and Resources.” In R. Costanza and S. E. Jørgensen (eds.), *Understanding and Solving Environmental Problems in the 21st Century: Toward a New, Integrated Hard Problem Science*, Amsterdam: Elsevier, 2002.)

number-one goal of nations and the central organizing principle of society, then citizens have a right to expect that the index by which we measure growth, GNP, ought to reflect general welfare more accurately than it does. Continued use of GNP as a proxy for welfare reminds us of the quote often attributed to Yogi Berra: “We may be lost, but we’re making great time.”

The objective, accurate scientific measurement of national costs and national benefits is not a realistic goal. Both costs and benefits of economic growth are spread out over time, and how we treat costs and benefits that affect future generations is an ethical issue, not a scientific one. The use of a particular discount rate to address intertemporal distribution is, for example, clearly a value-laden decision. Ecosystem change and evolution are not predictable, and how we treat the resulting uncertainty is also an ethical issue. Even using monetary measures of

Box 13-1 GROSS NATIONAL COST?

Years ago, Kenneth Boulding suggested that the GNP be relabeled GNC for gross national *cost*. While we are sympathetic to Boulding's tongue-in-cheek plea, we can't help feeling a certain nostalgia for the good old days when newscasters regaled us with quarterly changes in the GNP. Now we are subjected to hammer-banging, gong-clanging reports of hourly changes in the Dow Jones and Nasdaq stock price indices—numbers that are an order of magnitude further removed from either welfare or income than GNP is! This is because stock market values are forward-looking, based on expectations of future earnings (even on speculators estimates of the expectations of others). By contrast, GNP is backward-looking, a historical record of what has already happened. Since the past is better known than the future, GNP is inherently a more trustworthy number than stock market values.

What should be done about GNP? Boulding's suggestion to count all of it as cost was an exaggeration, but no more so than the current practice of counting it all as benefit. A more sensible procedure would be to disaggregate GNP into two separate accounts: a national benefits account and a national costs account. As the scale of the economy grows, both benefits and costs will increase. We should be comparing those benefit and cost increases at the margin to find the optimal scale (see Figure 13.2).^a It makes absolutely no sense to add them together!

^aFor an effort in this direction for Australia, see P. A. Lawn, *Toward Sustainable Development: An Ecological Economics Approach*, Boca Raton, FL: Lewis Publishers, 2001.

market goods is not objective; markets will yield different monetary values depending on the initial distribution of the wealth, and what comprises a desirable initial distribution is an ethical judgment. Monetary values for a given resource also vary depending on the amount of the resource society is using—for example, the price of oil depends primarily on current rates of extraction of oil. Oil is such an important input into so many economic processes that all prices are affected by how much oil we are using. Using prices determined by resource use in this period to decide the appropriate amount of a resource to use is therefore a case of circular reasoning; you can't do it on a computer spreadsheet, and you can't do it in real life. Efforts to put monetary values on nonmarket goods such as ecosystem services not only compounds these ethical issues with serious methodological problems, but also implicitly assumes that natural capital and manmade capital are perfect substitutes, a position that most ecological economists strongly reject.

■ BEYOND CONSUMPTION-BASED INDICATORS OF WELFARE

Personal consumption is not an end in itself, but merely one means toward achieving the end of enhancing human welfare. GNP is not only inadequate as a proxy for income, but income is only one element among many that provides human welfare. For example, the ecosystem services that increasing GNP inevitably encroaches upon are at least as important as GNP in providing welfare.¹³

Relative Wealth and Well-Being

While in the past, consumption was arguably one of the most limiting factors in enhancing welfare, abundant evidence suggests that this is no longer the case. Above a certain level, resource consumption and wealth may be “positional”—that is, our welfare is determined by comparing our position with that of others. Social psychology has long recognized that a person’s sense of well-being is in part determined by his or her sense of relative affluence or deprivation. In this view, we are engaged in a never-ending wealth and consumption race, where greater consumption by our reference group demands greater consumption on our part simply to maintain the same relative position. If this is true, then doubling GNP with no change in distribution might have no positive impact on welfare. Moreover, a profoundly negative impact may result if the doubling of GNP comes at the expense of vital ecosystem services.

In reality, with current economic growth patterns leading to greater concentration of the wealth in the hands of the few, the majority of the population may be falling behind in this race. The wealthy generally compare themselves with each other and not with the poor, and therefore they may not be achieving greater welfare either. To the contrary, the blind pursuit of positional wealth and consumption places substantial demands on our time and resources, and leaves us with a decreasing ability to meet our other human needs.¹⁴

Human Needs and Welfare

Do other factors not yet discussed contribute to our welfare? It is reasonable to assume that welfare is determined by the ability to satisfy one’s

¹³See, e.g., R. Costanza et al. The Value of the World’s Ecosystem Services and Natural Capital, *Nature* 6630: 253–260 (1997), in which the value of global ecosystem goods and services is found to outweigh global GNP. While this paper does put monetary values on natural capital for purposes of comparison with manmade capital, it also explicitly discusses many of the problems with this approach.

¹⁴E.g., F. Hirsh, *The Social Limits to Growth*, Cambridge, MA: Harvard University Press, 1976; R. Frank, *Luxury Fever: Why Money Fails to Satisfy in an Era of Excess*. New York: Free Press, 1999.

needs and wants. What are our needs? Absolute needs are those required for survival and are biologically determined. Some 1.2 billion individuals globally and 28% of the population in the Third World currently live in extreme poverty and have difficulty meeting even these absolute needs.¹⁵ For this group, greater consumption is probably very closely correlated to greater welfare.

Once absolute needs have been met, as is the case for the remaining four-fifths of the world's population, then welfare is determined by the satisfaction of a whole suite of primary human needs. Numerous researchers have proposed a variety of human needs, typically claiming that they are pursued in hierarchical order—Maslow's hierarchy (1954) (in which consumption only meets the lowest rung on the needs ladder) being the most famous. The hierarchical ordering, though generally not seen as rigid by these researchers, still leaves something to be desired. Even the 1.2 billion people living in absolute poverty seek to fulfill other needs than mere subsistence.

Manfred Max-Neef¹⁶ has summarized and organized human needs into nonhierarchical axiological¹⁷ and existential categories (Table 13.1). In this **matrix of human needs**, needs are interrelated and interactive—many needs are complementary, and different needs can be pursued simultaneously. This is a better reflection of reality than a strict hierarchy in which we only pursue higher needs after lower ones have been fulfilled. Also important in Max-Neef's conception, needs are both few and finite. This stands in stark contrast to the assumption of infinite wants, or the nonsatiety axiom in standard economics.

If we are to evaluate the success of economic policies both now and in the future (assuming that providing a high level of welfare for humans for the indefinite future is our economic goal), then we must develop measurable indicators that serve as suitable proxies for needs fulfillment and welfare.

To state the obvious, we cannot precisely measure welfare, which in the present context is equivalent to quality of life (QOL). In the words of Clifford Cobb:¹⁸

the most important fact to understand about QOL indicators is that all measures of quality are proxies—indirect measures of the true condition we are

¹⁵D. E. Bloom, D. Canning, B. Graham, and J. Sevilla, "Out of Poverty: On the Feasibility of Halving Global Poverty by 2015" (Discussion Paper No. 52). Consulting Assistance on Economic Reform (CAER II), World Bank, 2001. Social Capital for Development: What Is Social Capital. Online: <http://www.worldbank.org/poverty/scapital/whatsc.htm>. World Bank.

¹⁶M. Max-Neef, Development and Human Needs. In P. Ekins and M. Max-Neef, *Real-life Economics: Understanding Wealth Creation*. London: Routledge, pp. 197–213, 1992.

¹⁷Axiology is the study of the nature of values and value judgments.

¹⁸C. W. Cobb, *Measurement Tools and the Quality of Life: Redefining Progress*, Oakland, CA. Online: http://www.rprogress.org/pubs/pdf/measure_qol.pdf.

■ **Table 13.1**

MAX-NEEF'S MATRIX OF HUMAN NEEDS

Axiological Categories	Existential Categories			
	Being	Having	Doing	Interacting
Subsistence	Physical health, mental health, equilibrium, sense of humor, adaptability	Food, shelter, work	Feed, procreate, rest, work	Living environment, social setting
Protection	Care, adaptability, autonomy, equilibrium, solidarity	Insurance systems, savings, social security, health systems, rights, family, work	Cooperate, prevent, plan take care of, cure, help	Living space, social environment, dwelling
Affection	Self-esteem, solidarity, respect, tolerance, generosity, receptiveness, passion, determination, sensuality, sense of humor	Friendships, family, partnerships with nature	Make love, caress, express, emotions, share, take care of, cultivate, appreciate	Privacy, intimacy, home, space of togetherness
Understanding	Critical conscience, receptiveness, curiosity, astonishment, discipline, intuition, rationality	Literature, teachers, method, educational policies, communication policies	Investigate, study, experiment, educate, analyze, meditate	Settings of formative interaction, schools, universities, academies, groups, communities, family
Participation	Adaptability, receptiveness, solidarity, willingness, determination, dedication, respect, passion, sense of humor	Rights, responsibilities, duties, privileges, work	Become affiliated, cooperate, propose, share, dissent, obey, interact, agree on, express opinions	Setting of participative interaction, parties, associations, churches, communities, neighborhoods, family
Idleness	Curiosity, receptiveness, imagination, recklessness, sense of humor, tranquility, sensuality	Games, spectacles, clubs, parties, peace of mind	Daydream, brood, dream, recall old times, give way to fantasies, remember, relax, have fun, play	Privacy, intimacy, space of closeness, free time, surroundings, landscapes
Creation	Passion, determination, intuition, imagination, boldness, rationality, autonomy, inventiveness, curiosity	Abilities, skills, method, work	Work, invent, build, design, interpret	Productive and feedback settings, workshops, cultural groups, audiences, spaces for expressions, temporal freedom

Continued

■ Table 13.1

MAX-NEEF’S MATRIX OF HUMAN NEEDS (CONTINUED)

Axiological Categories	Existential Categories			
	Being	Having	Doing	Interacting
Identity	Sense of belonging, consistency, differentiation, self-esteem, assertiveness	Symbols, language, religion, habits, customs, reference groups, sexuality, values, norms, historical, memory, work	Commit oneself, integrate oneself, confront, decide on, get to know oneself, recognize oneself, actualize oneself, grow	Social rhythms, everyday settings, settings which one belongs, maturation stages
Freedom	Autonomy, self-esteem, determination, passion, assertiveness, open-mindedness, boldness, rebelliousness, tolerance	Equal rights	Dissent, choose, be different, run risks, develop awareness, commit oneself, disobey	Ability to come in contact with different people at different times in different places

The column of Being registers attributes, personal or collective, that are expressed as nouns. The column of Having registers institutions, norms, mechanisms, tools (not in material sense), laws, etc. that can be expressed in one or few words. The column of Doing registers locations and milieus (as time and spaces). It stands for the Spanish *estar* or the German *befinden*, in the sense of time and space. As there is no corresponding word in English, Interacting was chosen for lack of something better.

Source: M. Max-Neef, “Development and Human Needs.” In P. Ekins and M. Max-Neef, *Real-Life Economics: Understanding Wealth Creation*. London: Routledge, 1992, pp. 197–213.

seeking to judge. If quality could be quantified, it would cease to be quality. Instead, it would be quantity. Quantitative measures should not be judged as true or false, but only in terms of their adequacy in bringing us closer to an unattainable goal. They can never directly ascertain quality. (p. 5)

Objective Measures

Numerous efforts have been made to objectively measure welfare. The problem is that these studies have found only weak relationships between objective measures of welfare and the subjective assessments of the same by the subjects concerned.¹⁹ However, both these studies and the various types of national accounts seem to include a relatively narrow range of objective indicators, often placing what we consider to be an excessive emphasis on consumption. Quite possibly the problem is that welfare is too rich a gumbo for us to recapture its flavor with so few ingredients. An important research agenda in economics is to develop a methodology for

¹⁹B. Haas, A Multidisciplinary Concept Analysis of Quality of Life. *Western Journal of Nursing Research* 21(6):728–743 (1999).

measuring access to “satisfiers” (the means by which we satisfy a given need) for Max-Neef’s axiological and existential categories of human needs as indicators of welfare. With sufficient ingredients, we can produce something reasonably close to the flavor of welfare.

Max-Neef’s human needs matrix as the basis of a welfare measure is a dramatic departure from existing national accounts, as well as from most of the proposed alternatives, differing even in its theoretical underpinnings. Neoclassical economics and GNP are explicitly utilitarian. Within utilitarian philosophy, individual welfare is determined by the degree to which individuals can satisfy their desires, and it is generally accepted that the goal of society is to provide the maximum amount of utility for its citizens. As utilitarian philosophy has been operationalized by NCE, citizens are the best able to determine what provides utility. Because it is extremely difficult to measure utility directly, economists have taken to using revealed preferences as a proxy. Preferences are revealed by people’s objectively measurable choices in the market. In the market economy, preferences are revealed through market decisions, and market decisions can only be made with money. Under this conception of utilitarianism, the philosophy values only end-states and requires only “having” such things as possessions and experiences. Sustainable income accounting and measurements of economic welfare are basically just extensions of this philosophy, and they similarly value only “having.”²⁰

In Max-Neef’s framework, having things is important, but it is just one of the elements required to meet our needs. Thus, a benevolent dictator with the resources to provide us with all the physical things we require for happiness would fail to meet our existential needs for being, doing, and interacting, as well as our axiological needs for creation, participation, and freedom. Also, within Max-Neef’s conception, people are not always best able to determine what contributes to their quality of life—for example, advertising may falsely convince people that consumption satisfies their need for affection, freedom, or participation.

This approach, which values human actions independently of their outcomes, has been dubbed the “human development” approach to welfare. Its main proponents include Nobel Prize-winning economist Amartya Sen and Martha Nussbaum. In a similar tone to Max-Neef, they argue that “capabilities” and “functionings” are critical to welfare.²¹ Roughly speaking, “functionings” correspond to human needs, while

²⁰C. W. Cobb, *Measurement Tools and the Quality of Life* (Redefining Progress, Oakland, CA). (2000). Online: http://www.rprogress.org/pubs/pdf/measure_qol.pdf.

²¹Ibid.; M. Nussbaum, *Aristotelian Social Democracy*. In R. B. Douglass, G. M. Mara, and H. S. Richardson (eds.), *Liberalism and the Good*, New York: Routledge, pp. 203–252, 1990; R. Sugden, *Welfare, Resources, and Capabilities: A Review of Inequality Reexamined by Amartya Sen*. *Journal of Economic Literature* 31 (December): 1947–1962.

“capabilities” include both states of being and opportunities for doing, and are therefore analogous to access to satisfiers for these needs in Max-Neef’s matrix (see Table 13.1). In utilitarian theory, we might have several different options, of which we choose one. If all options but that one were eliminated, it would not affect our welfare. In the human development approach, losing options restricts our capabilities and would therefore affect our welfare. The human development approach is less concerned with the actual choices that people make than with the options they are free to choose from, and the marketplace is only one of many spheres in which choice is important.

Operationalizing Human Needs Assessment as a Measure of Welfare

Measuring the extent to which human needs are satisfied is, of course, an exceptionally difficult task and highly subjective. Following the lead of Sen and Nussbaum, it would be most useful to measure capabilities, that is, the extent to which individuals have access to satisfiers. However, as noted by Max-Neef, specific satisfiers may vary by culture, and the difference in satisfiers required to meet a human need may indeed be one of the key elements that defines a culture. This means that objective “welfare accounts” must be very culture-specific. Second, some satisfiers might help fulfill several human needs, while other needs require several satisfiers. Further complicating matters, satisfiers may change through time. And humans are social creatures who inhabit a complex environment; needs are satisfied not only in regard to the individual, but also in regard to the social group and environment.²² Furthermore, while needs are different and distinct, they are also interactive and may complement each other, and therefore may not be additive. Abundant access to satisfiers for one set of needs does not compensate for a lack of satisfiers for another set of needs. This suggests that separate “accounts” should be kept for access to satisfiers to different needs.

In developing welfare accounts based on **human needs assessment (HNA)**, it would be useful to test measurements of satisfiers empirically in studies comparing these objective measures against subjective assessments of welfare to determine their effectiveness. These empirical tests, as well as efforts to operationalize HNA accounts, must involve people in dialogues to confirm or refute the validity of the needs Max-Neef specifies, as well as the validity of the satisfiers we use to assess the degree to which needs are met. Such dialogues would almost certainly elicit additions and alternatives to the generic satisfiers, the entries in the columns of Table

²²Op. cit., Max-Neef (1992).

13.1.²³ While the average person may not always know exactly what satisfiers will best meet their needs, interactive discussion with people is nonetheless essential to select and test appropriate indicators. We would also need to develop group-based methodologies to determine the effectiveness of our indicators in a social setting.

It is clear that Max-Neef's approach is very difficult to operationalize, even if his concept is theoretically more compelling than GNP or even ISEW. The debate over which approach to take to national accounting—theoretically sound measures or ease of accounting—is old. As Irving Fisher argued back in 1906, the appropriate measure, even of income, is one that captures the psychic flux of service (i.e., satisfaction of needs and wants) and not simply the final costs of goods and services.²⁴ And at the time Fisher wrote, the absence of suitable data for calculating either psychic flux of service or final costs no doubt led many to ignore the debate as entirely academic. The widespread use of GNP indicates that in practice, Fisher lost this earlier debate. However, measures such as the ISEW suggest that the GNP is becoming increasingly incapable of measuring economic welfare, much less general human welfare. Even if we can never quantify access to satisfiers as precisely as we currently quantify GNP, as Sen suggests, perhaps it is better to be vaguely right than precisely wrong.²⁵

Accepting Max-Neef's human needs matrix as a framework for the specific elements of human welfare, and access to satisfiers as potentially the best objective indicator of welfare, has profound implications with respect to scale, distribution, and allocation. First, most of the possible indicators suggested by Max-Neef require few, if any, material resources beyond those required to sustain human life, and hence are not subject to physical exhaustion. Thus, for most elements of human welfare, increases for one person or one generation do not leave less for others. Second, explicitly accepting that there is a limit to material needs implies that we can limit consumption greatly with little, if any, sacrifice of welfare. This result is critical, because the laws of thermodynamics make it impossible to uncouple physical consumption from resource use and waste production. Abundant evidence suggests that current levels of consumption could not

²³E.g., food and shelter are specific dimensions of "having" that are satisfiers of the need for "subsistence." How we actually meet our needs for food and shelter are culture-specific. A traditional Inuit might be satisfied with walrus blubber and an igloo, while a New Yorker would require hamburgers and a high-rise apartment.

²⁴H. Daly and J. Cobb, *For the Common Good: Redirecting the Economy Toward Community, the Environment, and a Sustainable Future*. Boston: Beacon Press, 1989.

²⁵D. Crocker, "Functioning and Capability: The Foundations of Sen's and Nussbaum's Development Ethic, Part 2." In M. Nussbaum and J. Glover, eds. *Women, Culture, and Development: A Study in Human Capabilities*, Oxford, England: Oxford University Press, 1995.

be sustainably met with renewable resources alone, and we must therefore limit consumption or else threaten the welfare of future generations.

The difficulty of operationalizing Max-Neef’s framework may actually be a point in its favor. Why do we want to measure welfare in the first place? It’s not just to track its rise or fall, but to help us create policies to improve it. Simply providing statistical data on welfare doesn’t help us achieve this end. However, applying Max-Neef’s framework would require extensive surveys asking people to think deeply about what their needs really are, and how they can satisfy them. Ultimately, improving welfare falls to decisions by political, cultural, and religious groups about what they want and how they want to achieve their goals, and making the correct decisions will require people to think deeply about what it is they ultimately desire.

BIG IDEAS to remember	
■ Fallacy of composition	■ Defensive expenditures or “anti-bads”
■ General equilibrium model versus aggregate macroeconomics	■ Natural capital consumption
■ Optimal scale of macroeconomy	■ Sustainable income
■ Gross national (or domestic) product	■ MEW and ISEW
■ Total welfare = economic welfare + noneconomic welfare	■ Gross national cost
	■ Relative wealth and welfare
	■ Human needs and welfare
	■ Matrix of human needs (Max-Neef)
	■ Human needs assessment (HNA)



Money

Money ranks with the wheel and fire as ancient inventions without which the modern world could not function. Probably more people today are “run over and burned” by out-of-control money than by out-of-control wheels and fires. Money is mysterious. Unlike matter and energy, it *can* be created and destroyed, evading the laws of thermodynamics. Private citizens (counterfeiters) are sent to jail for making even small amounts of it, yet private commercial banks make almost all of it, and we pay them for it! Sometimes money is a costly commodity (gold) and sometimes a costless token (paper notes). It is easily transferable into real assets by individuals, but the community as a whole cannot exchange its money into real assets at all, since someone in the community ends up holding the money. Some economists think the money supply should be determined by fixed rules, others think it should be manipulated by public authorities. And some people think the love of money is the root of all evil! Anyone who is not confused by money probably hasn’t thought about it very much.

Money functions as a medium of exchange, a unit of account, and a store of value. The functions are interrelated but worth considering separately. To measure exchange value, we need a unit—call it a dollar, a peso, a franc, or a yen. If the unit is stable over time (no inflation or deflation), then money automatically serves as a store of exchange value. To function as a medium of exchange and let us escape the inconvenience of barter, money must hold its value at least long enough to effect both sides of the transaction, which in barter, of course, are simultaneous. A moment’s reflection shows how tremendously inefficient barter is, and consequently how efficient money is. In barter there must be a coincidence of wants—it is not enough that I want what you have to trade; you also have to want whatever it is that I have to trade, and we have to find each other. Money

provides a common denominator that everyone wants simply because everyone else is willing to accept it. It is a standard, well-defined commodity (or later a token) that breaks the two sides of a difficult barter arrangement into two separate and easy transactions.

Karl Marx analyzed transactions as follows. First we have simple barter, which he denoted as:

$$C—C^*$$

Commodity C is exchanged for commodity C*. You have C and prefer C*; I have C* and prefer C. We are both better off after the transaction. We both increase the use value of what we own. Exchange value is not separated from use value. No money is needed, but we were lucky to have found each other.

Next for Marx comes “simple commodity production”:

$$C—M—C^*$$

Now we have money functioning as a medium of exchange. Exchange value, the sum of money, M, is entirely instrumental to bringing about an increase in use values by facilitating the exchange. The process begins and ends with commodity use values. The goal is to increase use value, not exchange value.

For Marx the critical change comes in the historical shift from simple commodity production to “capitalist circulation,” which he symbolized as:

$$M—C—M^*$$

The capitalist starts with a sum of money capital, M, uses it to make commodity C, and then sells C for the amount M*, presumably greater than M. Thus:

$$M^*—M = \Delta M$$

ΔM is profit, or surplus value in Marxist terms. For us the important thing is not Marx’s notion of surplus value, which is tied up with his very problematic labor theory of value, but the simple observation that in moving from C—M—C* to M—C—M* the driving motive has shifted from increasing use value to increasing exchange value.

Use value arises from the actual use of commodities, it is concrete and physically embodied. **Exchange value** is abstract and inheres in money. It has no necessary physical embodiment.¹ Real wealth—commodities—obey the laws of thermodynamics. Money, a mere symbolic unit of account, can be created out of nothing and destroyed into nothing. There is a physical limit to the accumulation of use values. There is no obvious

¹Though, of course, exchange value is only real if something exists for which money can actually be exchanged.

limit to the accumulation of exchange value. Fifty hammers are not much better than two (one and a spare) as far as use values are concerned. But in terms of exchange value, fifty hammers are much better than two, and better yet in the form of fifty hammers' worth of fungible money that can be spent on anything, anywhere, and at any time.

Box 14-1 DIAMONDS-WATER PARADOX

The distinction between use value and exchange value goes back to Aristotle and was used to “resolve” the “diamonds-water paradox”—the paradox that although water is a necessity it has a low price, while diamonds are practically useless but have a high price. Economists dealt with this conundrum by declaring that there are two basic kinds of value, use value and exchange value, and one has nothing to do with the other. In the late 1800s the marginalist revolution in economic thinking resolved the paradox as follows: Exchange value is determined by marginal utility and use value is determined by total utility; that is, exchange value equals marginal use value. Water has enormous total utility, but it is so plentiful that at the margin we use it for trivial satisfactions. This marginal utility determines exchange value. How do we know that? If you want to buy a gallon of water from me, what determines how much you will have to give me in exchange? If I give you a gallon of water, I won't stop drinking and go thirsty, nor will I stop bathing and be dirty. I'll probably water my petunias less often. The petunias are my least important use value, my marginal utility of water, my opportunity cost for a gallon of water. Since the *marginal* utility of water is what I will sacrifice by trading away a gallon, that's what determines the exchange value of water. Exchange value is determined by the least important use value, the value sacrificed. Water is abundant so its marginal utility is very small; diamonds are scarce, so their marginal utility is still high.

A hoard of hammers takes up space and is subject to rust, termites, fire, and theft. Fifty hammers' worth of money is not subject to rust, rot, and entropy, and far from costing a storage fee will earn interest from whomever gains the privilege of “storing” it for you. Production for use value is self-limiting. Production for the sake of exchange value is not self-limiting. Since there is no limit to the accumulation of abstract exchange value, and since abstract exchange value is convertible into concrete use value, we seem to have concluded that there must not be any limit to concrete use values either. This has perhaps led to the notion that exponential growth, the law of money growing in the bank at compound interest, is also the law of growth of the real, or material, economy.

■ VIRTUAL WEALTH

Frederick Soddy summarized all this by carefully distinguishing wealth from debt.² He noted that “a weight, although it is measured by what it will pull up, is nevertheless a pull down. The whole idea of balancing one thing against another in order to measure its quantity involves equating the quantity measured against an equal and opposite quantity. Wealth is the positive quantity to be measured and money as the claim to wealth is a debt” (p. 103).³ Monetary debt, the measure of wealth, is negative wealth, say minus two pigs. It obeys the laws of mathematics, but not of physics. Wealth, on the other hand, plus two pigs, obeys the laws of thermodynamics as well as mathematics. Positive pigs die, have to be fed, and cannot reproduce faster than their gestation period allows. Negative pigs are hyper-fecund and can multiply mathematically without limit. As Soddy put it, “you cannot permanently pit an absurd human convention, such as the spontaneous increment of debt (compound interest), against the natural law of the spontaneous decrement of wealth (entropy)” (p. 30).

The holding of token money by the public to avoid the inconvenience of barter gives rise to the curious phenomenon that Soddy called **virtual wealth**, which he defined as the aggregate value of the real assets that the community voluntarily abstains from holding in order to hold money instead. Individuals can always convert their money holdings into real assets, but they choose not to in order to avoid the inconvenience of barter. This raises the question of whether money should be counted as a part of the real wealth of the community? Yes, if money is a commodity like gold that circulates at its commodity value. No, if it is token money like a dollar bill whose commodity value is nil but whose exchange value is significant. Even though each person can at an instant convert his money into real assets, it is impossible for the community as a whole to do this, as we have previously noted.

Money, therefore, represents not real wealth but, in Soddy's term, “virtual wealth.” More exactly, it is the magnitude of virtual wealth that determines the value of money. What happens if the government puts into circulation more money than people currently want to hold? People will exchange money for real assets and drive up the price of real assets. As the price of real assets rises, the real value of money falls until it again coincides with the virtual wealth of the community. If there is too little money, people will exchange real assets for money, thereby

²F. Soddy, *Wealth, Virtual Wealth, and Debt*, London: George Allen & Unwin, 1926.

³When banks create money by providing someone with a loan (see below), they actually create a debt as the first step. On the assets side of the accounting books, the banker enters a debt for the amount of money borrowed (to be paid off with interest). This borrowed money is then placed in a bank account, which is listed in the bank's books as a liability.

driving down the price of real assets. As the price of real assets fall, the value of money increases until it again equals the virtual wealth of the community. The value of a dollar, then, is the virtual wealth of the community divided by however many dollars are in circulation. It follows that the value of a unit of token money is not determined by the total wealth of a community, nor by its annual GNP, but by its virtual wealth relative to the money supply.

Box 14-2 VIRTUAL WEALTH AND FIDUCIARY ISSUE

Nobel laureate economist James Tobin comes very close to Soddy's concept of virtual wealth in his explanation of the "fiduciary issue":

The community's wealth now has two components: the real goods accumulated through past real investment and fiduciary or paper "goods" manufactured by the government from thin air. Of course, the nonhuman wealth of such a nation "really" consists only of its tangible capital. But as viewed by the inhabitants of the nation individually, wealth exceeds the tangible capital stock by the size of what we might term the fiduciary issue. This is an illusion, but only one of the many fallacies of composition which are basic to any economy or society. The illusion can be maintained unimpaired as long as society does not actually try to convert all its paper wealth into goods."^a

^aJ. Tobin, "Money and Economic Growth," *Econometrica* (October 1965), p. 676.

SEIGNIORAGE

Who owns the virtual wealth? Since it does not really exist, we might say that no one owns it. It is a collective illusion. Yet individuals voluntarily hold money instead of real assets, and they behave as if money were a real part of their individual wealth, even if they understand that collectively it is only "virtual" or illusory. Every member of the community who holds money had to give up a real asset to get it—except for the issuer of money. The one who creates the money and is the first to spend it gets a real asset in exchange for a paper token. The difference between the monetary value and the negligible commodity value of the token, the profit to the issuer of money, was and still is called **seigniorage**, in recognition of the lordly nature of this privilege. Who is this fortunate person? Historically it was the feudal lord, or the king, the sovereign, who issued money within his domain. One might expect that this privilege would have been passed on to the sovereign's legitimate heir, the democratic state. To some extent this is the case, because only governments can issue currency or legal tender. However, over 90% of our money supply today is not currency but demand deposits created by the private commercial banking

system.⁴ They are created out of nothing and loaned into existence by the private commercial banks under rules set up by the government. Who gets the seigniorage? Seigniorage from currency goes to the government. Seigniorage from demand deposits goes to the private sector, initially to commercial banks. To the extent there is competition among banks for savings, they will redistribute some of the seigniorage to depositors. Sectors of society too poor to save will receive nothing.

What does money consist of in our economy? A further mystery of money is that it has several definitions. The most restrictive is “currency plus demand deposits in the hands of the nonbank public.” More expansive definitions include savings deposits and even credit card debt. Most of our money supply bears interest as a condition of its existence. Whoever borrowed it into existence must pay back what he borrowed plus interest. Thus, a requirement for growth (or else inflation) is built into the very existence of our money supply. Moreover, the money supply, *ceteris paribus*, expands during boom times when everyone wants to borrow and invest, and contracts during recessions when loans are foreclosed, thereby aggravating cyclical instabilities.

On learning for the first time that private banks create money out of nothing and lend it at interest, many people find it hard to believe. Indeed, according to Joseph Schumpeter, as late as the 1920s, 99 out of 100 economists believed that banks could no more create money than cloakrooms could create coats. Yet now every economics textbook explains how banks create money. We will explain how it works in a minute, but first we’ll let the strangeness of it sink in. Of course, this is not the only way to create money. Nonetheless, most economists today accept this situation as normal. But the leading economists of the early part of the twentieth century, Irving Fisher and Frank Knight, thought it was an abomination. And so did Frederick Soddy.

Box 14-3

LOCAL CURRENCIES AND LOCAL EXCHANGE TRADING SYSTEMS

Currencies are created not exclusively by governments. A variety of non-government legal currencies exist in countries throughout the world, and a closer look at local currencies can provide important insights into money. There are three ways to design a currency system. Most national currencies are created by fiat. There is nothing to back up fiat money but faith that someone else will accept it in exchange for goods (“in God we Trust”

⁴Demand deposits are ordinary checking accounts from which money is payable “on demand” to the bearer of your check.

or as the Ithaca HOUR says “in Ithaca we trust”). Second, a currency can be valued in terms of a commodity, and may or may not be redeemable in terms of that commodity. For example, the Constant, one of the earliest alternative currencies and a forerunner of today’s local currencies, was introduced in the 1970s on an experimental basis in Exeter, New Hampshire. The Constant was designed to maintain a constant value against a basket of 30 different commodities. Finally, a currency can be backed by a commodity, which means it can freely be exchanged for that commodity. Such was the case for U.S. currency in the nineteenth century when money holders could theoretically exchange gold-backed dollars for gold at any time, and the necessary gold reserves were physically available to do this.^a

The city of Ithaca, New York, has one of the best-developed local currency systems in the world. The currency is known as Ithaca HOURS. An individual can participate in the HOURS system simply by agreeing to accept HOURS in exchange for the goods or services she produces. New money must be issued to chase this greater supply of goods and services. Where does this new money come from?

Published backers of the HOUR directory, which is considered a service provided to Ithaca HOURS, are paid 2 HOURS (the equivalent of approximately \$20 US) on first participation and again when they renew their commitment. Technically speaking, the participant is being paid for publicly backing HOURS, but one could also say that in the HOURS system, the person who agrees to generate new goods and services earns the right to seigniorage. While at first glance it may seem strange that one would be entitled to money for simply agreeing to accept money, new money must clearly come from somewhere, and it’s reasonable for part of it to go to the person responsible for creating the new wealth.

Theoretically, the amount of new money created times the velocity with which the money circulates should equal the amount of new goods and services being offered. So far it seems that new participants have on average offered more than enough goods and services to use up their 2 new HOURS. Several mechanisms are used to increase the money supply and prevent deflation. Residents of Ithaca may request interest-free loans of HOURS, organizations may request grants of HOURS, employees of member businesses can accept HOURS as a regular part of their pay, and people may purchase hours into circulation with dollars, from the HOUR bank. Additional money is created to finance administrative costs of the system. The circulation committee of Ithaca HOURS is responsible for deciding how many HOURS to create. So far, Ithaca HOURS are holding their own against the U.S. dollar, and they continue to trade at a ratio of 1 HOUR to 10 U.S. dollars.^b

^aR. Swann and S. Witt, *Local Currencies: Catalysts for Sustainable Regional Economies*. Revised 1988 Schumacher lecture, 1995/2001. Online: <http://www.schumachersociety.org/currencypiece.html> (E. F. Schumacher Society).

^bSee also <http://www.ithacahours.com>. Paul Glover, the founder of Ithaca HOURS, was also very helpful in providing information.

■ THE FRACTIONAL RESERVE SYSTEM

What allows banks to create money is our fractional reserve system. If banks had to keep 100% reserves against the demand deposits they create, then there would be no creation of money. Hence, the reform called for by Soddy, Fisher, Knight, and others was for a 100% reserve requirement. Banks would still provide the convenience of checks and safekeeping, and they would charge for these services. They could still lend other people's money for them and make a profit. But they could not create money any longer.

Exactly how does the fractional reserve system enable banks to create money? Suppose the law required banks to keep 10% reserves against their demand deposits (actually, it is much less). Reserves are either cash or deposits with the Federal Reserve Bank owned by the commercial bank. The bank needs reserves only to settle the difference between daily deposits and withdrawals, which nearly always balance to within a few percent. Therefore, the bank feels that keeping 100% reserves is excessively cautious. It can keep only 10% reserves and meet all imbalances that are statistically likely to ever happen. The "excess reserves" can be loaned at interest, thereby increasing the bank's profits. The government has concurred in this practice and made it legal; it is known as **fractional reserve banking**. It works as long as all depositors do not demand their money at once, as happens in a bank panic (when depositors doubt the solvency of the bank and all rush to get their money out at the same time). To avoid panics, the government set up the Federal Deposit Insurance Corporation (FDIC). If depositors are insured against loss when a bank fails, then they will be less likely to panic and cause the very failure they feared. (They will also be less likely to demand prudence from their bank, but that's another story we leave for later.)

How do banks actually create money? Let's first consider a monopoly commercial bank. Because it is the only bank, it knows that any check drawn against it in one branch will be deposited with it in another branch. When it clears its own check, there is no transfer of money, of reserves, to another bank. Therefore, if it has a new cash deposit of \$100 that counts as reserves, and the reserve requirement is 10%, it can lend out in newly created demand deposits an amount of \$900. People and businesses borrow only what they intend to spend, so it is certain that this \$900 will be spent. Its total additional demand deposits are \$100 in exchange for the new cash deposit, plus \$900 new loans, giving \$1000 in new demand deposits backed by \$100 in new reserves, thus satisfying the 10% reserve requirement. Net addition to the money supply is \$900 worth of demand deposits.

Now let's consider a competitive banking system rather than a single

monopoly bank. Suppose Bank A receives a new cash deposit of \$100. Unlike the monopoly bank, Bank A cannot lend out \$900 because nearly all of the checks written on that amount of new demand deposits will be deposited in other banks, not Bank A. Clearing will necessitate a transfer of reserves to other banks. If it had lent out \$900, it would surely soon have to transfer almost that amount to other banks. But it only has \$100 in new reserves and thus will not be able to meet its legal reserve requirement of 10%.

So how much can Bank A lend as a result of a new cash deposit of \$100? If it safely assumes that all checks drawn on its loans will be deposited in other banks, it can only lend out \$90. Therefore, it still creates money—\$90 in new demand deposits above the \$100 demand deposit in exchange for the \$100 cash. But the process does not stop here. The \$90 of excess reserves safely lent by Bank A end up being transferred to Bank B, which can now safely lend 90% of that, or $0.9 (\$90) = \81 . So now the money supply has gone up by $\$90 + \$81 = \$171$. But then the new \$81 excess reserves of Bank B get transferred to Bank C, which can create new deposits of $0.9 (\$81) = \72.90 . And the process continues in an infinite series, the sum of which turns out to be—can you guess it? Exactly \$900 of new money, as with the monopoly bank, or \$1000 of new demand deposits, remembering the exchange of \$100 cash for a \$100 demand deposit that started the whole process.⁵ The whole process works in reverse when someone withdraws cash (reserves) from the bank. The net result of simultaneous processes of money creation and destruction determines the net growth of the money supply.

■ MONEY AS A PUBLIC GOOD

Money is a collective phenomenon, not a privately owned resource. In a peculiar but very real way, money is a true public good. You might think that if you own money, you can exclude others from using it, but if you did so completely, your money would have no value whatsoever. Money only has value if everyone can use it. And money is certainly nonrival, in that my spending a dollar in no way decreases the value of that dollar for the next person. Since money is a public good, one would expect seigniorage to be public revenue, not private. The virtual wealth of the community could be treated as a publicly owned resource, like the atmosphere or electromagnetic spectrum. But that is not the case. The money supply is privately loaned into existence at interest. The fact that most of our money

⁵If r is the required reserve ratio, then the demand deposit multiplier is this infinite series:

$$1 + (1 - r) + (1 - r)^2 + (1 - r)^3 + \dots + (1 - r)^n = 1/r$$

was loaned into existence and must be paid back at interest imparts a strong growth bias, as well as cyclical instability, to our economy. There is no economic reason why the monetary system must be linked with the private commercial activity of lending and borrowing.

What are the alternatives? Soddy offered three reforms. His first proposal was to gradually raise the reserve requirement to 100%. That would put the private banks out of the money creation business and back into the business of borrowing and lending other people's real money, providing checking services, and so on. Control of the money supply would then belong to the government. How, then, would the government regulate the money supply? Soddy's second policy suggested an automatic rule, based on a price level index. If the price level index is falling, the government should finance its own activities by simply printing new money and spending it into existence. If the price level is rising, the government should cease printing money and tax more than it spends, that is, run a surplus. This would suffice for a closed economy, but for an open economy, one that engages in international trade, the domestic money supply can be increased or decreased by international payments balances. Soddy's third proposal (back in 1926, under the gold standard) was freely fluctuating exchange rates. Currencies would trade freely and directly against each other; an equilibrium exchange rate would eliminate any surplus or shortage (deficit) in the balance of payments, and consequently any international effect on the domestic money supply. Remember our discussion of surplus and shortage in Chapter 9.

Of course, this is not what we have now.

The gold standard has been abandoned, and fixed exchange rate regimes have given way to flexible exchange rates, but not to freely floating exchange rates, which are thought (rightly or wrongly) to be too volatile and disruptive of international trade. (We return to the topic of exchange rates in Chapter 19.) The money supply is largely determined by the commercial banking system, subject to some manipulation, but not control, by the Federal Reserve (the Fed). The **Federal Reserve System** is a coordinated system of district central banks in the U.S. that influences interest rates and money supply.

The Federal Reserve has three tools for manipulating the money supply. First, the Fed can set the reserve requirements, within limits prescribed by law, and thus reduce or expand the supply of money created by banks, as explained above. This tool is used infrequently, because it has large impacts on the financial sector. Second, the Fed can change the interest rate it charges to lend reserves to the commercial banks (known as the discount rate), thus making it more or less profitable for the commercial banks to lend to their customers, and in doing so expand (or limit the expansion of) the money supply. Third, the Fed can conduct open market

operations, directly increasing or decreasing the money supply by buying and selling government securities in the open market. When the Fed buys government securities, it does so by crediting the bank account (at Reserve Banks) of securities dealers. This directly increases the available supply of money by the amount of the purchase. The deposit also increases the bank's reserves, allowing the bank to make more loans and create even more money. When the Fed sells government securities, the money supply contracts.

■ MONEY AND THERMODYNAMICS

Frederick Soddy was a Nobel Prize winner in chemistry, and a great believer that science should be used to benefit humankind. He doubted that this would happen, however, and even predicted back in 1926 the development of the atomic bomb. Why are the fruits of science often badly used? Because, thought Soddy, we have a flawed and irrational economic system. Unless we reform that system, scientific progress will serve to help us destroy the world faster. Soddy spent the second half of his 80-year life studying the economic system. He understood thermodynamics and entropy and the biophysical basis of economics, and forcefully called attention to this interdependence. But he focused his attention mainly on money. Why? Because money was the one thing that did not obey the laws of thermodynamics; it could be created and destroyed. And yet this undisciplined, imaginary magnitude was used as a symbol and counter for real wealth, which has an irreducible physical dimension, and cannot be created or annihilated. Money is the problem precisely because it leads us to think that wealth behaves like its symbol, money; that because it is possible for a few people to live on interest, it is possible for all to do so; that's because money can be used to buy land and land can yield a permanent revenue, therefore money can yield a permanent revenue.

Because of this fallacy, M. King Hubbert recently had to remind us that **exponential growth**—growth at a constant percentage rate—is a transient phase in human history.⁶ The classic example of the power of exponential growth is the story about putting a grain of wheat on the first square of a chessboard, two grains on the second, four on the third, and so on. At the next-to-last, or 63rd, square the board contains 2^{63} grains of wheat, far more than the world's whole wheat crop, and the last, or 64th, square will by itself contain that much again. Hubbert's conclusion was that the world cannot sustain 64 doublings of even a grain of wheat. In our world, many populations are simultaneously doubling—populations

⁶M. King Hubbert, "Exponential Growth as a Transient Phenomenon in Human History," in H. Daly and K. Townsend, eds., *Valuing the Earth*, MIT Press, Cambridge, MA, 1993.

of people, livestock, cars, houses—things much bigger than a grain of wheat. How many times can each of these populations double? How many times can they all double together? A few tens at most, was Hubbert's answer. Our financial conventions, on the other hand, assume that this doubling will go on forever.

This expectation gets played out in reverse when we discount future values to an equivalent present value. We simply run the exponential calculation backward, asking: How much would we have to deposit in the bank today at today's interest (discount) rate in order to have the given future amount at a given future date? This discounting procedure is, as we have seen, at the heart of the financial model of present value maximization, which has displaced the more traditional economic model of profit maximization. The error that bothered Soddy is deeply ingrained in present economic thinking. We have already encountered it in our discussion of why renewable resources are driven to extinction.

It is convenient to dismiss Soddy as a “monetary crank” and to remark what a pity it was that such a brilliant chemist wasted so much of his time on a topic that he was unqualified to think about! This is exactly the treatment that Soddy was given. It was harder to dismiss Irving Fisher and Frank Knight, who also called for 100% reserve requirements, because they were the leading economists of their generation. But their ideas on money were simply classed separately from the rest of their economics, treated as a peccadillo, and were ignored.

Our previous statement—that money does not obey the laws of thermodynamics—needs some qualification. Exchange value is hardly a value if there is nothing for which it can be exchanged. If money is issued without real wealth to back it up, spending that money simply drives up the prices of goods and services, causing inflation, and bringing “real money” back closer into line with real wealth (more on inflation later).

What about virtual wealth? Are there limits to the amount of real wealth people are willing to forego in order to hold money? If not, then the amount of real money in circulation can continue to grow independently of the production of real goods and services. Financial assets are neither money nor real wealth, but they are bought and sold in the market, and people will hold more money to be able to meet their demand for transactions in these assets. In addition, people trade in money itself, using one national currency to buy another, and this similarly increases the demand for money. Both currency speculation and growth in financial assets have increased dramatically in recent years.

The $M-C-M^*$ equation previously showed how money has become less a means for facilitating exchange, more an end in itself. In reality, the $M-C-M^*$ equation has itself been dwarfed by pure currency speculation and trading in financial paper. John Maynard Keynes warned back in

the 1930s:⁷ “Speculators may do no harm as bubbles on a steady stream of enterprise. But the position is serious when enterprise becomes the bubble on a whirlpool of speculation. When the capital development of a country becomes a by-product of the activities of a casino, the job is likely to be ill-done.” While global production of marketed goods and services is roughly on the order of \$30 trillion per year,⁸ the trade in paper purchasing paper (or, more accurately these days, electrons purchasing electrons) with no intervening commodity is almost \$2 trillion *per day*.⁹ This means that the buying and selling of paper assets and currencies, $M \rightarrow M^*$, is more than 20 times greater than exchanges in the real economy! Real enterprise has indeed become a bubble on the whirlpool of speculation. As no productive activity intervenes in these speculative purchases, the sole result seems to be a magical growth in money. But is such growth actually possible indefinitely?

Growth in money is meaningless unless there is a corresponding increase in real wealth, so now we must ask: Does financial speculation lead to growth in real wealth? Some paper-paper purchases are purchases of new stock offerings that do provide financial capital, which can mobilize physical factors of production, but this is only an estimated 4% of stock purchases. Speculation in currency, in which millions of dollars are traded back and forth for very small margins over short time scales, clearly produce nothing. Indeed, such transactions almost certainly contributed to the crises in several South East Asian economies in 1997–1998 as speculators sold off regional currencies, and these crises meant dramatic *decreases* in production from those economies. Yet such speculation would not be undertaken unless some profits were being made somewhere. For example, George Soros, who participated in the financial speculation in South East Asia, is reported to have earned 1 billion pounds speculating on England’s currency in 1995.¹⁰ The only possible explanation is that if those who produce nothing are earning, through speculation, more money that entitles them to more real wealth, then those who actually do produce something must be becoming entitled to increasingly less wealth.

In summary, it would appear that the illusion that money can grow without physical limits results from three things. First, as long as the

⁷J. M. Keynes, *The General Theory of Employment, Interest and Money*. Orlando, FL: Harcourt Brace, 1991, p. 159.

⁸Official estimates based on purchasing power parity (PPP) are on the order of \$40 trillion per year; the numbers for speculation are in nominal dollars, not PPP.

⁹D. Korten, *The Post-Corporate World: Life After Capitalism*, San Francisco: Berrett-Koehler, 1998.

¹⁰W. Greider, *One World, Ready or Not: The Manic Logic of Global Capitalism*, New York: Simon & Schuster, 1997.

production of real goods and services increases, more money is required to pursue them, so growth in money is justified. But such growth cannot, of course, continue forever on a finite planet. Second, as the number or price of financial assets grows, such as through speculative bubbles, then demand for money grows as well, and supply can increase to meet this demand. The fact is, however, that financial bubbles inevitably burst. Third, holders of financial capital see their capital grow because speculation can serve to transfer resources from those who produce to those who merely speculate. Such transfer of wealth has limits, though the limits are obscured by continued economic growth. Thus, the appearance that money is exempt from the laws of thermodynamics is an illusion that can only be maintained while scale is increasing, or the financial sector is expanding relative to the real sector. It remains impossible for real money to grow without limit.

THINK ABOUT IT!

What do you think would happen if a national government tried the same approach to seigniorage as Ithaca HOURS? For example, the government could impose 100% reserve requirements to prevent banks from creating money, award every new entrant to the economy some lump sum of money (perhaps by providing 18-year-olds sufficient money to pay for a college education or start a business), and lend money into existence at 0% interest for socially desirable projects.

BIG IDEAS to remember

-
- | | |
|---|------------------------------------|
| ■ Money as: medium of exchange, unit of account, store of value | ■ Seigniorage |
| ■ Barter; simply commodity production; capitalist circulation | ■ Fractional reserve banking |
| ■ Exchange value vs. use value | ■ Money creation |
| ■ Virtual wealth | ■ Money as public good |
| | ■ Federal Reserve System |
| | ■ Money and laws of thermodynamics |
| | ■ Local currencies |
-



Distribution

We have emphasized that ecological economics is concerned with three issues: the allocation of resources, their distribution, and the scale of the economy. We have seen how the ecological sustainability of the Earth is related to the size or scale of the macroeconomy. We have also explored the economist's meaning of efficient allocation in our discussion of microeconomics and the basic market equation. We then looked at the macroeconomic allocation problem in Chapter 16. But the second issue, distribution and the fairness thereof, has remained largely in the background.

■ PARETO OPTIMALITY

In dealing with allocation, we saw that economics defines efficiency as the Pareto optimal allocation of resources by the market. This definition assumes a given distribution of wealth and income. More specifically, an efficient allocation is one that best satisfies individual wants *weighted by the individual's ability to pay*—that is, by her income and wealth. Change the distribution of income and wealth, and we get a different set of efficient prices (since different people want different things), which define a different Pareto optimum. Because different Pareto optima are based on different distributions of income and wealth, economists are reluctant to compare them; one optimum is as good as another. We saw that a major reason for scale expansion—economic growth—has been to avoid the issue of distributive equality. As long as everyone is getting more from aggregate growth, then the distributive issue is less pressing, at least as a cure for poverty. Besides, the efficiency of the allocation of aggregate growth loses its well-defined meaning (Pareto optimality) once we accept the legitimacy of changing distribution in the interest of fairness. Consequently,

economics has tended to address distribution out of logical necessity, but quickly sets it aside in the interests of political convenience.

Box 15-1

DOES A PARETO OPTIMAL ALLOCATION ASSUME A GIVEN SCALE AS WELL AS A GIVEN DISTRIBUTION?

If we take the concept of scale literally, as in the scale model of a house, to involve only a proportional change in all linear (scalar) dimensions, then we might say that a scale change is simply an increase or decrease in which all proportions remain constant. All relative prices, measuring unchanged relative scarcities, would also remain constant, defining an unchanging Pareto optimal allocation in terms of proportions. This seems to be what standard economists often have in mind. But is it possible to have everything grow in proportion? No, for two reasons. First, if something is fixed, it obviously cannot grow proportionally to everything else. What is fixed from the ecological economist's perspective is the size of the total ecosystem. As the economic subsystem grows, albeit proportionally in terms of its internal dimensions, the ecosystem itself does not grow. The economy becomes larger as a proportion of the total system—what we have called an increase in its scale, meaning size relative to the ecosystem. Natural capital becomes more scarce relative to manmade capital.

Of course, if the economy were to expand to encompass Earth's entire ecosystem (the model of "economic imperialism" in Chapter 3), the scale issue would disappear. In this sense the neoclassical economist's claim that if only all externalities were perfectly internalized then prices would automatically solve the scale problem (in the process of allocating everything in creation) makes sense. But it's a rather utopian point—like Archimedes' boast that he could move the Earth, if only he had a fulcrum and a long enough lever!

The second difficulty, long noticed by biologists and some economists, is that if you scale up anything (increase all linear dimensions by a fixed factor), you will inevitably change the relative magnitudes of non-linear dimensions. Doubling length, width, and height will not double area; it will increase area by a factor of four and volume by a factor of eight. Biologists have long noted "the importance of being the right size." If a grasshopper were scaled up to the size of an elephant, it could not jump over a house. It would not even be able to move, because its weight (proportional to volume) would have increased eightfold, while its strength (proportional to a cross-sectional area of muscle and bone) would have increased only fourfold.

Returning to our example of a house, doubling the scale will increase surfaces and materials by fourfold and volumes to be heated or cooled by eightfold. Relative scarcities and relative prices cannot remain the same.

The answer to our question, does the notion of Pareto optimal allocation assume a given scale as well as a given distribution, appears to be yes. Size cannot increase proportionally because (1) there is a fixed factor, namely the size of the ecosystem, and (2) it is mathematically impossible even for all relevant internal dimensions of the subsystem to increase in the same proportion. In sum, it seems quite true that an optimal allocation assumes a given scale, just as it assumes a given distribution.

Economics prides itself on being a “positive science.” Allocative efficiency is thought to be a positive, or empirically measurable, issue, even though, as we just saw, it presupposes a given distribution. Whether or not the scale of the economy is sustainable is also considered to be a positive issue involving biophysical constraints, although normative questions of conservation for the future and other species are not far below the surface. Distributive equity, on the other hand, is a normative issue. This is the main question addressed to distribution: Is it fair, not is it efficient, or is it ecologically sustainable? The question “Is it fair?” is directly and unavoidably normative, and for that reason alone given minimal attention by the positivist tradition of economics.

But like other sciences, economics assumes certain cultural values. First, the very criterion of objective efficiency, Pareto optimality, embodies an implicit normative judgment—namely, that malevolence or invidious satisfactions are not acceptable. If everyone but you becomes better off and you remain no worse off, the Pareto criterion tells us that is an objective increase in social welfare. But if everyone else is better off except you, and you are an envious person, then you will be less happy than before, even though your absolute situation is no worse. Economists must either make the (false) positive judgment that people are in fact not invidious and jealous, or the (true) normative judgment that envy at another's good fortune is a moral failing rather than a welfare loss.

There is a second reason that economics is less positive than some think. Redistribution can be efficient in the sense of increasing total social utility, yet economists make the value judgment that this kind of efficiency should not count. For example, redistributing a dollar from the low marginal utility use of the rich to the high marginal utility use of the poor will increase total utility to society and is in that sense efficient. The Pareto criterion forbids such interpersonal comparisons and summations of utility. Some argue that the major function of the Pareto criterion was precisely to sterilize the egalitarian implications of the law of diminishing marginal utility¹—a law that economics cannot afford to give up, as we saw in our discussion of demand curves (see Chapter 9).

¹J. Robinson, *Economic Philosophy*, Middlesex, England: Penguin, 1962.

If we admit interpersonal comparisons of utility, then distribution has efficiency implications as well as fairness implications. The extreme individualism of economics insists that people are so qualitatively different in their hermetical isolation one from another that it makes no sense to say that a leg amputation hurts Smith more than a pin prick hurts Jones. If we are all isolated individuals, we can rule out such obviously realistic human characteristics as envy and benevolence. Man as atomistic individual is the *Homo economicus* of neoclassical economics. Ecological economics' concept of the nature of man is "person-in-community," not isolated atom. Community here means community both with other humans and with the rest of the biosphere.

■ THE DISTRIBUTION OF INCOME AND WEALTH

Ecological economics distinguishes between the distribution of income and the distribution of wealth, and between the functional and the personal distribution of income.

Wealth is a stock of assets, measured at a point in time, that is, cash in the bank, plus the market value of bonds, corporate shares, land, real estate, and consumer durables as of a given date. Income is the flow of earnings from these assets, plus the earnings of your own labor power (or human capital), between two dates, that is, over a period of time, usually a year. Labor power is not usually counted as capital because one cannot sell it all at once to another person (short of slavery) but can only rent it for certain durations. Income and wealth are thus two different magnitudes, measured in different units, and distributed differently over the population.² Wealth is usually more concentrated than income. And financial wealth is even more concentrated than wealth in general. In 1989, the top 1% owned 48% of financial wealth. Virtually all of the growth in wealth between 1983 and 1989 in the U.S. went to the top 20%. The bottom 80% was excluded from this growth, and the bottom 40% saw their wealth decline in absolute terms (Table 15.1).

Economics has a theory that explains income, as discussed next, and one that explains the prices of assets (though not entirely, as the "price" of entrepreneurship is a residual), but no theory at all to explain the distribution of wealth among individuals. It is the historical result of whose ancestors got there first, of marriage, of inheritance, plus individual ability and effort, and just plain luck.

²Wealth is measured in dollars (for example) and income in dollars/time. These magnitudes are as different as miles (distance) and miles per hour (speed).

■ **Table 15.1****U.S. PERCENTAGE SHARE OF WEALTH AND INCOME BY PERCENTILE GROUP**

	Percentile Shares			
Year	Top 1%	Next 19%	Bottom 80%	Gini Coefficient
Net Worth (Wealth)				
1983	33.8	47.6	18.7	0.799
1989	37.4	45.3	16.2	0.832
1992	37.2	46.6	16.3	0.823
1995	38.5	45.8	16.1	0.828
1998	38.1	45.3	16.6	0.822
Income				
1983	12.8	39.0	48.1	0.480
1989	16.4	39.0	44.5	0.521
1992	15.7	40.7	43.7	0.528
1995	14.4	40.8	44.9	0.518
1998	16.6	39.6	43.8	0.531
Source: E. N. Wolff, Top Heavy, The Twentieth Century Fund Report, New York: New Press, 1995, p. 67 (years 1983–1992) and E. N. Wolff, Recent Trends in Wealth Ownership, 1983–1998, Working Paper No. 300, Table 2, Jerome Levy Economics Institute, April 2000.				

■ THE FUNCTIONAL AND PERSONAL DISTRIBUTION OF INCOME

Income distributed among people, regardless of its source, is called the personal distribution. Income is also distributed according to how much of total income goes to wages, interest, rent, and profit—the functional distribution. The idea behind the functional distribution is that income is not first created, then distributed. Rather, it is distributed as it is created among the factors combining to create it.

Remember from the circular flow diagram (see Figure 2.4) that supply and demand in the factors market determine the prices of factors—wages, interest, rent, with profit as a residual. Factor prices times the total amount of each factor used yields the functional distribution, usually expressed as percentage of total income going to landowners (rent), laborers (wages), capitalists (interest), and entrepreneurs (profit). Prices of each factor times the amount of the factor owned by each individual yields the personal distribution of income. The amount of each factor owned by each person, including labor power, is the personal distribution of wealth. Therefore, the personal distribution of wealth times the rental price of each type of wealth asset determines the personal distribution of income.

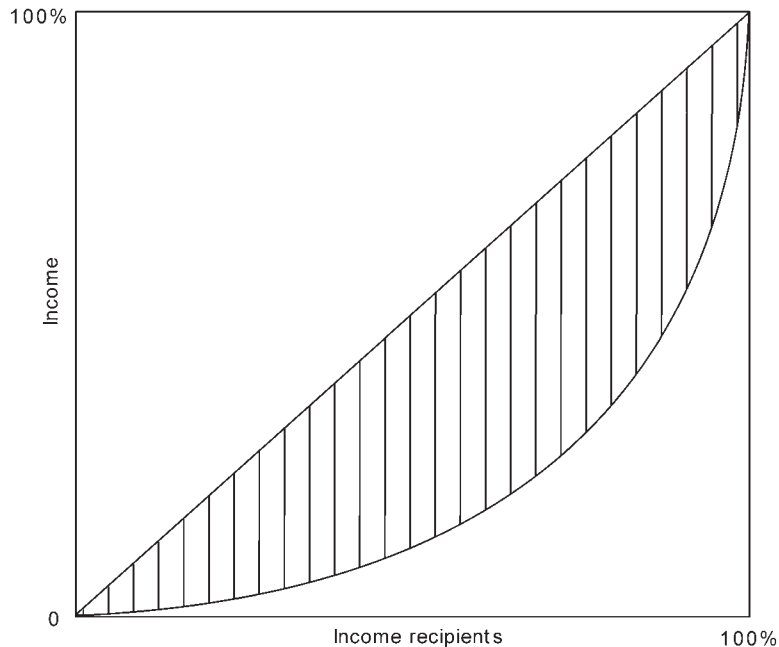


Figure 15.1 • The Lorenz curve. Because the Lorenz curve is in percentages, its shape does not depend on units of measure. It is therefore useful for making comparisons across countries and over time.

■ MEASURING DISTRIBUTION

Although economists have no good theory by which to explain the distribution of wealth and income, they do have useful ways of measuring and describing it statistically.³ One useful representation is the **Lorenz curve**, shown in Figure 15.1. The X-axis shows the number of income recipients in terms of cumulative percentages, from lowest to highest income. The Y-axis shows the percentage of total income. The lengths of the axes are equal, so that when closed in, they make a square.

The Lorenz curve plots the percentage of total income going to each percentage of income recipient. We know that 0% of income recipients will get 0% of the income, and that 100% of income recipients will get 100% of the income, so we already know the two extreme points on any Lorenz curve. If each percentage of the population received the same percentage of the income (i.e., the bottom 20% got 20% of total income, the bottom 70% got 70% of income), we would have perfect equality. The Lorenz curve would be the 45-degree line connecting (0, 0) with (100, 100). But suppose the bottom 80% of recipients get 44% of the income.

³For a clear and insightful exposition, see J. Pen, *Income Distribution*, New York: Praeger, 1971.

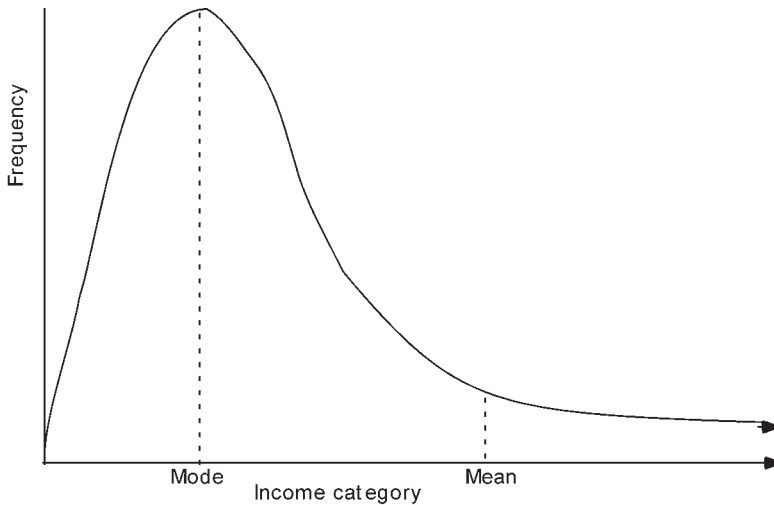


Figure 15.2 • The frequency distribution of income.

That gives us another point, one that lies well below the 45-degree line. If we fill in many points between the extremes, we get a curve shaped like the one in Figure 15.1. The closer the curve to the 45-degree line, the more equal the distribution; the farther away, the less equal. The shaded area defined by the curve and the 45-degree line measures inequality. In the limit, if one person got 100% of income and everyone else got 0%, the Lorenz curve would coincide with the axes and look like a backwards L.

The ratio of the shaded area (between the curve and the 45-degree line) to the total triangular area under the 45-degree line is called the **Gini coefficient**. For perfect equality the shaded area is zero, and consequently the Gini coefficient is 0; for perfect inequality the shaded area takes up the whole area under the 45-degree line, and consequently the Gini coefficient is 1. Values of the Gini coefficient for U.S. wealth and income distribution are given in Table 15.1.

A more familiar statistical description is the common frequency distribution, shown in Figure 15.2. The X-axis shows income category and the Y-axis shows number of members in each income category (frequency). Income distribution does not follow a normal distribution, as does height or many other personal characteristics. Rather, it is highly skewed, with the mode well below the mean, and a very, very long tail to the right needed to reach the top income.

If we wanted to show the maximum income on Figure 15.2, we would need a fold-out extending the horizontal axis by the length of a football field. Graphical representations generally do not capture the extreme

The *Gini coefficient* is used to measure the inequality of the distribution of wealth or income across a population. A Gini coefficient of 1 implies perfect inequality (one person owns everything) and a coefficient of zero indicates a perfectly equal distribution.

inequality at the upper income range. Income categories are frequently truncated at a maximum category of “\$100,000 and over,” where “over” means four orders of magnitude over.

Moreover, these data are just for the United States. The distribution of wealth and income *between* countries is far greater than that found *within* countries.

What is the proper range of inequality in the distribution of income? Surely it is impossible to have one person owning everything, and everyone else owning nothing. Maybe we could have everyone else getting a subsistence wage, and the fortunate one person enjoying the entire social surplus above subsistence. But most people would not consider that fair, even though possible. At the other extreme, few people think a perfectly equal distribution—a Gini coefficient equal to zero—would be fair either. After all, some people work harder than others, and some jobs are more difficult than others. Fairness in a larger sense would require some income differences. There is a legitimate case to be made that differences in distribution provide a socially useful incentive for industriousness and innovation.

Is there a legitimate range of inequality, beyond which further inequality becomes either unfair or dysfunctional? What might such a range be? Plato thought that the richest citizen should be four times wealthier than the poorest. Ben Cohen and Jerry Greenfield, of Ben and Jerry's ice cream fame, at one time reportedly pledged that the highest paid executive would receive no more than five times the salary of the lowest paid employee. Maybe Plato, Ben, and Jerry were wrong, though, and maybe a factor of ten would be better. Or 20 or 50. Currently the acceptable ratio is not defined, and in 1999 in the United States, the *typical* CEO earned 475 times the *typical* worker.⁴ (Of course, comparing the *top* CEOs to the *bottom* workers would yield even a higher ratio.) Ecological economics does not accept the current notion that real total output can grow forever. If the total is limited, then the maximum for one person is implicitly limited. The issue of a proper range of inequality in distribution is therefore critical for ecological economics, even though it has not yet received due attention. The standard economist's effort to keep distribution at bay forever by eternal growth is not a satisfactory solution.

Finally a word on the functional distribution of income. For industrial countries, the division varies around the following: wages = 70%, profits = 20%, interest = 8%, and land rent = 2%. For ecological economics, what is striking is that essentially none of the value of the total product is at-

⁴J. Reingold and F. Jespersen, Executive Pay: It Continues to Explode—and Options Alone Are Creating Paper Billionaires, *Business Week*, April 17, 2000. Online: <http://www.businessweek.com/careers/content/jan1990/b3677014.htm>.

tributed to natural resources or services. Even land rent is mainly a locational premium, not a payment for resources *in situ* or natural services—one more piece of evidence that the flow of low entropy from nature is treated as a free good. If we think of two social classes struggling to divide the pie, we have laborers getting 70% and capitalists, business owners, and landowners together getting about 30%. This division represents a kind of balance of power in the social struggle. Neither side wants to include nature as a participant in production, which would require that nature's services be paid according to their scarcity and productivity.

Even if one wanted to pay for nature's contribution, who would collect on nature's behalf? There is no social class analogous to labor or capital that has an interest in seeing to it that nature's services are properly accounted and paid for. Historically the landlord class may to some extent have played the role of defender of nature's services, but that class hardly exists anymore, and few lament its demise. The government is the biggest landholder in the U.S., and it has followed a policy of cheap resources in order to benefit and ease the tensions between labor and capital. The existing classes, labor and capital, see it in their mutual interest not to share with a third party. Since in reality there is no third party, all that would be necessary is to pay into a fund a scarcity rent for natural resources and services, and then redistribute the fund back to labor and capital, perhaps on the same 70-30 division. This would get the cost accounting and prices right, and improve the efficiency of allocation, without necessarily affecting the distribution. Alternatively, since many of the goods and services provided by natural capital are nonmarket goods, the scarcity rent could go toward supplying other nonmarket goods. The government could either do this directly or could subsidize the private production of such goods. The rent could also be redistributed progressively by financing the abolition of regressive taxes.

■ CONSEQUENCES OF DISTRIBUTION FOR COMMUNITY AND HEALTH

The existing distribution of wealth is not only a precondition for efficient allocation; it is also a fundamental dimension of justice in society. As such, it affects us more directly than we might at first think. Evidence indicates that inequality of income distribution (independently of absolute poverty) has a substantial effect on rates of morbidity and mortality.⁵ The relatively poor have higher incidences of death and sickness than the relatively rich, regardless of the absolute level of income of the relatively poor. The main reason investigators suggest is the extra stress associated with being

⁵See. G. Wilkinson, *Mind the Gap*, New Haven, CT: Yale University Press, 2001.

relatively poor, being at the bottom of a dominance hierarchy. This extra stress is caused by less control over the circumstances of one's life, greater risks of job loss, a lower level of social standing and respect, more frequent experiences of disrespect and shame, with consequent anger and violence. Life at the bottom is more threatening, and the threat often comes from stressful relations with people higher up, including bosses, landlords, and government officials. Stress, of course, has well-known negative direct physiological effects on health.

In addition to these direct effects, inequality has indirect social effects on health. It is more difficult to form friendships across wider income gaps, as well as more difficult to form civic associations when wealth levels and economic interests are very disparate. Lack of friends and civic cohesion is also correlated with ill health. Treating people as atomistic, isolated individuals, unaffected by social relationships, literally makes them sick. As seen from our discussion of Max-Neef's human needs matrix (see Table 13.1), we are persons-in-community, related to each other internally—that is, our personal identity is largely constituted by our relation to others in the community. We are not independently defined entities held together only by external relations of the cash nexus. When these identity-constituting social connections become strained and corrupted by excessive inequality, we get sick more often and we die younger. We are also less happy.

■ INTERTEMPORAL DISTRIBUTION OF WEALTH

Every bit as important as the distribution of wealth and income within a generation is the distribution of resources between generations. However, while people have pondered the distribution of resources within a generation for millennia, the concern for distribution between generations is more recent. For the vast majority of human history, natural resources appeared limitless and technological advance was slow. People had approximately the same resource endowment as their great grandparents had enjoyed, and they expected their great-grandchildren to inherit the same endowment as well. As the pace of technological change accelerated with the Industrial Revolution, change became noticeable from one generation to the next, and people began to expect a better life for their children than they themselves had enjoyed. The "Protestant work ethic" asked people to work hard and invest for their children. At least up through the 1960s, the question most economists asked was: How much consumption should this generation sacrifice for the ever-growing well-being of the next?⁶

⁶E.g., J. Robinson, *Essays in the Theory of Economic Growth*, London: Macmillan, 1968; E. Phelps, Second Essay on the Golden Rule of Accumulation, *American Economic Review*: 793–814 (1965).

However, the onset of the atomic age made it apparent that technological advance had the capacity to bring harm as well as good. Growth in population and per-capita consumption raised the specter of resource depletion. Worsening pollution caused alarm, and ecologists began to worry that many systems were nearing irreversible, catastrophic thresholds. The relevant question was no longer: How much should we sacrifice to make the future even better off? Now it was: How much should we sacrifice to keep the future from being worse off than the present? Curiously, at least in the United States, a culture change was occurring at about the same time. The work ethic was no longer “work hard, live frugally and invest in the future” but rather “work hard and consume as much as possible now.” As a result, savings rates in the U.S. are currently at historic lows and rapidly approaching zero. For much of the population, savings rates are negative.

Should people strive to make the future better off than the present? Do we have at least an obligation to make sure it is not worse off than the present? There are no easy answers to the “appropriate” distribution of wealth between generations. Even a brief survey of philosophies is beyond the scope of this text. We will, however, quickly examine two alternative approaches: the ecological economics approach, based on ethical judgments concerning obligations to future generations (intergenerational justice), and the more mainstream approach in economics that argues for an “objective” decision-making rule (intergenerational allocation).

The Normative Approach of Ecological Economics

Ecological economists generally take the position that intergenerational resource distribution is an ethical issue. The generation into which someone is born is entirely based on chance. There is therefore no moral justification for claiming that one generation has any more right to natural resources, the building blocks of the economy, than any other. At the very least, future generations have an inalienable right to sufficient resources to provide a satisfactory quality of life. The current generation thus has a corresponding duty to preserve an adequate amount of resources. What is adequate depends on both technological and ecological change, both of which are characterized by pure uncertainty (ignorance). How we choose to deal with uncertainty is also an ethical decision.

What does this mean in practical terms?

Renewable and nonrenewable resources are fundamentally different and must be treated separately. An equal distribution of finite nonrenewable resources among a virtually infinite number of future generations would imply no resource use by any single generation. But there is no point in leaving resources in the ground forever, never to do anyone any good, so an upper limit to exhaustible resources for any one generation

might be determined by the waste absorption capacity of the environment. As long as the use of the resource generates waste no faster than the ecosystem can absorb it, the use of exhaustible resources by one generation will not reduce renewable natural capital. Keeping fossil fuel use within such limits would automatically limit our ability to extract other mineral resources.

Even with a limited ability to extract nonrenewable but recyclable resources, each generation would have a further obligation to either efficiently recycle such resources, or at least minimize the generation and dispersion of “garbo-junk” as much as possible to make such resources as intergenerationally nonrival as possible. If existing technologies makes our well-being dependent on nonrenewable resources—as is currently the case—then we are simply obliged to develop substitutes for these resources. One option would be to capture marginal user costs, the unearned income from nonrenewable resources, and invest them toward developing such substitutes.⁷

Renewable resources as fund-services provide essential life-support functions, and these functions clearly must be maintained. Renewable resources as stock-flows must also be harvested at sustainable levels. No one created renewable resources, and therefore no single generation has the right to reduce the amount of the resource a future generation can sustainably consume, suggesting resource stocks must be at least as large as that which provides the maximum sustainable yield. As we saw in Chapter 12, sustainable management of renewable resources in a manner that “optimizes” both stock-flow and fund-service benefits will in general maintain these resources far from any catastrophic ecological thresholds. It is worth bearing in mind that as nonrenewable resources are finite, the exhaustion of these resources is a finite loss to future generations. Renewable resources, as both stock-flows and fund-services, produce a finite flow over an immeasurable number of future generations, and their irreversible loss therefore imposes a perpetual cost to the future.

The “Positive” Approach of Neoclassical Economics

Conventional economists, in contrast, favor an objective decision rule to determine the intergenerational allocation of resources. The problem thus becomes simply a technical one of comparing future benefits and costs with those that occur in the present. Fortunately, the market can tell us the value of things in the future relative to things today, and therefore the market can solve the problem of intergenerational allocation.

⁷For practical guidelines on investing scarcity rents, see S. El Serafy, “The Proper Calculation of Income from Depletable Natural Resources.” In Y. J. Ahmad, S. El Serafy, and E. Lutz, eds., Washington, DC: *Environmental Accounting for Sustainable Development*, World Bank, 1989.

Intertemporal Discounting. How does the market reveal future values? Where adequate financial markets exist, people can borrow money today at interest, which requires them to pay back more money in the future. The fact that people engage in this activity reveals that people prefer things now rather than in the future, and economics must respect people's preferences.

There are three basic reasons why people might prefer things now to things in the future. First, people may simply be impatient. Anyone who goes into interest-bearing debt to purchase something is willing to sacrifice a greater quantity in the future for a smaller quantity now. Some of this impatience may come from uncertainty—no one knows for sure if he or she will be alive tomorrow, so why not eat, drink, and be merry today? This rationale for discounting is known as the **pure time rate of preference (PTRP)**.

Second, for things that reproduce, it makes sense that a given quantity in the future would be worth less than a given quantity now. For example, a handful of seed corn now can become a bushel of marketable corn a few months from now, so if growing corn was risk-free and required no resources or effort, then a handful now would be at least as valuable as a bushel in a few months. Of course, growing corn is risky and requires land, labor, and resources. However, market goods (in this case, seed corn) can be sold for money. Investing the money earned from the sale of the seed corn in an insured bank is not very risky, and for the individual investor basically requires no further resources or labor. As we explained in Chapter 10, this rationale for discounting is known as **opportunity cost**, the lost opportunity to invest. If money is a substitute for any other resource, then we should give more weight to any resource today over the same resource tomorrow.

Third, the economy has grown fairly steadily for hundreds of years. People therefore expect that they will be richer in the future than they are today. Just as an extra \$1000 provides less utility to Bill Gates than to a pauper, the law of diminishing marginal utility means that money in the future will be worth less than the same amount of money today. This is sometimes referred to as the “richer future” argument.

In general, this process of valuing the future less than the present is known as **intertemporal discounting**, introduced in Chapter 10. Business people explicitly discount the future when making investment decisions, and mainstream economists argue that people automatically apply this concept to all of their purchase decisions. As a result, they conclude, the market efficiently allocates goods between the present and future.

What's more, if intertemporal discounting leads to allocative efficiency in the market, then it should also be applied to nonmarket investments. For example, one of the biggest nonmarket decisions we face today is how

to deal with global climate change. Virtually all economic analyses of climate change place a lower weight on future costs and benefits than on present ones. These analyses look at different policy scenarios and for each sum up the present costs and benefits with discounted future costs and benefits to arrive at a **net present value (NPV)**. NPV basically tells us what present and future costs and benefits are worth to us *today* (not to the future *tomorrow*), which implies that future generations have no particular right to any resources, and we have no obligation to preserve any. Under this type of benefit-cost analysis, the higher the NPV relative to required investments, the better the project.

Such analyses can carry a great deal of weight as society decides how to address some of the most pressing problems we now confront. The central importance of the discount rate in determining the outcome of such analyses means the topic deserves our attention.

Box 15-2

INTERTEMPORAL DISCOUNTING AND GLOBAL CLIMATE CHANGE

Policy makers seeking an objective decision-making tool for resolving the problems of global climate change have turned to economists. Economists typically respond to the problem by creating complex models of future costs and benefits of climate change, and compare these to the costs of mitigation measures in a cost-benefit analysis designed to calculate net present value. Not surprisingly, analyses using a fairly high discount rate find that future damages from global warming do not justify efforts today to reduce greenhouse gases. The 6% discount rate used in one study would have us believe that we should not invest \$300 million today to prevent \$30 trillion (a rough estimate of today's global GNP) in damages in 200 years.^a A similar study using a 2% discount rate in contrast finds that we should make substantial investments now to reduce the impacts of global warming in the future. Similarly, what we decide to do with an old-growth forest that supplies a small but steady flow of benefits forever if left intact or a large, one-time return if it is clear-cut will depend on the discount rate we choose.

A frequently asked question is: Does a higher or lower discount rate favor the environment? For a given fishery or mine, as we have seen, higher discount rates increase the intensity and rate of exploitation and are thus bad for the environment. But a higher interest rate (discount rate) slows down aggregate growth in GNP and throughput, thus easing pressure on the environment. In terms of evaluating a given project, a high discount rate favors projects whose costs are mainly in the future and whose benefits are in the present, and penalizes those whose costs are in the present with benefits in the future. Most issues in economics are not simple, and that certainly holds for discounting.

In many such models, the choice of a discount rate may be the single most important factor, yet respected economists addressing the same problem use dramatically different rates and arrive at dramatically different results. Are such models actually objective decision-making tools?

"30 trillion is a number that's hard to wrap your mind around. Putting it into perspective: 30 trillion seconds is slightly less than a million years (951,294 years, to be precise)!"

Discounting Reconsidered

We have already explained why intertemporal discounting can make sense for the individual and for market goods. We must now examine whether it also makes sense for society and for nonmarket goods.

We saw why individuals might have a pure time rate of preference: People are impatient; they don't live forever; possessions can be lost, destroyed, or stolen, and opportunities disappear. A reasonable individual may discount the future for any one of these reasons—why should I pay money now to reduce damages from global warming that will only occur after I am dead?—but the same logic does not apply to society. Relative to the individuals of which they are composed, societies are immortal, and uncertainties are averaged out. For this reason, there is, in fact, fairly wide consensus within the economics profession that social discount rates should indeed be lower than individual discount rates. The **social discount rate** is a rate of conversion of future value to present value that reflects society's collective ethical judgment, as opposed to an individualistic judgment, such as the market rate of interest.

When it comes to the opportunity cost of capital, however, the consensus changes. Financial capital does function as a productive asset, and if we have it now instead of in the future, we have the opportunity to invest it in productive activities that will increase the quantity of market goods in the future. There are a number of important issues we must bear in mind, however.

First, the real value of money can only grow if the production of goods and services that money can acquire also grows, and we know that the production of goods and services cannot grow forever on a finite planet. While there may always be some areas that are growing, justifying a discount rate for the individual, the economy as a whole cannot grow indefinitely, in which case a social discount rate into the indefinite future may be inappropriate.

Second, we must recognize that many investments are "profitable" because we ignore many of the costs of production. We know that all human productive activities use up natural resources and return waste to the environment, and these costs of production are often ignored. Many of these

costs, such as contributions to global warming, have greater impacts on future generations. Thus, ignoring costs to future generations allows us to earn higher returns on investments. We then use these higher returns to justify the fact that we ignore costs imposed on future generations! Even in the short run, then, it seems that market-determined interest rates are not suitable discount rates.

Related to the opportunity cost of capital is the argument that the future will be richer than the present because of investments we make now. Of course, if the economy does not continue to grow, the future will not be richer, and if we deplete our natural resource stock, there is every chance the future will be poorer. In fact, measures such as the ISEW suggest that society is already growing poorer, not richer, if we take into account external costs. Also, if we believe that natural capital must be treated separately from manmade capital (because they are complements rather than substitutes and natural capital has become the limiting factor), then the decline in natural capital, coupled with the law of diminishing marginal utility, suggests we should apply a negative discount rate to natural capital. At the very least, we might consider applying a positive discount rate only to those goods and services that are actually highly fungible with money—that is, that can be converted into money and back again with little effort. Basically, this would mean that we should only discount market goods and services.

Third, there are only finite opportunities for productive investment in an economy; investments, like other things, show diminishing marginal returns. For example, someone borrows money to explore for oil, and someone else borrows money to build a car factory. The next person to borrow money to explore for oil will have fewer places to explore, and therefore will expect lower returns. The next person to borrow to build a car factory will face a more saturated market, and therefore expect to sell fewer cars. As more and more people borrow to invest and opportunities are used up, the returns on investments can be expected to decrease, ultimately falling to zero. More likely, if interest rates are determined in the market by the supply and demand of money for borrowing and returns on investments, a balance will be reached in which investors cannot afford to pay high enough interest rates for consumers to be interested in deferring consumption. Theoretically then, in a perfect market situation, the opportunity cost of capital at the margin will just equal the PTRP of the existing generation. (Obviously, future generations cannot take part in financial markets any more than they can in any other market.) However, a high PTRP means consumption will be high and investment and growth low,⁸ and a low PTRP implies the op-

⁸Returns on investments will be high, but the total amount invested will be low, and hence growth will be low.

posite. Thus, if we allow the market interest rate to determine the discount rate, there would theoretically be an inverse correlation between discount rate and economic growth, the exact opposite of what would justify the “richer future” rationale for discounting.

Box 15-3 DISCOUNTING, PSYCHOLOGY, AND ECONOMICS

Economists argue that economics is the science of human preferences, so human preferences must be respected. If people value the present more than the future, we must respect that. The question is: Do people *exponentially* discount the future? While it is true you may prefer to have something now rather than the same thing in 5 years, how do you value something that happens 100 years in the future compared to 105 years? If you heard that global warming was going to result in the deaths of 50 million Bangladeshis in 125 years, would that make you feel only half as bad as finding out it would actually kill those 50 million Bangladeshis in 100 years? If you are like most people, you would feel equally bad in either case, yet influential economic models of the impacts of global warming really do assume we would care only half as much about those deaths if they occurred 25 years later.

Empirical studies show that people do discount the future, but do not do so exponentially. We might give more weight to what happens now than to what happens in the near future, but we are nearly indifferent between the same outcome occurring at different times in the more distant future. One approach to modeling this type of behavior mathematically is known as **hyperbolic discounting**. While this precise formulation of intertemporal discounting may not be perfect, considerable evidence suggests it is far more representative of human preferences than exponential discounting. While the approach was first introduced over 30 years ago and has gained increasing attention in the last few years, it is still fairly rare to see it in use.

An increasing number of studies in the area of Psychology and Economics are finding that the traditional economic assumptions of human behavior are often seriously flawed. If economics is serious about becoming the science of human preferences, it would do well to pay more attention to how humans really behave.^a

^aFor a good introduction to the field of Psychology and Economics, see M. Rabin, *Psychology and Economics*, *Journal of Economic Literature* 36(1): 11–46 (March 1998).

Finally, many economists argue that technology is the driving force for economic growth. Not only does technology ensure we won't run out of resources and the economy won't stop growing, but it offers yet another reason to discount the value of resources in the future. Technology is

likely to develop substitutes for natural resources. When these substitutes become available, the resources they replace will lose value. Therefore, they will be worth less in the future than they are today. After all, hasn't oil largely replaced coal, and haven't fiber optics replaced copper in many uses? However, technology ultimately complements resources and can never completely replace them. Some 150 years ago, oil had little value. Today, it is an integral part of an overwhelming number of industrial processes and products. As we saw in Chapter 5, we are actually developing new uses for oil and other raw materials faster than we are developing replacements, again suggesting that the value of raw materials will increase in the future, not shrink.

What in the end can we say about discount rates? They do make sense for individuals in the short run. For some small-scale, short-term social projects, they may also make sense. However, justifications for discounting the future on a large scale and over long-time horizons are questionable at best. **Intertemporal allocation** is the apportionment of resources across different stages in the lifetimes of basically the same set of people (the same generation). Discounting can make sense for someone efficiently allocating resources intertemporally. But as we lengthen the time period we are more and more talking about different people (different generations) and less and less about the same people at different stages in their lives. **Intertemporal distribution** is the apportioning of resources across different generations, across different people. Distribution is fundamentally different from allocation, and, consequently, justice replaces efficiency as the relevant criterion for policy when time periods become intergenerational.

BIG IDEAS to remember

-
- | | |
|--|---|
| ■ Pareto optimality | ■ Gini coefficient |
| ■ Role of scale and distribution in defining Pareto optimal allocation | ■ Inequality and health |
| ■ Income distribution vs. wealth distribution | ■ Intertemporal distribution vs. intertemporal allocation |
| ■ Functional vs. personal distribution | ■ Discounting and net present value |
| ■ Social limits to range of inequality | ■ Pure time rate of preference (PTRP) |
| ■ Lorenz curve | ■ Individual vs. social discount rates |
-



The IS-LM Model

We have now explored three of the major issues with which macroeconomics is concerned: gross national product (GNP), money, and distribution. We questioned the appropriateness of GNP as the desirable end for economic policy, and emphasized the importance of a just distribution as a desirable end, but said little about policies for attaining these ends. In this chapter, we examine the policy tools at the macroeconomist's disposal that can help us attain an economy with sustainable scale, just distribution, and efficient allocation.

Of course, to know how policies work, we have to know how the macroeconomy works. One way of doing this might be to build on microeconomic principles to construct a model in which supply and demand of all goods and services balances simultaneously. This approach would extend the basic market equation presented in Chapter 8— $MU_{xn} \cdot MPP_{ax} = MU_{yn} \cdot MPP_{ay}$ —into a general equilibrium model encompassing all goods (x, y, z, \dots), all commodities (a, b, c, \dots), and all consumers (n, m, o, \dots). Such a model can easily become overwhelming. A thousand simultaneous equations with a thousand unknowns is hard to come into mental contact with. It does show that everything depends on everything else, which is interesting and usefully humbling, but it is also crippling from a policy perspective to have to face the implication that in order to predict anything, you have first to know everything. But a smaller system of two or three or five especially important aggregate sectors interacting through two or three or five simultaneous equations that reflect key behavior can aid the understanding and give basic policy insights. This is the kind of model that most macroeconomists have sought. They still look at the whole economy, but they divide it into fewer but larger aggregate

sectors than does the general equilibrium model of microeconomics.¹ A model of this type, first offered in 1937 by Sir John Hicks² and now called the IS-LM model, has proven to be a good “two-digit” compromise between completeness and simplicity. It has become the “workhorse” model in macroeconomics. Below we will explain this model, and then discuss its applications to ecological economics.

The model divides the economy into two sectors: the real sector (dealing with national income, savings, investment, rates of productivity of capital, government spending, taxation, etc.) and the monetary sector (money supply, interest rates, demand for liquid cash balances). The real sector reflects the theories and insights of classical economics, and the monetary sector reflects the insights of John Maynard Keynes, which in 1937 were still quite new. The model seeks to explain how the interdependent behavior of consumers and savers, lenders and borrowers, and monetary authorities interact to determine the level of national income and the rate of interest.

Box 16-1 THE QUANTITY OF MONEY THEORY OF INCOME

Another way of relating the real and monetary sectors in an aggregate way is through the “identity of exchange,” $MV = PQ$, where Q is quantity of final commodities sold to households, P is average price of exchange, M is stock of money, and V is velocity of circulation of money (number of times an average dollar is spent per year on final goods and services). Since by definition $V = PQ/M$, the equation of exchange is an identity or truism. To the extent that V is a constant or slow to change, reflecting stable payment habits and settlement periods, the identity becomes the “quantity of money theory of income,” stating that changes in PQ are proportional to changes in M . If the economy is at full employment, it will be very hard to increase Q in the short run, and the change in PQ will be mainly a change in P —i.e., inflation. Historically M and P have often moved in direct proportion, yielding a quantity of money theory of the price level.

¹Our measures of the two most basic magnitudes of macroeconomics, GNP and money, are too dialectical and uncertain to be able to support exact calculations implicit in complicated models. As Oskar Morganstern remarked in his classic *On the Accuracy of Economic Observations*, “economics is a two-digit science.”

²J. Hicks, Mr. Keynes and the “Classics,” *Econometrica* 5(2) (April 1937).

■ IS: THE REAL SECTOR

Let's begin with the real or classical sector. The real sector is in equilibrium when the supply of goods by firms is just equal to the demand for goods by households (the lower half of the circular economy in Figure 2.4). Of course, the demand for goods by households is determined by their income—the money firms pay households for their factors of production (e.g., labor), and the supply of goods is determined by the firms' employment of those factors of production (the upper half of the circular economy in Figure 2.4). In equilibrium, income (Y) equals output (GNP). Remember from the circular flow diagram in Figure 2.5 that the equilibrium condition for the continued flow of national income at a given level is that leakages equal injections. In the simplest case, the leakage is savings (S) by households, the new injection is investment (I) by firms. Therefore, the equilibrium condition for the real sector is $S = I$.

But how do S and I get determined? Let r be the interest rate and Y be national income (GNP). In equilibrium, income paid to the factors of production will just equal the output of goods provided by those factors of production, and the income will be used to purchase the output. Savers (i.e., households) will save more if their income Y is higher than if it is lower. Also savers will save more with a higher interest rate r than with a lower one. Investors (i.e., firms) will borrow and invest more if the interest rate is lower, and if income is higher. In other words, savings is some function of the interest rate and national income, say $S = S(r, Y)$. Likewise, investment is some different function (representing the behavior of firms instead of households) of the same two variables, say $I = I(r, Y)$.

In equilibrium,

$$S = I$$

or

$$S(r, Y) = I(r, Y)$$

The above equation is satisfied for all combinations of r and Y such that $S = I$, that is, such that savers and investors are both satisfied.

There are many such combinations of r and Y —we have only one equation with two unknowns. Plotting all the combinations of r and Y that result in $S = I$ gives us Hicks' so-called IS curve, short for $I = S$ (Figure 16.1). To reiterate, this is the combination of r and Y that leads to equilibrium in the real sector: leakages (savings) equal injections (investment), and the demand for goods is just equal to the supply.

Why is the IS curve drawn with a negative slope? Businesses will only borrow money to invest if they can make sufficient returns from the investment to pay off the loan plus interest and still have money left over for profit. A businessperson would not borrow money at 6% interest to invest

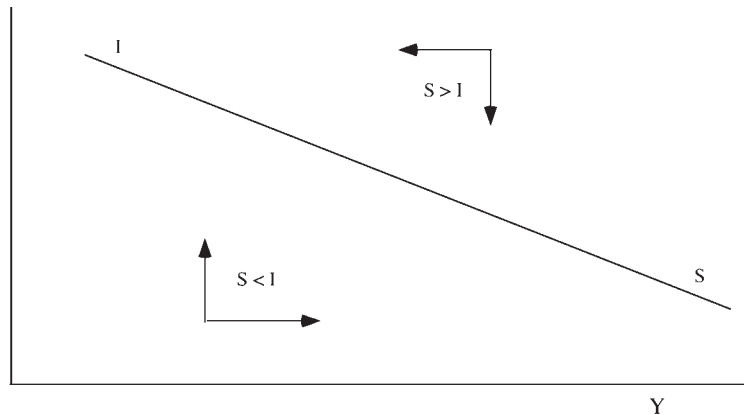


Figure 16.1 • The IS curve: At low (high) levels of income Y , there is a correspondingly low level of savings. At high (low) rates of interest r , there is a low (high) demand for investment. Therefore, at low (high) levels of Y , savings and investment will only be in equilibrium when r is high (low). If interest rates are too high for a given level of income, savings (leakages) will be greater than investment (injections). Firms producing more goods than people consume reduce production and the economy shrinks. Firms with excess capacity borrow less, so the price of borrowing (the interest rate) falls to clear the market. The converse is true when investment is greater than savings.

in a project expected to return 5% annually on the investment, but would borrow at a rate of 4%. As interest rates go down further, more and more investments become profitable, and therefore more investments are made. More investment leads to higher Y . Therefore, high interest rates lead to low rates of investment and low income, while low interest rates lead to high rates of investment and high income. Savings, in contrast, are probably determined more by income than by interest rates.³ When income is low, all money has to be spent simply to meet basic consumption needs, and none is available to save. As income increases, basic consumption needs require a smaller percentage of income, and more is left over to save, so in general higher incomes lead to greater savings.

Combining these two tendencies, we would expect that at high levels of income when lots of money is being saved, investors will only borrow all that money to invest if interest rates are low. At low levels of income, savings are low, and unless interest rates are high, businesses will demand more money than is being saved. For some readers, a diagrammatic explanation for the negative slope of the IS curve will be easier to follow (Figure 16.2).

³Savings rates should also increase as interest rates increase, as under these circumstances savings yield higher returns, and consumption has a higher opportunity cost. But empirical evidence does not support this. One reason may be that if savers are motivated by attainment of a target future amount, a higher interest rate would mean less saving is needed to reach the future target.

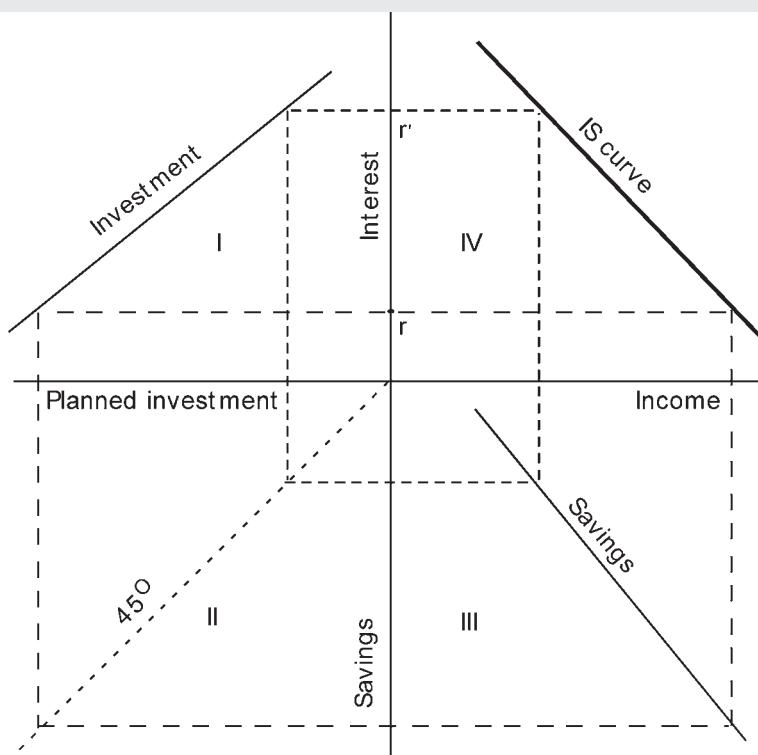
Box 16-2 A GRAPHIC DERIVATION OF THE IS CURVE

Figure 16.2 • A graphic depiction of the derivation of the IS curve.

Figure 16.2 illustrates one way to derive the IS curve. Quadrant I shows the basic relationship between interest rates and investment—high interest rates lead to low levels of investment, and low interest rates lead to high levels of investment. There is a negative correlation between interest rates and investment, as depicted on the graph. Quadrant III shows the relationship between savings and income. Poor people must spend all their income to meet their basic needs and cannot afford to save anything. As income increases, people begin to save, so there is a positive correlation between income and savings, as depicted on the graph. We know that in equilibrium (which is what the IS curve depicts), investment equals savings. Quadrant II contains a 45-degree line that allows us to translate a given rate of investment from quadrant I to an identical rate of savings in quadrant III. Quadrant IV shows the relationship between income and interest in a real sector equilibrium.

If we start with interest rate r , we can see from quadrant I that this

will correspond to level of investment I . Dropping a line down from I in quadrant I to the 45-degree line then across to quadrant III lets us determine the equilibrium level of savings S in quadrant III. We can see from quadrant III that this level of savings corresponds to income Y . The point in quadrant IV where income Y meets interest rate r gives us one point on the IS curve. If we do the same for interest rate r' , we have two points on the IS curve. We see that a low level of interest leads to equilibrium only when income is high, and a high level of interest leads to equilibrium when income is low. Perhaps the simplest way to remember this relationship is that at low interest rates, investment will be high, and high investment leads to high income.

Macroeconomics does not assume that the economy is always in equilibrium, but it does assume that it is at least moving in that direction. For example, we know that if r rises, then savers will try to save more, and investors will be less willing to borrow and invest, leading to a condition in which planned $S > I$. In other words, savers want to save more than investors want to invest at the new higher r . This will have two impacts. First, leakages will be greater than new injections, causing income to fall. Second, savers earn interest on their savings because investors are willing to pay that interest to borrow the money. Interest is the price of money. When the supply of savings is greater than the demand for savings, the interest rate must fall. The mechanism is the same as for any other good, as explained in Figure 9.2. At a lower Y savers save less, and at a lower r investors borrow more, and both r and Y continue to fall until I again equals S at a lower income (Y) and higher interest rate (r) than before. If the interest rate falls, then investment will become greater than savings, and adjustment will occur in the opposite manner. These dynamics are indicated by the arrows in Figure 16.1.

■ LM: THE MONETARY SECTOR

We turn now to the monetary sector and the LM curve, which shows the levels of income (Y) and interest (r) at which the demand for money balances (money held by people) equals the supply of money. We must first ask why individuals want to hold money balances when they could easily exchange them for real assets. From our earlier discussion the answer is clear—people hold cash balances to avoid the inconvenience of barter. Keynes referred to this as the **transaction demand for money**. He also spoke of a related **liquidity preference**, meaning that, other things being equal, people prefer liquid assets to “frozen” assets because they are so easily convertible into anything else, therefore fungible. Money is the most liquid of all assets. But of course other things are seldom equal, and the cost

of holding wealth in the form of fungible money is to forego the interest that could be had by lending the money, or the utility from spending it on a real asset or commodity. Yet if too much of your wealth is tied up in nonliquid forms, you will have difficulty making necessary transactions in a timely manner and meeting unexpected contingencies. The higher the national income, the more need for transactions and consequently the more money everyone will need (a higher transactions demand for money), and the higher the interest rate will have to be to induce owners of those transactions balances to sacrifice liquidity by lending them.

The demand for money balances (DM) thus depends on r and Y , by means of a relation of liquidity preference (L). Thus:

$$DM = L(r, Y)$$

The equilibrium condition is that the demand for money equals the supply of money (SM):

$$DM = SM$$

What determines the supply of money? In earlier times it was the geology and technology of gold or silver mining (a part of the real sector!), but today we have not real commodity money, but fiat or token money, controlled by the government through the private banking sector, as discussed in Chapter 14. For simplicity, the model usually takes SM as given by the government, equal to M . Thus:

$$L(r, Y) = M$$

is the equilibrium condition for the monetary sector, and the LM curve consists of all those combinations of r and Y such that the aggregate demand for cash balances is equal to the given money supply (Figure 16.3).

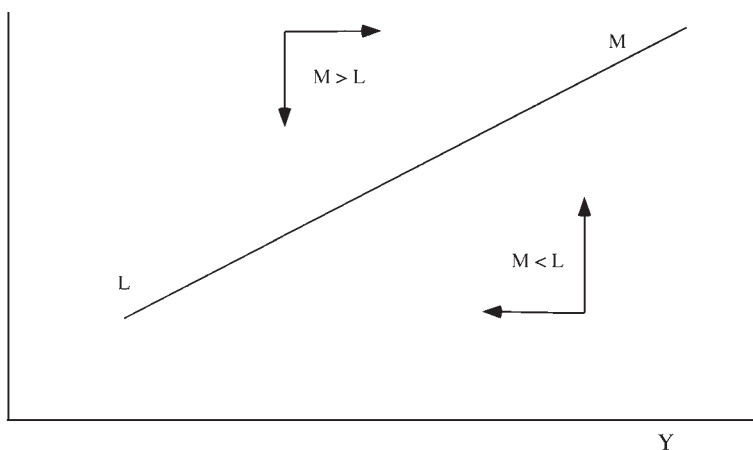


Figure 16.3 • The LM curve.

Since we have one equation with two unknowns, we cannot get unique values of the unknowns, but we can determine all those combinations of r and Y that satisfy our one equation.

Box 16-3 THE FEDERAL RESERVE BANK

In the U.S., money is *not* controlled by the democratically elected government but rather by the Federal Reserve Bank (the Fed), a nonelected “branch” of government. Decisions concerning monetary policy are decided upon by a seven-member board of governors with lesser influence by the directors of the 12 regional Federal Reserve Banks. Members are appointed by the President (with Senate approval) for 14-year staggered terms, and the chair and vice chair are appointed for 4-year terms. Despite the importance of monetary policy in the functioning of our economy, the system is specifically designed to insulate the Fed from pressure by democratically elected politicians! The Fed is not expected to respond to voters. This does not mean that the Fed does not have a constituency to which it feels responsible, as we shall discuss later.

Why is the LM (short for $L = M$) curve drawn with a positive slope? Let's ask ourselves what are the consequences on the interest rate (r) of an increase in income (Y). A larger Y means a larger volume of transactions and will cause a greater demand for transactions balances. This will lead to a higher r to compensate for the loss in liquidity from lending those balances. Thus, a higher Y will require a higher r for money holders to again be satisfied (for L to equal M). Hence the positive slope of the LM curve. This relationship seems to be sufficiently clear that a more detailed graphic explanation is unnecessary.

When the monetary sector is out of equilibrium, what specific mechanisms drive it toward equilibrium? Say the monetary authority increases the money supply, so there is more money available than people actually desire to hold at the existing interest rate—that is, $M > L$. Excess money is used to buy bonds and other nonliquid interest-bearing assets (which we will refer to jointly as “bonds” for convenience). More money chasing the same number of bonds will drive up their price.

There are many types of bonds, but in the simplest case, when someone buys a bond, they are paying something now to receive a fixed amount when the bond matures. For example, if I pay \$50 today for a \$100 bond that matures in 10 years, my rate of return is about 7.2%. An increase in the money supply might drive the price of the bond up to \$60, which provides a rate of return of only 5.24%. The higher the price for a bond, the

lower the interest rate on that bond. Hence, an increase in the supply of money increases the demand for bonds and drives down the interest rate.

At lower interest rates, there is less opportunity cost to holding money, and hence a higher demand for money. Lower interest rates also stimulate investment, leading to economic growth, which further stimulates the demand for money. The result is a new equilibrium at lower interest rates and higher income. A decrease in the money supply of course leads to the opposite result. These forces are illustrated by the arrows in Figure 16.3.

■ COMBINING IS AND LM

Putting the IS and LM curves together lets us determine a unique combination of r and Y (namely r^* , Y^*) that satisfies both the $S = I$ condition of the real sector and the $L = M$ condition of the monetary sector (Figure 16.4). The point of intersection is the only point common to both curves, the only point that gives equilibrium in both real and monetary sectors. Basically we now have two simultaneous equations determining two unknowns, r and Y .

The **IS-LM model** is used in a comparative statics⁴ way to analyze the effect on r and Y of changes in the underlying determinants—namely,

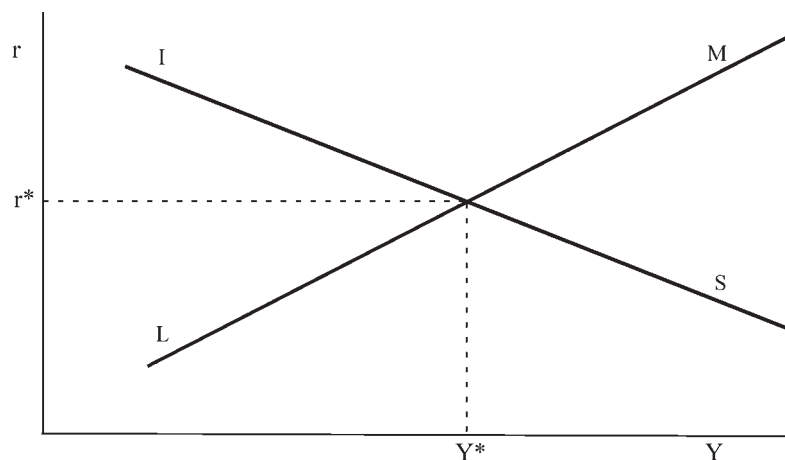


Figure 16.4 • The IS-LM model.

⁴Comparative statics is the analysis of what happens to endogenous variables in a model (in this case, r and Y) as a result of change in exogenous parameters (in this case, propensity to save, efficiency of capital investment, and liquidity preference). It compares the new equilibrium variables with the old ones, without explaining the precise dynamic path leading from the old to the new equilibrium.

propensity to save, the efficiency (productivity) of capital investment, and liquidity preference. Of particular interest to policy makers is the impact of policy variables on r and Y —namely, government expenditure, taxation, and the money supply. Each of these changes results in a shift in one of the curves, and consequently in a move along the other curve to a new intersection point. What we are really interested in, then, is how the economy moves toward equilibrium after policies or outside (exogenous) changes push it away.

■ EXOGENOUS CHANGES IN IS AND LM

First let's look at some exogenous changes, those that are basically independent of fiscal and monetary policy, and therefore outside the IS-LM model. Consider an increase in the marginal propensity to save. Such a change in savings rate might result from fears of an economic downturn that would lead to lower wages and greater unemployment. (We might hope that people might one day simply decide to consume less in order to protect the environment!) In either case, people decide to save more and spend less of their extra income. This means that now $S > I$ for all the combinations of r and Y on the IS curve. We need a new IS curve for which $S = I$ again. If people save more at every r , this means $S > I$, or leakage greater than injection, so the flow of income will fall to the level at which $S = I$ again.

Even though people are saving a larger fraction of their income, they will end up having a smaller income out of which to save, with the result that S will be the same, only now the product of a higher rate of savings per dollar times a lower number of dollars in income. The very act of saving more and spending less will have caused a fall in aggregate income to the extent that the lower income times the higher fraction saved gives the same total savings, equal to the unchanged level of investment. So every r will be paired with a smaller Y in the new IS than with the old one. The IS curve will have shifted to the left. The new intersection with LM will occur at a lower r^* and lower Y^* than before. An increase in the marginal propensity to save will therefore result in a fall in national income and a fall in the interest rate.

Of course if the marginal propensity to save increases, then the marginal propensity to consume must decrease. As people consume less, businesses will be unable to sell their goods, leading to unplanned accumulations of inventory. This in turn will lead businesses to reduce planned investment and production—perhaps laying people off. Unemployment resulting from layoffs further decreases consumption, requiring another round of adjustment, lowering Y still more. The final result, when $S = I$ again, may well be that S will be lower than the level at which

we started. Thus, the effort of everyone to save more in the aggregate could result in everyone actually saving less—the so-called **paradox of thrift**. In such a case, a higher savings rate induced by fear of recession could itself cause a recession—a self-fulfilling prophecy.

Now suppose there's an increase in the efficiency of investment (an increase in the marginal productivity of capital), thanks to a new invention. For example, many people claim that this is exactly what has happened in today's "new economy," in which information technology is said to have increased productivity. This would increase I , so that $I > S$ now along the old IS curve. With $I > S$, injections are greater than leakages out of the circular flow, so the flow of income will grow until $S = I$ again. The new IS curve will have a higher Y for each r . The curve shifts rightward. The new equilibrium occurs with a higher Y^* and a higher r^* . An improvement in the marginal efficiency of capital raises both income and the interest rate.

Finally, turning to the LM curve, suppose there was an increase in liquidity preference, so that $L > M$. Such a change could result from increasing uncertainty over future economic conditions, and a desire by people to be prepared for the unforeseen with cash on hand. Alternatively, the deregulation of banking in the United States during the mid-1970s allowed certain checking accounts to pay interest. This reduced the opportunity cost of holding money, and therefore probably increased the liquidity preference as well. In either case, for any income and associated level of needed transactions balances, there is a greater willingness to hold those balances, to hold more than strictly needed. It takes a higher r to induce holders of money to lend. Consequently each level of Y will be associated with a higher r on the new LM than on the old one. The new LM will shift upward. The new equilibrium will occur at a higher r^* and lower

Box 16-4 JUNK BONDS AND TIMBER COMPANIES

Seemingly abstract things like interest rates on bonds and Wall Street transactions can affect real economic production and the provision of environmental services. For example, during the 1980s, hostile takeovers and the introduction of junk bonds on Wall Street led to deforestation on the West Coast. How did this happen? Mergers, when two companies join together, and acquisitions, when one company purchases another, are a normal part of corporate activity in the U.S. Sometimes, however, one company does not wish to be taken over by another. For example, mergers and acquisitions (M&A) focused primarily on short-term profits

can weaken or destroy the company being acquired, leading to massive layoffs.

Reasonably enough, managers not eager to be laid off will be opposed to a merger, and under such circumstances, takeover attempts are “hostile.” One company acquires another through the purchase of a controlling share of stocks. As soon as someone starts buying enough stock to control a company, the stock price rises. A company threatened by a hostile takeover can attempt to defend itself by repurchasing its own stock, driving up the price of stock even further. To get enough money for a hostile takeover, the company attempting the takeover can offer high-yield, high-risk bonds known as “junk bonds” in Wall Street jargon.^a

The best target for takeover is a company that has lots of assets in a nonliquid form that can be liquidated after takeover to pay off the junk bonds, but that cannot be sold quickly to defend against the takeover. Timber companies have valuable assets in the form of forests that can be liquidated after takeover, but cannot be sold quickly to buy back stock and prevent a hostile takeover. This made them popular takeover targets during the 1980s.

A classic example is Charles Hurwitz’s acquisition of Pacific Lumber in the mid-1980s. Pacific Lumber fought the takeover, but using a combination of junk bonds and short-term loans, Hurwitz won out, acquiring with the company’s 196,000 acres of forest, including the largest unprotected stands of virgin redwood in the world. Hurwitz was saddled with an enormous debt and crushing interest payments. To repay the debt, Hurwitz liquidated much of the forest stock, including many old-growth redwood groves. Some illegal cuts were conducted on weekends and holidays to avoid state regulators. Wall Street innovations during the 1980s accelerated the decimation of the nation’s last remaining virgin forests, and of the environmental services those forests once provided.^b

^a*Different companies (and cities and countries) have different credit ratings based on their financial soundness. Bonds from financially sound companies are themselves very sound, bonds from less sound companies are not, and the risk of default is higher.*

^b*N. Daly, “Ravaging the Redwood: Charles Hurwitz, Michael Milken and the Costs of Greed,” Multinational Monitor, 16(9) (1994). Online: http://multinationalmonitor.org/hyper/issues/1994/09/mm0994_07.html.*

Y*. An increase in liquidity preference raises the interest rate and lowers national income.

The above analysis of changes in propensity to save, efficiency of investment, liquidity preference, and money supply is summarized in Figure 16.5.

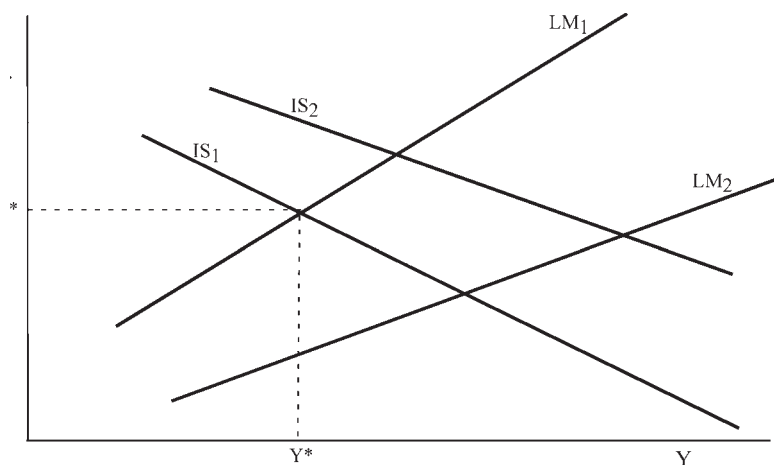


Figure 16.5 • Shifting of the IS and LM curves. The shift from IS_1 to IS_2 could be the result of either a decrease in the marginal propensity to save or an increase in the marginal efficiency of investment. The shift from LM_1 to LM_2 is caused by either a reduction in liquidity preference or an increase in the money supply. Can you work out the changes in r and Y resulting from an increase or decrease in each of the four parameters, others remaining constant? What about changes in two or more parameters at the same time?

■ IS-LM AND MONETARY AND FISCAL POLICY

Changes in the propensity to save, the efficiency of investment, and liquidity preferences are not brought about directly by policy interventions; they are affected by psychology and technology and as a result are difficult, if not impossible, to predict. However, policy makers do have two sets of economic levers by which they can influence these variables: monetary policy and fiscal policy.

What does the IS-LM model tell us about different monetary and fiscal policy levers? The analysis of monetary and fiscal policy in macroeconomics can be worked out by tracing the effects on IS or LM of changes to the money supply and of government taxing and spending.

Monetary policy basically affects the money supply. When the monetary authority (the Federal Reserve, in the U.S.) increases money supply, the LM curve shifts downward and $M > L$, which drives interest rates down, as explained earlier. Lower interest rates stimulate the economy, and income grows. If the money supply is increased by too much, there can be too much money chasing too few goods, and inflation threatens. Reducing the money supply drives interest rates up, shrinks the economy, and can help control inflation.

Fiscal policy is basically government expenditure and taxation. When the government spends money, industry has to produce more goods and

services to meet the increased demand. This drives up income and also increases the demand for investments, driving up interest rates. The IS curve shifts to the right. Decreasing government spending has the opposite effect.

There are three ways the government can finance expenditures. First, it can impose taxes. When the government increases taxes, people have less to spend, decreasing demand, and leading to less investment. The economy shrinks, and interest rates go down. However, if the government spends the entire tax increase, the stimulus of increased expenditure outweighs the contraction caused by taxes, since some of the tax money now being spent would have been saved. Second, the government can use debt financing and borrow money. The government borrows money by selling bonds. An increase in the supply of bonds drives down the price and drives up the interest rate. Third, the government can use its right to seigniorage to simply print and spend money, which increases the money supply. As we noted previously, the increased money supply will further stimulate the economy but will have a countervailing impact on interest rates.

Seigniorage-financed fiscal policy seems the logical choice for stimulating the economy, but it carries the threat of inflation. Governments could dramatically increase their ability to use seigniorage if they increased reserve ratios to 100%, as suggested by Frederick Soddy so long ago. The government would then be able to print and spend money when the price index started to fall, and tax and destroy money if inflation threatened to become a problem. The government would also be able to target monetary policy much more effectively, using it to address issues of scale, distribution, and allocation.

The impact of fiscal and monetary policy depends on how much excess capacity exists in the economy. Consider a bowling alley in a small, isolated town where the government is undertaking a large project to stimulate economic growth. When unemployment is high, wages may be fairly low, and few people have disposable income to spend on bowling. As a result, the bowling alley is virtually empty. If the government funds a large project in town, some people are directly employed by the project, and they spend much of their money in town, inducing other local businesses to hire to meet the increased demand. People use their extra income to go bowling, and the bowling alley's income grows.

Now imagine that the government implements the same project in a town with very low unemployment. Bowling is popular, and the alley is full every night. The government needs employees for the project, but increasing demand when supply is low drives up wages, the price of employees. Disposable income increases, but every new bowler at the alley simply "crowds out" another bowler, who would have to leave the alley. The alley might like to expand, but the government is borrowing money

to finance its project, driving up interest rates, making it too expensive for the alley owner to expand. The alley can raise its prices with the increased demand, but it must also pay higher prices for its labor force, and therefore can only break even. When an economy is at full employment, the bowling alley owners might be much better off with an expansionary monetary policy that lowered interest rates so they could expand. In contrast, if the government lowered interest rates when the alley had considerable excess capacity, expansion would do the owners no good at all.

The failure of lowering interest rates to stimulate economies with low demand is known as a **liquidity trap**. In general, the economy is somewhere between the extremes of depression and operation at full capacity (i.e., most bowling alleys are full sometimes, but very few are always full). While increased government expenditure leads to some degree of crowding out and increased interest rates, it also increases income.

Table 16.1 summarizes the impacts of fiscal and monetary policy on interest and income. In each case, the impact is obviously the opposite for the opposite policy.

Inflation and Disinflation

If we looked only at the IS-LM model, and if our goal was continued economic growth, the superior policy option would be clear: keep increasing the money supply to lower interest rates and stimulate investment, and use fiscal policy when necessary to stimulate demand. However, when we first presented the LM model, we saw that the real money supply is equal to the nominal money supply divided by prices—that is, the real money supply equals M/P . There is, therefore, another path toward equilibrium between supply and demand for money in response to an increase in money supply—price inflation. A larger nominal money supply divided by higher prices can lead to no change in real money supply. The closer the system is to full output, the less output is likely to increase in response to lower interest rates, and the more likely that monetary expansion will result in inflation. **Inflation** is an increasing general level of prices (not a state of high prices).

Why are governments and monetary authorities so worried about inflation? Are their concerns justified? How does inflation affect the real economy? The first point to make is that people appear not to like inflation, which alone is some justification for trying to avoid it. Many economists argue that inflation is regressive, but empirical support for this argument is difficult to find.⁵ Empirical evidence does show, however,

⁵A. Bulir and A.-M. Gulde, *Inflation and Income Distribution: Further Evidence on Empirical Links*, IMF Working Papers, no. 95/86. Washington, DC: International Monetary Fund, 1995.

■ **Table 16.1****EXPECTED IMPACTS OF BASIC MONETARY AND FISCAL POLICIES ON INTEREST RATE AND INCOME**

Policy	Interest Rate	Income
Monetary expansion can be accomplished by: <ul style="list-style-type: none"> • Reduced reserve requirements • Selling bonds on the open market • Lowering the discount (interest) rate 	(–) When economy is weak (high unemployment, low investment), monetary policy may have little to no impact on interest rates.	(+) When economy is weak, no impact on income. Known as the liquidity trap.
Tax increase	(–) Taxes (especially progressive income taxes) help stabilize the economy.	(–) Taxes collect more money when income grows and less when it shrinks.
Increased government expenditure	(+) Can be spent on market or nonmarket goods.	(+)
<ul style="list-style-type: none"> • Financed by deficit spending 	(+) Impact on interest rate may be small when economy is weak, large when economy is operating at full capacity.	(Probably +) Income will increase when economy is weak, but may not increase when it is already operating at full capacity; latter condition is known as crowding out.
<ul style="list-style-type: none"> • Financed by taxes 	(+) Increase in interest rate is less than occurs with deficit spending.	(+) Growth rate is less than occurs with deficit spending.
<ul style="list-style-type: none"> • Financed by seignorage 	0	+ Likely to cause inflation under crowding out conditions, with no real growth in income.

that real wages can fall substantially during prolonged episodes of high inflation.⁶ In addition, during episodes of high inflation, it is likely that the wealthy and educated are better able to take advantage of investments and contracts that protect their money than the poor. Thus, with continuous high inflation, the poor may well lose ground to the rich.

Hyperinflation, often defined as inflation greater than 50% per month, can also destabilize the economy. In hyperinflation, money fails not only as a store of value, but also as a medium of exchange. Impacts of moderate inflation depend to a large extent on whether it is expected or unexpected.

⁶B. Braumann, High Inflation and Real Wages, Western Hemisphere Department Series: Working Paper WP/01/50, May 1, 2001.

If everyone expects a certain rate of inflation, and their expectation comes to pass, then inflation is incorporated into contracts and causes very few problems. The only groups one would expect to lose from an expected inflation are holders of money (which pays no interest) and people on fixed incomes. However, with expected inflation, most people will hold less money, and incomes are likely to be inflation adjusted. **Disinflation** is a decrease in the rate of inflation. Deflation is a decline in the overall price level. Unexpected inflation, disinflation, and deflation have entirely different outcomes than expected inflation. The most useful way to assess the impacts of these unexpected changes is to look at debtors versus creditors.

Unexpected inflation has entirely different outcomes. The most useful way to assess the impacts of unexpected inflation and disinflation is to look at debtors versus creditors. When there is unexpected inflation, any loans with nominal interest rates (i.e., interest rates that are not pegged to inflation) will be worth less and less every year. Debtors benefit and creditors suffer. For example, people in the 1960s got 30-year house mortgages at around 6%. When inflation in the 1970s climbed over 12%, some home owners ended up paying back less than they originally borrowed. In general, unexpected inflation systematically redistributes wealth from creditors (generally the rich) to debtors (generally the poor). The government is a net debtor, and therefore benefits, as do the future generations that are expected to pay off the government's debts. However, a country cannot have unexpected inflation forever—eventually it becomes expected, or else becomes hyperinflation, with its accompanying problems.

What happens when the government tries to cause disinflation or deflation? Obviously, just as unexpected inflation benefits debtors, unexpected disinflation must benefit creditors. In 1980, a 30-year mortgage at 14% didn't look so bad when inflation was 13% annually, and people expected their incomes to rise by at least that rate. By 1986, however, inflation (and wage increases) had fallen to less than 2%, and creditors were collecting a 12% annual real return on their loans. Thus, existing debtors suffer and existing creditors benefit from disinflation.

Other impacts of disinflation depend on whether it is brought about by fiscal or monetary policy. Theoretically inflation can be reduced by decreasing aggregate demand or increasing aggregate supply, but policy usually acts on demand. Fiscal policy can only decrease aggregate demand through greater taxation or reduced expenditure, both of which should lower the real interest rate to the benefit of new debtors. Other distributional impacts depend on the specific policy used. For example, demand could be reduced by reducing subsidies for big business or by reducing transfer payments to the poor.

THINK ABOUT IT!

Under President Reagan there was a big emphasis on supply-side economics, increasing income by providing incentives for production (i.e., supply). Policy measures for achieving this include investment subsidies, reduced capital gains taxes, and reduced taxes for the rich. Can you explain why these policies would theoretically increase supply and reduce inflation?

The monetary authority, on the other hand, can only act to reduce demand by reducing the money supply, which increases real interest rates, to the detriment of debtors. Interest-sensitive sectors of the economy, such as farming and construction, also lose out. If losers are forced into liquidation or bankruptcy, they may be forced to sell their assets at bargain prices, and it is the well-to-do who maintain the liquidity necessary to purchase those assets. Thus, recessions may generate corporate mergers and increased concentration of the means of production.

The claim made for disinflationary policies is that in the short term the economy suffers, but in the long term stable money allows for steady growth and higher real wages. The problem is that short-term suffering can be severe, especially when monetary policy is used to decrease demand. While the jury is still out on the distributional impacts of moderate inflation, the distributional impacts of unemployment caused by disinflationary policies, as we will see below, are clear.

Unemployment

In the world of microeconomics, involuntary unemployment should not exist. Prices are set by supply and demand, and when the demand for labor is low, the price falls. At a lower wage, fewer people are willing to work, and supply falls accordingly, returning the system to equilibrium. Clearly, however, unemployment is a persistent problem in modern economies. We particularly want to examine two issues: the link between unemployment and inflation, and the implications of unemployment for distribution.

Some unemployment is inevitable. People are constantly entering and leaving the labor market, changing jobs, and moving from place to place. Businesses go bankrupt, or suffer downturns and lay people off. It always takes time to find a new job. This is known as “frictional” or “natural” unemployment. According to theory, if policy makers tried to reduce unemployment below this level, the result would be greater demand for a fixed number of workers. Workers would have more bargaining power and would demand higher wages, thereby causing inflation.⁷ Thus, a wide-

⁷This theory was originally introduced by Milton Friedman in his 1967 American Economics Association presidential address.

spread euphemism for “natural” unemployment is NAIRU, the non-accelerating inflation rate of unemployment.⁸ There is considerable disagreement over what NAIRU actually is. James K. Galbraith argues that economists are really quite practical—their estimates of NAIRU simply reflect actual unemployment.⁹

But the link between low unemployment and inflation is not clear empirically. Why not? We offer two explanations. First, in the era of globalization, large corporations are free to move their capital and production to other countries. Even when unemployment is low, corporations can counter demands for higher wages by a local workforce with the threat of moving to a lower-wage country. This explains how the low unemployment of the 1990s in the U.S. was accompanied by stagnant wages and a diminished share of national income going to wage earners.¹⁰ Second, we must point out that income from production is divided between wages, profit, and rent. Increased bargaining power by wage earners need not lead to “wage-push” inflation—it could instead simply increase the share of income going to wage earners and decrease the share going to rent or profit. Does increased bargaining power by owners lead to “profit-push” inflation?

In summary, then, low unemployment increases the bargaining power of wage earners, which translates into higher wages (though this effect is diminished by globalization). Higher wages can cause inflation, which then erodes the higher wages, or it can change distribution patterns between wages and profit. High unemployment, in contrast, increases the bargaining power of corporations and leads to redistribution toward the owners of capital. Whatever the validity of the theory behind NAIRU, it is quite clear that monetary authorities pay close attention to unemployment as an indicator of inflationary pressures. For example, when unemployment falls too low, the Fed tends to raise interest rates to reduce investment, employment, and demand. Distributional impacts of inflation are uncertain, but unemployment caused by disinflationary policies has clearly negative impacts on some of the poorest sectors in society.

Finally, it is worth noting that increasing unemployment can set up a vicious cycle. As people lose jobs, they lose money to purchase goods and services. With less demand, businesses respond by reducing supply, perhaps laying off more workers to do so, and further reducing demand.

⁸In the 1960s, economists found an inverse empirical relationship between unemployment and inflation, which was dubbed the Phillips curve. But during the 1970s, a number of economies experienced increasing unemployment and increasing inflation simultaneously.

⁹J. K. Galbraith, Well, Excuuuuse Me! *The International Economy*, December 1995.

¹⁰R. J. Gordon, The Time-Varying NAIRU and Its Implications for Economic Policy, NBER Working Paper No. W5735, May 1997.

Many fiscal policies such as welfare payments, unemployment insurance, and other transfer payments are designed to diminish this impact, adding stability to the economy. Economic stability is a public good, and an important policy objective.

The Impact of Policies on Scale, Distribution, and Allocation

Now that we understand the basic elements of fiscal and monetary policy, we can turn to their particular applications. How we apply these policies, of course, depends on what we wish to achieve. Mainstream macroeconomists primarily pursue continuous economic growth, with a lesser emphasis on distribution. Allocation is left to microeconomic forces. Ecological economists are primarily concerned with the impact of macroeconomic policies on scale (i.e., growth) but with a different goal than mainstream economists—to make sure that the costs of additional growth in material throughput are not greater than the benefits. Ecological economists assume that eventually the costs will exceed the benefits if they haven't already. They also place much more importance on distribution than mainstream economists. In short, ecological economics strives to create an economy in which there is no growth in physical throughput, while avoiding the suffering caused by recession or depression. The allocation of resources between market and nonmarket goods and services can play an extremely important role in this regard.

Macro-allocation

As we have discussed earlier, free markets work very well at allocating resources among market goods but very poorly at allocating nonmarket goods, typically failing to provide them in satisfactory quantities. Many policy makers already recognize this point, as can be clearly seen in government budgets, the bulk of which are spent on public goods such as defense, health care, education, road systems, bridges, streetlights, national parks, and so on.¹¹ In fact, few institutions besides government allocate resources toward nonmarket goods, and only the government is able to use policy to reduce demand and hence expenditure for market goods and shift it toward nonmarket goods.

For simplicity, we refer to the allocation between market and nonmarket goods as **macro-allocation**, and allocation among market goods as **micro-allocation**.

Macro-allocation is the allocation of resources among market and nonmarket goods and services.

¹¹National defense is generally considered a public good, though arms races, nuclear weapons, and excessive defense expenditures may do more to undermine national security than to ensure it. To the extent that disease is communicable and individuals are made uncomfortable by the suffering of others, health care is also a public good.

In the private sector, monetary policy directly affects only the *market* economy, by stimulating or discouraging investment in the production and consumption of *market goods* for profit. Why is this so? Monetary policy acts primarily through its impact on interest rates and hence on borrowing and lending. The private sector invests little in nonmarket goods, since such goods generate no profit that can be used to pay back loans. Therefore, lower interest rates will not affect the production of nonmarket goods by the private sector. Not only will monetary expansion do nothing to provide public goods and open access resources, it can actually increase the degradation of these resources if the production of the market goods is accompanied by negative externalities affecting the environment. Return to our example of the bowling alley. If lower interest rates induce it to expand, it will not expand into a void and may expand into some ecosystem—a wetland, for example—that currently provides valuable nonmarket services to the local community. As we discussed earlier, such negative externalities are an inevitable outcome of market production.

Therefore, if our policy objective is sustainable scale, monetary expansion is very problematic. Even if the economic scale is well within the constraints imposed by the ecosystem, monetary expansion acts on only one type of good, market goods. Market goods do not always offer the highest marginal contribution to human well-being, however. The microeconomic law of the equimarginal principle of maximization thus applies not only to the scale of the economic system relative to the ecosystem that sustains it, but also to the division of market and nonmarket goods produced by an economy. In ecological economics, macro-allocation is every bit as important as micro-allocation.

Theoretically, federal money in a democratic society will be directed toward the goods and services that provide the greatest marginal utility for society as a whole. As we have discussed, an important role of government expenditure is to provide nonmarket goods.

It is important to distinguish between two classes of nonmarket goods, which have different effects on scale. Manmade nonmarket goods affect scale to the same degree as market goods. If the government project in the bowling alley town is a big government building, it may also encroach upon some valuable ecosystem and destroy the services it provides. In contrast, what if the government project in the bowling alley town created a park that provides recreation and restores wetlands that serve as filters for the town's water supply and as a buffer against catastrophic floods? Protecting and restoring the ecosystems that provide nonmarket environmental services can effectively decrease scale, or at least help ensure that we do not surpass optimal scale. As the world becomes more full, the marginal benefits from protecting and restoring ecosystem funds, and hence the nonmarket services they provide, will increase relative to those

from market goods and manmade public goods. As this happens, and if politicians come to understand the benefits and public good nature of ecosystem services better, more and more federal money should be allocated toward providing such services.

It is important to recognize, however, that government expenditure on ecosystem funds can still increase scale. How is this so? We must remember that once the initial expenditure enters the economy, the multiplier effect takes hold. Money spent to restore ecosystem funds will in turn be spent by its recipients on market goods—workers restoring the wetland may spend their money on bowling, pressuring the bowling alley to expand. The larger the multiplier, the larger the impact on the market sector of the economy, and the less control the government has over composition. As most macroeconomics textbooks explain, tax increases decrease the multiplier and also reduce income. A smaller multiplier increases the ability of the government to affect macro-allocation, and reduced income reduces scale. Taxes can also be used to discourage undesired behaviors, such as pollution, and transfers can be used to encourage desired ones, such as environmental preservation. The full impact of taxation on scale and macro-allocation depends on how the taxes are spent, but taxes can certainly play an extremely important role in achieving an optimal scale—a point we will examine at greater length in Chapter 21.

Another important point must be made here. Under traditional analysis of the IS-LM curve, fiscal policy when the economy is operating at full capacity results in crowding out (remember the full bowling alley) and should be avoided. However, in terms of macro-allocation and scale, full output conditions can increase the effectiveness of fiscal policy. With full employment, if the government spends money to create a park and restore wetlands, interest rates and labor costs go up, and it is more difficult for the bowling alley to expand. (Fortunately, the park offers a recreational alternative to bowling that does not displace ecosystem services.) Government expenditure on restoring ecosystems under such conditions will therefore have an unambiguous impact on reducing scale.

What are the distributional impacts of fiscal and monetary policy? Fiscal policy in the form of taxation and government transfers can be easily and effectively distributed as desired. Government transfers such as welfare, unemployment insurance, Medicare, Medicaid, and Social Security all play an important role in distribution. Corporate welfare programs (which outweigh transfer payments to the poor¹²) affect distribution in the opposite direction. Public goods are available to all, and their provi-

¹²C. M. Sennott, "The \$150 Billion 'Welfare' Recipients: U.S. Corporations," *Boston Globe*, July 7, 1996.

sion improves distribution. In terms of income, progressive taxation can help reduce gross inequalities in income distribution, a necessary condition morally and practically if we are to achieve a sustainable scale. Monetary policy can also play an important but narrower role in distribution. High interest rates caused by tight monetary policy can lead to unemployment, and they favor creditors over debtors, as discussed earlier in the section on inflation and disinflation. Low interest rates have the opposite effect.

In summary, in terms of ecological economic goals, monetary policy is a blunt instrument directed only toward the production and consumption of market goods, with limited flexibility in terms of distribution and macro-allocation. Expansive monetary policy increases scale. Fiscal policy has far greater flexibility in terms of scale, distribution, and macro-allocation.

■ IS-LM IN THE REAL WORLD

While the IS-LM model is very useful, it has important limitations.¹³ The model is deceptively simple and does an inadequate job of conveying the real-life complexity of monetary and fiscal policies. While the model shows the general impacts of such policies, it fails to incorporate the issues of uncertainty, time lags, and structural changes, as well as the difficulty of choosing the appropriate policy variables to manipulate.

Economists typically have a poor understanding of what is happening in the economy at any given moment. Is unemployment too high? Is the economy growing too fast, threatening inflation? Are we headed for recession? Weathermen may be inaccurate at predicting the future, but at least they can look outside right now and tell you what the weather is. Economics is less advanced. Viewing the same data concerning the economy, economists frequently disagree on how to interpret them. For example, in the United States, the Federal Reserve Bank raised interest rates in May 2000, fearing that the economy was growing too fast, and continued to voice fears of inflation through November 2000. Less than 2 months later, the Fed initiated the first of several interest rate cuts designed to ward off recession. Part of the problem is that the economic system is evolving rapidly in response to technological, environmental, cultural, and structural changes.

¹³See J. R. Hicks, IS-LM: An Explanation, *Journal of Post-Keynesian Economics* III (2): 139–154 (Winter 1980–1981). Hicks, the originator of the model, expresses reservations about how well it fits the real world once expectations and dynamics are recognized. He considers the model useful for understanding the past, but less so for understanding the future. We agree, but feel that many of Hicks' caveats apply to all equilibrium models, and that, rare though it is among scholars, Hicks was too hard on himself!

Compounding the difficulty of an inadequate understanding of the economy are the time lags involved in policy. There are two types of lags: lags in decision making (the inside lag) and lags between the time the decision is made and the policy takes effect (the outside lag). In fiscal policy, decisions such as tax cuts and expenditure increases are typically debated at length. Both legislative and executive branches must agree, and appropriate legislation must be passed. The decision lag can therefore be substantial. Once the decision to increase or decrease expenditure has been made and carried out, the outside lag may be relatively short, as such policies have an immediate effect on aggregate demand (though the full effect of the multiplier will take some time). Tax cuts or increases, on the other hand, have much slower results and are often not even felt until the next tax year.

The Fed, in contrast, generally has a much shorter decision-making lag. The Federal Open Market Committee (FOMC), responsible for Fed Policy, meets about eight times a year.¹⁴ Policy is generally decided at the meeting, and open market transactions can take place almost immediately. However, the most relevant impact of these policies is on interest rates and their effect on investment and consumption decisions. Investment decisions are rarely spur of the moment; they generally have a long gestation period. Thus, the Fed has a short decision-making lag, and a long lag before the policy takes effect.

These lags are very important to consider when deciding on a policy. It is quite possible that by the time a decision is made and the resulting policy takes effect, the problem the policy was designed to address will have disappeared, and a policy with the opposite effect may even be required.

Another problem is disagreement over what type of policy should be pursued, and what the impact will be, especially for monetary policy. The Fed usually tries to manipulate one of two targets: the money supply or the interest rate. Not only is there considerable debate over which course the Fed should pursue, there are serious obstacles to achieving either goal. For example, as Fed chairman Alan Greenspan admitted in congressional testimony: “. . . we have a problem trying to define exactly what money is . . . the current definition of money is not sufficient to give us a good means for controlling the Money Supply. . . .”¹⁵

Psychology can also make it difficult to manipulate interest rates. As we discussed earlier, interest rates are ultimately determined by the bond markets. Bonds, of course, mature in the future, and the amount someone

¹⁴The Federal Reserve Act mandates that the FOMC meet at least four times a year, and since 1981 it has met eight times a year.

¹⁵Congressional testimony, February 17, 2000.

is willing to pay for a bond depends on their *expectation* of future inflation. The Fed might implement an expansionary monetary policy to bring down interest rates, but if bond marketers believe this expansion will instead induce inflation or force monetary contraction in the near future, it could paradoxically serve to drive interest rates *up*.

A final problem with policy in countries with independent monetary authorities is the difficulty in coordinating between monetary and fiscal policy. This problem can become acute when the monetary authority and the government have different policy objectives. The elected government is mainly concerned with growth and employment, two issues that affect voters, and hence their elected representatives. In contrast, the Fed is mainly concerned with “sound” money (i.e., no inflation) and has frequently pursued this policy in the past, even when it has caused significant hardship in the form of unemployment and worse.

Box 16-5 WHY IS THE FED SO ANTI-INFLATION?

From our discussion of inflation, it would seem that inflation is less harmful than unemployment induced by anti-inflationary policies. Why then is the Fed so anti-inflation? In answering this question, it is worth bearing in mind who the natural constituency of the Fed is. Most members of the FOMC are bankers or Wall Street professionals, and the Fed seems to listen closely to the concerns of these groups. These two groups form the bulk of the wealthy creditors who benefit from low inflation and disinflationary episodes, and who are unable to increase their share of national income as readily during inflationary periods.^a

^aW. Greider, *Secrets of the Temple: How the Federal Reserve Runs the Country*, New York: Simon & Schuster, 1987.

Despite its shortcomings, the IS-LM model is a vast improvement over prior models. It is a two-sector general equilibrium model, the sectors being the real sector and the monetary sector. Before Hicks' model, economists often tried to explain the interest rate as a purely monetary phenomenon (liquidity preference and money supply) or a purely real phenomenon (savings and investment). There was an investment rate of interest and a money rate of interest, and confusion about which set of factors “really” determined the interest rate. Hicks showed that the real and monetary sectors simultaneously interact to determine both the interest rate and national income. But Hicks said nothing about the ecosystem and biological rates of growth. In 1937 the world was still considered

“empty.” Thus, the IS-LM model treats all economic growth as identical—it does not distinguish between government expenditures on market goods, manmade public goods, or investments in ecological restoration, nor does it address distribution.

■ ADAPTING IS-LM TO ECOLOGICAL ECONOMICS

How might the IS-LM model be adapted to ecological economics? Remembering our basic vision of the macroeconomy as a subsystem of the finite and nongrowing ecosystem, the most obvious suggestion would be to impose an external constraint on the model representing the biophysical limits of the ecosystem. For example, we could assume a fixed throughput intensity per dollar of Y (i.e., GNP), so that a given Y in money terms implied a given physical throughput. Then we could estimate the maximum ecologically sustainable throughput, convert that into the equivalent Y , and impose that as an exogenous constraint on the model. Based on Figure 16.4, it would be represented by a vertical line at the Y corresponding to maximum sustainable throughput. It would not be a function of the interest rate at all.¹⁶ Let's call the vertical line EC for “ecological capacity.” It reflects a *biophysical* equilibrium, not an *economic* equilibrium. It is ignored by the actors whose behavior is captured in the IS and LM curves.¹⁷

The most obvious approach is not always the best, but it is usually a good place to start. Also, this approach closely parallels the macroeconomist's representation of full employment of labor as a perpendicular at the level of Y corresponding to full employment at an assumed labor intensity of GNP. Our EC line represents “full employment” of the environment at an assumed throughput intensity of GNP. Later we will discuss further the assumption of fixed throughput intensity.

Let's consider the three possible positions of the biophysical equilibrium relative to the economic equilibrium, shown in Figure 16.6. The first case represents the “empty world” scenario. The biophysical limit is not binding. The distance Y^*C may be thought of as excess carrying capacity.

¹⁶A. Heyes, A Proposal for the Greening of Textbook Macro: “IS-LM-EE,” *Ecological Economics* 32 (1) (2000); P. Lawn, *Toward Sustainable Development: An Ecological Economics Approach*, Boca Raton, FL: Lewis Publishers, 2001. Heyes and Lawn have proposed an EE curve corresponding to the ecological limits discussed here that would be a function of the interest rate. Several technologies produce income, some of which require or degrade more natural capital than others. Less natural capital-intensive technologies require investments and are thus more likely at lower rates of interests. One problem is that the investments themselves would require natural capital.

¹⁷However, it need not be ignored by government, which affects the IS curve. The government is perfectly capable of investing in environmental services produced by natural capital and other nonmarket goods. But it is completely ignored by monetary policy (the LM curve), which acts on the economy through its effect on interest rates and hence market goods.

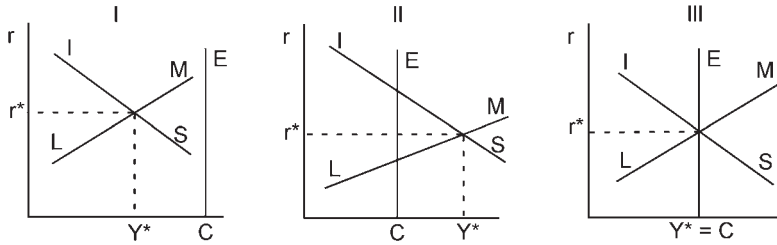


Figure 16.6 • The biophysical equilibrium relative to the economic equilibrium.

Most macroeconomists who use the IS-LM model would have this case in mind, if indeed they thought at all about EC. If the distance Y^*C is large then there is for practical purposes of short run policy no point in conceiving or drawing the EC line.

The second case is the “full world” (or overfull) scenario. The economic equilibrium has overshoot the biophysical equilibrium. The distance CY^* , the overshoot, can be given two interpretations. In neither case, however, would it represent real income. In the first case we might think of it as a purely monetary phenomenon, inflation. After hitting EC the real sector is effectively at the end of its tether. Even though the monetary sector keeps on churning, it only generates price increases for the same real income C . The other interpretation assumes that the real sector keeps producing real output, but by unsustainable drawdown of natural capital. Thus, CY^* would represent capital consumption counted as income. As natural capital is consumed, the EC line eventually has to shift even farther to the left, increasing the overshoot. Most ecological economists believe the second case to be a rather accurate description of the present state of affairs. We seem to be avoiding short-term inflation by long-term capital drawdown. Most conventional economists do not worry about long-term capital drawdown and shifting the EC curve farther to the left because they believe that knowledge is shifting EC to the right and thereby restoring the empty world situation.

The third case represents a big coincidence under our assumptions. For the economic equilibrium to coincide with the biophysical equilibrium would require either extraordinary good luck, or purposeful coordination and planning. There is nothing in the model to make it happen, just as, currently, there seems to be nothing in our institutions or behavior that would make it happen. In Chapters 20–23, we will discuss policy changes that could theoretically lead to this outcome.

Recall that we previously discussed the concept of full employment, which we might represent by an FE limit for labor similar to the EC limit for natural capital. Ideally, a FE labor line should coincide with the ISLM

equilibrium point—make $IS = LM$ at full employment. If FE is beyond the intersection of IS and LM , then policy makers might pursue FE through growth in Y . But what if FE is beyond EC ? The problem is no longer to pursue FE by growth in Y , but instead through structural change, such as shifting factor intensity away from fossil fuels and manmade capital (both of which rapidly draw down natural capital) and toward labor. We have already explained that when $IS = LM$ beyond EC , we are likely to draw down natural capital, and it is implicit in the acronym $NAIRU$ (the non-accelerating inflation rate of unemployment) that going beyond FE results in inflation. Why the difference? Why doesn't moving beyond EC also simply cause inflation? The answer is that natural resources are either free or cheap to begin with; they are not appropriately priced by the market mechanism, and excessive use therefore does not affect the price signal.

It remains true, however, that the assumption of constant throughput intensity of Y is troublesome. We know that throughput intensity of Y changes with new technology, and with shifts in the mix of goods that make up Y , even if probably not with factor substitution of capital funds for throughput flows. Differing assumptions about throughput intensity of Y can at least be represented by a shift in the EC perpendicular. (Curve shifting is, as you have no doubt realized by now, not an uncommon device in economic analysis!) However, in terms of practical policy recommendations, perhaps the best approach would be simply to impose the ecological constraint as a limit on throughput. For any given technology, a fixed limit on throughput will also limit Y , but over time, new technologies and a different mix of goods and services can allow Y to increase without increasing throughput, thereby threatening the life-support functions of the ecosystem.

BIG IDEAS to remember

-
- | | |
|---|---|
| ■ Macroeconomic model vs. general equilibrium | ■ IS-LM analysis of monetary and fiscal policy |
| ■ Real vs. monetary sectors | ■ Comparative statics |
| ■ $IS = LM$ | ■ Crowding out |
| ■ $MV = PQ$ | ■ Inflation and disinflation |
| ■ Transactions demand for money | ■ Unemployment |
| ■ Liquidity preference | ■ Macro-allocation |
| ■ Relation of bond prices to interest rate | ■ IS-LM adapted to show economic and biophysical equilibria |
-



Conclusions to Part IV

Chapters 13–16 offered a whirlwind tour of the macroeconomy. The real economy consists of the physical transformation of low-entropy matter-energy from nature into forms that enhance human welfare. GNP is an inadequate measure of the real economy, because it lumps together both goods and bads. The monetary sector of the economy functions as a lubricant that helps us allocate what is produced. The law of diminishing marginal utility tells us that distribution, both within and between generations, can be an important tool for increasing human welfare. Macroeconomic policy levers include government expenditure, taxes, money supply, and interest rates. These policy levers must be used to attain the goals of sustainable scale and just distribution—goals essentially ignored by the market microeconomy. We next turn our attention to international economics, to see how it affects the policy levers we just discussed.

PART V

International Trade





International Trade

It is a lot easier to grow grapes and make wine in Italy than in Norway. It is easier to hunt reindeer in Norway than Italy. To the extent that Norwegians want Chianti and Italians want reindeer steaks, there is an obvious basis for international trade. Economists never needed to prove that. Nor, given climate and geography, is it hard to understand why Italians traditionally drink Chianti and eat fish, while Norwegians traditionally drink aquavit and eat reindeer and whales. But we all like to experience other peoples' traditions, tastes, and capacities. International trade allows that, to the mutual benefit of all. In this view, national production for national consumption is the cake; international trade is the icing. Of course, no one objects to trade in this sense. But this is not an accurate picture of either the reality of trade today, or the trends and goals of globalization. However, before we tackle the difficulties of globalization, let's look more closely at the classical argument for free trade, an argument that goes well beyond the common-sense picture just sketched.

■ THE CLASSICAL THEORY: COMPARATIVE ADVANTAGE

The case for free trade, like the case for exchange in general, hinges on the assumption that trade is voluntary. Both parties to a voluntary exchange must be better off, in their own estimation, after the trade than before—otherwise they would not have made the trade. Under what conditions is this happy result likely to be the case? The most obvious condition is that of **absolute advantage** in cost differences reflected in the example above. If country A can produce something at a lower absolute cost than country B, while B can likewise produce something else at a lower absolute cost than A, then there is a reason for voluntary exchange, assuming A wants some of B's product and B wants some of A's product in the first place.

A country has an *absolute advantage* if it can produce the good in question at a lower absolute cost than its trading partner. It has a *comparative advantage* if it can produce the good in question more cheaply relative to other goods it produces than can its trading partners, regardless of absolute costs.

Initially economists thought that if one country produced all goods more cheaply than another, there was no basis for mutually advantageous trade between them—the more efficient country would only harm itself by trading with the less efficient country. David Ricardo demonstrated that this was not so—that both countries could benefit from trade even if one country had an absolute advantage in all tradable goods.¹ The key to understanding this is to focus on “comparative advantage” rather than absolute advantage.

Ricardo demonstrated this using a two-country, two-goods numerical example. We will retrace his logic using just such an example shortly. But first it is worthwhile to whet our appetite by noting how proud economists are of this result. Trade theorist R. Findlay referred to comparative advantage as the “deepest and most beautiful result in all of economics.”² Surveys have shown that about 95% of economists support the policy of “free trade.” MIT economist Paul Krugman said that “If there were an Economist’s Creed it would surely contain the affirmations ‘I believe in the Principle of Comparative Advantage,’ and ‘I believe in free trade.’”³ With that fanfare, let’s bring the comparative advantage argument onstage.

Consider a world in which we have two countries, A and B, each producing two goods, coal (C) and wheat (W). Since we want to demonstrate the gains from trade without appealing to absolute advantage, confining our reasoning only to comparative advantage, let’s impose on ourselves a veil of ignorance about absolute cost differences. Specifically, let’s say that country A measures costs in terms of a-units, and country B measures costs in terms of b-units, and that we know nothing about the relation of a to b. Perhaps $a > b$, or $a < b$, or $a = b$. Therefore, we cannot compare costs across countries, as required to know absolute advantage. But we can compare the cost of C and W within each country. The comparative or relative cost of W to C in country A can be calculated and compared to the comparative or relative cost of W to C in country B. That information will be sufficient to demonstrate the possibility of mutual gains from trade between A and B. The argument will depend only on comparative advantage (internal cost ratios), and not on cross country comparisons of cost (absolute advantage) because the latter have been made impossible by assuming incomparable units for measuring costs in the two countries. Table 17-1 summarizes the unit costs of C and W in A and B.

¹D. Ricardo, *On the Principles of Political Economy and Taxation*. 3rd edition. London: John Murray, Albemarle-Street, 1821 (originally published in 1817). Online: <http://www.systemics.com/docs/ricardo/principles.html>.

²R. Findlay, “Comparative Advantage.” In J. Eatwell et al., eds. *The New Palgrave: The World of Economics*, New York: Norton, 1991, p. 99.

³P. Krugman, Is Free Trade Passé? *Economic Perspectives* 1(2): 131 (1987).

■ Table 17.1

UNIT COSTS OF COMMODITIES			
	Country A	Country B	Total Output
Coal	1a	1b	2C
Wheat	1a	4b	2W
Total Resources	2a	5b	

Each country needs both coal and wheat, so before trade, each country allocates its resources between the two commodities and produces as much of each as permitted by its unit costs, as shown in the table. A has total resources of 2a. With unit costs of both coal and wheat equal to 1a, it can produce one unit of coal and one unit of wheat with its endowment of 2a resources. Country B has a total resource endowment of 5b. B allocates 4b resources to produce one unit of wheat, and with the 1b remaining can produce one unit of coal. Total world output before trade is $2W + 2C$.

Ricardo has a look at this situation and realizes that we can do better by allowing specialization and trade according to comparative advantage. He notices that W is four times as expensive as C in country B, and only one time as expensive as C in country A. In country B coal is cheaper relative to wheat than it is in country A. Therefore B has a comparative advantage in coal. Similarly, wheat is cheaper relative to coal in country A than it is in B. Consequently A has a comparative advantage in wheat.

Now let each country specialize in the production of the commodity in which it has a comparative advantage and trade for the other commodity. A will specialize in wheat, allocating all of its 2a resources to wheat, and with a unit cost of wheat of 1a will end up producing 2W. B will allocate all 5b units of its resources to production of coal, and with a unit cost of coal of 1b will produce 5C. Total world product after specialization and trade is $2W + 5C$. The world has gained an extra 3C with no extra resources. The world is better off. And we have shown this without ever having to compare costs of wheat in country A with costs of wheat in country B, or costs of coal in country A with costs of coal in country B. Only comparative advantage (a difference in internal cost ratios) matters, not absolute advantage. Suppose we suddenly learned that in absolute terms $a = 5b$, so that country B has an absolute advantage in both C and W. Does that change our conclusion? Not a bit.

Although the world is better off by 3C, one might still wonder which country gets the 3C, or how it is divided between them. We can see that the world as a whole is better off, but how do we know that each country

is better off? Might not one country suffer while the other gains a lot? Ricardo had an answer. How the total gains from specialization and trade, $3C$, will be divided between A and B depends on the terms of trade, the relative price at which W exchanges for C between the two countries. That will depend on supply and demand and bargaining power, so we cannot say exactly how the gains will be divided. But we can be sure, says Ricardo, that neither country will be worse off after trade. How do we know this? Because, since trade is voluntary, neither country will accept terms of trade less favorable than the terms of its own internal cost ratio, the terms on which it can “trade with itself.” A will not trade $1W$ for less than $1C$, because it could do better by reallocating its own resources back from W to C again. Likewise B will not give more than $4C$ for $1W$. So the terms of trade will fall somewhere between the limits of $1C/1W$ and $4C/1W$. Anywhere between those ratios both countries gain a part of the extra $3C$.

■ KINKS IN THE THEORY

Ricardo’s demonstration is indeed interesting and impressive. To more than double coal production, with no sacrifice of wheat production and no additional resources, is a neat trick. Within the world of its assumptions, the logic of the comparative advantage argument is unassailable. But it is time to have a closer look at the assumptions. What are the assumptions, and might they veil some costs that need to be subtracted from the gain of $3C$?

First, it is not really true that the extra $3C$ is produced with no additional resources. There is the obvious increase in the rate of depletion of coal mines and of pollution resulting from burning the coal. What the trade economists mean by “no extra resources” is simply no additional labor or capital. Recall the neoclassical production function discussed earlier, in which output is a function only of labor and capital inputs. But, contrary to the neoclassical assumptions, the resource cost of extra output cannot simply be ignored. Therefore some deduction from the value of $3C$ is required to account for extra depletion and pollution. But let us assume that is done and there are still net gains from trade and specialization.

Second, we have neglected or abstracted from transport costs, implicitly assuming them to be zero. Wheat has to be shipped from A to B, and coal from B to A. If the energy costs of that transportation were $3C$ or greater then the world would gain nothing. It is worth noting that transportation is energy intensive, and currently energy is not only directly subsidized, but, in addition, many of its external costs are not internalized in its price. Consequently, international trade is indirectly subsidized by energy prices that are below the true cost of energy. But let us suppose that there is still a net gain from trade after deducting fully counted transport costs.

Third, we also assumed that in each country the cost of specialization was negligible. But there are two important costs to recognize. First, in A all coal miners must become wheat farmers, and in B all wheat farmers must become coal miners. Making such a shift is costly to all whose livelihood is changed. Also in the future the range of choice of occupation has been reduced from two to one—surely a welfare loss. Most people derive at least as much life satisfaction from how they earn their income as from how they spend it. Economists practically identify welfare with increased range of choice among commodities, but are strangely silent about the welfare effects of the range of choice among jobs. Second, after specialization countries lose their freedom not to trade. They have become vitally dependent on each other. One's access to essential items depends on the cooperation of people on the other side of the world, who, however, admirable they may be, have different customs, different values, different interests, and different types of government. Remember that the fundamental condition for trade to be mutually beneficial is that it be voluntary. The voluntariness of “free trade” is compromised by the interdependence resulting from specialization. Interdependent countries are no longer free not to trade and it is precisely the freedom not to trade that was the original guarantee of mutual benefits of trade in the first place.

True enough, as Ricardo pointed out, if the terms of trade become too disadvantageous the country getting the worse end of the deal can opt out and despecialize—put some of its wheat farmers back in the coal mines, reinvest in mining equipment, reactivate mining legislation, re-employ exporters and importers, and so on. But this is both costly and socially disruptive in reality. The model assumes specialization is reversible, while in fact it is closer to being irreversible—that is, reversible but at a high cost. Countries specializing in non-essential products—bananas, sugar, cocoa, and so on—are especially vulnerable to hardship from having lost their freedom not to trade. It is easier to drive a hard bargain if you are selling essentials and buying nonessentials than vice versa. Clearly there should be some further deduction from the 3C resulting from the above costs of specialization.

THINK ABOUT IT!

Many countries specialize in only one or few commodities for export. For example, Ecuador specializes in bananas, Ghana in cocoa, and Cuba in sugar. What happens to the price of these commodities when lots of other countries start producing them? What happens to the price when growing conditions produce a global bumper crop? What happens to export income for each country when some regions have great growing conditions, and other regions lousy ones? Look up price trends for these commodities. Do you think specialization in agricultural commodities is a good development strategy?

■ CAPITAL MOBILITY AND COMPARATIVE ADVANTAGE

But suppose we still have a net gain. Can we then conclude that the argument for free trade on the basis of comparative advantage holds true? In fact, there is an often-overlooked provision of the theory, one of great relevance today in the debate over globalization, that suggests that it does not hold true. Note that our numerical demonstration of comparative advantage implicitly assumes factor (labor and capital) immobility between country A and B. Only coal and wheat crossed borders. Labor and capital stayed at home and were reallocated domestically between coal and wheat according to the principle of comparative advantage. Since it is usually the capitalist who makes the investment allocation decisions let us just focus on capital and its mobility or lack thereof. Clearly our model implicitly assumes that capital is mobile between sectors within each country, but immobile between countries. Capitalists cannot even compare costs or profitability between A and B because, thanks to our veil of ignorance, they cannot compare a-units with b-units. Therefore they could not possibly know whether or not investment in the other country would be profitable or not.

One other way to show the implicit assumption of immobile capital in modern texts is to note that the examples, like ours, are usually in terms of barter—wheat for coal, with no money involved. Barter trade is always necessarily balanced. No monetary or short-term capital flows are required to balance the differences in exports and imports of bartered goods. The **current account** is the difference between the monetary value of imported and exported goods and services. If imports are greater than exports, the current account is in deficit. If exports are greater than imports, the current account is in surplus. **Capital accounts** are the difference between monetary flows, used to purchase various assets, into and out of a country. Such assets include stocks, bonds, real estate, and other assets that remain in the country. When more money flows into the country than out, the capital accounts are in surplus. The current account of the balance of payments is always balanced in barter, so there is no need for a compensating imbalance on capital account. Therefore, barter examples assume balanced trade, and balanced trade means no capital mobility.

Immobile capital does not mean that producer goods cannot be exchanged on current account. Machines and tractors, “materialist capital” can be traded just like shoes and sugar. What is immobile is capital in the “fundist” sense, money, or liens on the future product of the deficit country. Immobile capital in the fundist sense is the same thing as balanced trade on the current account, or trade that requires no compensating transfer of fundist capital on capital account. In other words, immobility

of capital does not prevent Brazil from importing machines and tractors and paying for them with exports of shoes and sugar. It just prohibits Brazil from importing machines and tractors *faster* than it can pay for them by exporting shoes and sugar, and thereby paying for the extra machines and tractors by issuing liens against its future production of shoes and sugar.

The capital immobility assumption is often hidden by examples in terms of individuals who specialize and trade, rather than nations. A favorite example is the lawyer and her secretary. The lawyer happens to be a champion speed typist. She has an absolute advantage over the secretary both in typing and of course in practicing law, since the secretary is not a lawyer. But the lawyer nevertheless finds it advantageous to hire the secretary to do the typing, and both benefit from this exchange. Although the lawyer is a better typist than the secretary, she is not much better. But she is a much better lawyer. So the secretary has a comparative advantage in typing (although an absolute disadvantage), and the lawyer has a comparative (and absolute) advantage in law. So they specialize according to comparative advantage. The lawyer is not so silly as to spend her time typing when she could be billing clients at \$300 per hour. Where is the assumption of capital immobility in this example? It is hidden by the obvious fact that it is impossible for the productive energy and capacity of the secretary to be transferred to the absolutely more efficient person of the lawyer. The lawyer is not a vampire that can suck the lifeblood and energy out of the hapless secretary in order to employ it more efficiently. In other words, productive capacity, “capital” is immobile between the lawyer and her secretary, so the logic of comparative advantage works.

Between countries, however, it is not impossible for productive capacity, capital, to be transferred from one country to another in response to absolute advantage. Capital mobility has to be explicitly ruled out for comparative advantage argument to work between countries. Although each country as a whole benefits under comparative advantage trade, not every citizen or group of citizens will benefit. Everyone could benefit, but only if winners were to compensate losers—that is, only if coal miners in B compensate wheat farmers, and wheat farmers in A compensate coal miners for the costs of changing their livelihood. As we move to absolute advantage (capital mobility internationally), some entire nations may lose, but the world as a whole will gain and could compensate the losing country, just as within countries the government could compensate the losing industry. Such compensation usually does not take place within nations, and almost never takes place between nations. Within nations there are at least institutions of community that could carry out internal transfers. At a global level there are no such

institutions. The move from comparative advantage to absolute advantage seems to be part of the general retreat from the Pareto to the Hicks-Kaldor (often referred to as “potential Pareto”) welfare criterion. The latter represents a retreat from the condition that no person be made worse off to the weaker condition that winners could, if they so choose, compensate losers and still be better off. Similarly, the former is a retreat from the condition that no nation be made worse off to the weaker condition that winning nations could compensate losing nations, if they choose to, and still be better off. In both cases the weaker criterion takes compensation as only potential, not actual, and in the case of trade the compensation condition is seldom even mentioned.

■ ABSOLUTE ADVANTAGE

Ricardo, to his credit, was very explicit about his assumption of immobile capital between countries. If capital were mobile internationally, it is obvious that capitalists would seek the greatest absolute profit and consequently the lowest absolute cost of production. In our example, if capital were mobile and the capitalists knew that $1a = 5b$, so that B had an absolute advantage in both coal and wheat, they would invest in B and forget about A. Comparative advantage would be irrelevant. If capital is mobile, absolute advantage is the relevant criterion. The only reason the capitalist would ever be interested in comparative advantage is if capital were immobile internationally. If capital cannot follow absolute advantage abroad, the next best thing is to follow comparative advantage specialization at home and trade for the foreign product. Comparative advantage is a clever second-best adaptation to the constraint of international capital immobility. But without that constraint, it has no reason to be, and absolute advantage is all that counts.

Why is the capital immobility assumption so important? Because it is utterly counterfactual in today's world. Capital is mobile all over the world in trillion dollar amounts at the speed of light.

Why is this overlooked? The nice thing about the comparative advantage argument is that both countries benefit from free trade, and gains are mutual—at least in theory, if not always in fact. The problem with absolute advantage is that both countries do not necessarily gain. If one country has an absolute disadvantage in both commodities, it will lose jobs and income as capital moves abroad. But under absolute advantage, world production will still increase. In fact, it would in theory increase by more than under comparative advantage. This is because in moving from comparative to absolute advantage, we relax a prior constraint on world product maximization—namely, the condition of capital immobility. Mo-

bile capital can seek out more productive opportunities than it could when it was confined to its country of origin. But while there is certainly a case for absolute advantage, it lacks the politically very convenient feature of guaranteed mutual benefit that was part of the traditional comparative advantage argument for free trade. All countries still could be better off if there were some regulatory institution whereby winners could compensate losers. But that would no longer be “free” trade. The focus of policy has shifted away from the welfare of nations to the welfare of the globe as a whole. And that brings us to the issue of globalization, or more specifically of globalization versus internationalization as alternative models of world community.

■ GLOBALIZATION VS. INTERNATIONALIZATION

Internationalization refers to the increasing importance of relations between nations: international trade, international treaties, alliances, protocols, and so on. The basic unit of community and policy remains the nation, even as relations among nations, and among individuals in different nations, become increasingly necessary and important.

Globalization refers to global economic integration of many formerly national economies into one global economy, by free trade, especially by free capital mobility, and also, as a distant but increasingly important third, by easy or uncontrolled migration. Globalization is the effective erasure of national boundaries for economic purposes. National boundaries become totally porous with respect to goods and capital, and increasingly porous with respect to people, viewed in this context as cheap labor, or in some cases cheap human capital.

Ricardo and the classical economists who argued for free trade based on comparative advantage were basically nationalists and retained their fundamental commitment to the nation even as they advocated internationalization.

In sum, globalization is the economic integration of the globe. But exactly what is “integration”? The word derives from *integer*, meaning one, complete, or whole. Integration means far more than “interdependence”—it is the act of combining separate albeit related units into a single whole. Integration is to interdependence as marriage is to friendship. Since there can be only one whole, only one unity with reference to which parts are integrated, it follows that global economic integration logically implies national economic disintegration—parts are torn out of their national context (disintegrated), in order to be reintegrated into the new whole, the globalized economy. As the saying goes, to make an omelet, you have to

break some eggs. The disintegration of the national egg is necessary to integrate the global omelet.

■ THE BRETTON WOODS INSTITUTIONS

At the end of World War II there was a conference of nations held in Bretton Woods, New Hampshire, for the purpose of reestablishing international trade and commerce, which had been disrupted by the war. The international diplomats and economists, led by John Maynard Keynes of England, and U.S. Secretary of the Treasury Henry Morgenthau and his aide Harry Dexter White, were successful in negotiating the charter that set up the **Bretton Woods Institutions**, the **International Monetary Fund (IMF)** and the **International Bank for Reconstruction and Development (World Bank)**. These two institutions are made up of member nations. They were founded on the federal model of internationalization, as just discussed, not the integral model of globalization. Their founding, almost 60 years ago, was a wonderful achievement of international cooperation and diplomacy. It symbolized the end of an era of economic depression followed by war and destruction, and the beginning of a hopeful era of peace and production in which “swords would be beaten into plowshares.”

The atmosphere of eager optimism was expressed by Morgenthau,⁴ who envisaged: “a dynamic world economy in which the peoples of every nation will be able to realize their potentialities in peace . . . and enjoy, increasingly, the fruits of material progress on an earth infinitely blessed with natural riches.” Morgenthau went on to say that “prosperity has no fixed limits. It is not a finite substance to be diminished by division.” The same note was sounded by Keynes: “In general, it will be the duty of the Bank, by wise and prudent lending, to promote a policy of expansion of the world’s economy. . . . By expansion we should mean the increase of resources and production in real terms, in physical quantity, accompanied by a corresponding increase in purchasing power.” The “empty world” vision of the economy was dominant. Notions of ecological limits to growth were not on the horizon, much less on the agenda. The founders of the Bretton Woods Institutions felt that they had far more pressing issues to deal with in 1945, and they were surely right. But the world has changed a lot in 60 years. Population has roughly tripled, and resource throughput has increased more than ninefold,⁵ moving us a long way from the “empty” toward the “full” world.

⁴Morgenthau and Keynes quotes are from B. Rich, *Mortgaging the Earth*, Boston: Beacon Press, 1994, pp. 54–55. For an interesting history of the Bretton Woods Conference see also the biography of Keynes by R. Skidelsky, *John Maynard Keynes*, Vol. III, “Fighting for Freedom, 1937–1946,” New York: Viking Press, 2000.

⁵Calculated from data in J. B. DeLong, *Macroeconomics*, Burr Ridge, IL: McGraw-Hill, 2002, chap. 5. The data are labeled GNP, but in the text DeLong refers to it as “material output,” what

The division of labor between these institutions was that the IMF would focus on short-term balance-of-payments financing (the current account of the balance of payments), while the World Bank would concentrate on long-term lending (the capital account).⁶ The word “reconstruction” in the World Bank’s proper name referred to reconstruction of war-torn countries. That function, however, was largely taken over by the Marshall Plan, leaving the World Bank to focus almost entirely on lending for the development of underdeveloped countries.

One of Keynes’ ideas, rejected by the conference in favor of the present IMF, was for an International Clearing Union for settling trade balances. In Keynes’ plan, all trading countries would have an account with the Clearing Union, denominated in an international monetary unit called the “bancor,” which would be convertible into each national currency at a fixed rate, like gold. Clearing of accounts would be multilateral; that is, if country A had a surplus with B and a deficit with C, and the two partially cancelled out, then A would have to settle only the difference with the Union.

The innovative feature of Keynes’ plan was that interest would have to be paid to the Union on credit balances as well as debit balances, at least when the balances exceeded a certain amount. In other words, there would be a penalty for running a balance of trade surplus, as well as for running a deficit. All nations would have an incentive to avoid both a surplus and a deficit, to balance their trade accounts, leading to less international debt and reduced capital flows. We think this proposal merits reconsideration today. It is superior to massive IMF bailouts of debtor countries that continue to run trade deficits. It also puts pressure on countries, like Japan, that are addicted to running huge trade surpluses.

■ THE WORLD TRADE ORGANIZATION

The **World Trade Organization (WTO)** is of more recent birth than the IMF and World Bank, having its origins in the General Agreement on Trade and Tariffs (GATT) rather than in the Bretton Woods conference.⁷ The purpose of GATT was to reduce tariffs and other barriers to international trade. The WTO is frequently lumped together with the World Bank and IMF, however, because the three institutions have common

we call throughput. There is not a 1:1 correlation between GNP and throughput, but Delong’s figures show a nearly tenfold increase in GNP from 1950 to 2000. A ninefold increase in throughput from 1945 to the present is probably in the ballpark.

⁶Current accounts and capital accounts are described in detail in Chapter 18.

⁷An international trade organization was proposed at Bretton Woods but was ultimately rejected by the U.S. and replaced by GATT.

policy goals of free trade, free capital mobility, and export-led growth—in other words, globalization. To the extent that the World Bank and the IMF push a policy of globalization, they run into conflict with the internationalist model of world community underlying their charter, a model very different from that of globalization, as we have already emphasized. The WTO's commitment to globalization is evident in the statement of its former director-general, Renato Ruggiero: "We are no longer writing the rules of interaction among separate national economies. We are writing the constitution of a single global economy."⁸ This is a clear affirmation of globalization and rejection of internationalization as defined earlier. Of course, different directors-general may alter policies.

But meanwhile, if the IMF-WB are no longer serving the national interests of their member countries, according to their charters, whose interests are they serving?

What are the advantages of an internationalized economic system over a globalized one? One is that nation-states that can control their own boundaries are better able to set their own monetary and fiscal policy, as well as such things as environmental standards and a minimum wage. Markets hate boundaries, but policy requires boundaries.

It is worth remembering that international free trade is not really trade between nations, but rather trade between private firms or individuals residing in different nations. Their transactions are carried out for the private benefit of the contracting parties, not for the larger benefit of their national communities. The policy of free trade represents the assumption that if these transactions benefit the private contracting parties, then they will also benefit the larger collectives (nations) to which each party belongs. "What's good for General Motors is good for the USA," to recall a famous statement.

Let's apply the assumption to another collective, the corporation. Do corporations allow their individual employees to freely contract with employees of another corporation in pursuit of their own self-interest? This happens, but if the employees are caught, they are usually sent to jail. All transactions initiated by employees are supposed to be in the interests of the corporation, not the employee, and must be judged so by officers of the corporation. Corporations regulate the trades negotiated by their employees. Why should nations not regulate trade negotiated by their citizens? Advocates of an internationalized economy argue that nations must have the ability to regulate their corporate citizens, just as a corporation regulates the behavior of its employees.

⁸From a speech to the United Nations Conference on Trade and Development's (UNCTAD) Trade and Development Board in October 1996. Online: <http://r0.unctad.org/en/special/tb43pr05.htm>.

Advocates of a globally integrated economy argue that nations are obsolete, and that they have been responsible for two world wars in the twentieth century. True enough, it is important to remember the real evils of nationalism. We agree with the Bretton Woods delegates, however, that the answer to nationalism is internationalism, not globalism. “A world with no boundaries” is a good song lyric, but taken literally, it makes policy in the interest of local community impossible. We see world community as a community of local and national communities—as a community of communities. Consequently, the problems we see with the Bretton Woods Institutions arise not from their historical charter, but from their institutional tendency to forget their charter and substitute globalism for internationalism.

■ SUMMARY POINTS

What are the critical points to take home from this chapter? First, globalization is entirely different from internationalization. Under globalization, all national economies are integrated into one global economy and must obey the laws laid down by a global economic institution—currently the WTO. Under internationalization, relations between countries grow increasingly important, but the nation remains the basic unit of community and policy.

Second, the concept of comparative advantage is largely irrelevant in a world of mobile financial capital. We must instead look at absolute advantage. While absolute advantage is likely to lead to greater overall gains, it will also produce both winners and losers on an international level. Without the elegant theoretical conclusion of comparative advantage that free trade is a win-win affair for all countries, it is necessary that we look at the empirical evidence regarding each country's absolute advantage to determine winners and losers under a regime of global integration. In the next chapter, we focus on globalization and its likely consequences.

BIG IDEAS to remember

- | | |
|---|---|
| ■ Comparative vs. absolute advantage | ■ Globalization vs. internationalization |
| ■ Capital mobility and comparative vs. absolute advantage | ■ Bretton Woods Institutions (IMF and World Bank) |
| | ■ World Trade Organization (WTO) |



Globalization

Why are so many smart people such ardent advocates of globalization? Mainly, it is because **globalization** results in more efficient use of resources, resulting in faster rates of global economic growth. But there are other consequences, also intended: (1) an increase in international competition, in which countries must compete against each other for a share of global markets; (2) more intense national specialization according to the dictates of competitive (not comparative) advantage; (3) worldwide enforcement of “trade-related intellectual property rights”; and (4) control over local and national affairs by an international institution. Empirical evidence suggests another important consequence, an unintended one: an increased concentration of wealth within and between countries. It is this last consequence, perhaps, that has sparked the strongest opposition to globalization.

In this chapter we look at each of these consequences in a bit more detail within the context of the policy goals of ecological economics: efficient allocation, just distribution, and sustainable scale.

■ EFFICIENT ALLOCATION

Advocates of globalization claim that free trade is efficient, producing the much-touted “gains from trade.” But efficiency depends on a number of critical assumptions and conditions, including these:

1. There must be a large number of nearly identical firms.
2. Information must be relatively freely shared.
3. There must be strong incentives to internalize costs.

Are the conditions for efficient production being met? We address this question for each condition listed above. We also ask whether

globalization maintains the opportunity for satisfying employment, and the role that opportunity plays in enhancing human welfare.

Perfect Competition vs. Transnational Corporations

The assumption of perfect competition, which requires a very large number of nearly identical firms, is a cornerstone of neoclassical economic theory. Perfect competition weeds out the inefficient and ensures the efficient allocation of resources within both national and international markets.

Globalization forces firms to compete against other firms worldwide. However, in general, only very large businesses have the resources to enter foreign markets. WTO rules actively forbid countries from promoting national small businesses if this can be construed as discrimination against large foreign corporations. Large corporations with economies of scale, or willing to accept low profits in an effort to gain market share, can easily underprice local businesses and either bankrupt or acquire them, thereby reducing the total number of firms. In fact, global mergers and acquisitions have been most intense in the areas of financial services and telecommunications, precisely those economic sectors in which WTO agreements have been completed.¹

Global competitiveness may therefore be incompatible with market competition in a given nation. As a rule of thumb, many economists agree that if 40% of a given market is controlled by four firms, the market is no longer competitive. Such concentration is not at all unusual in the agricultural sector: in the U.S. Midwest, four firms control well over 40% of the trade in most major agricultural commodities,² and the top four agrochemical corporations reportedly control over 55% of the global market.³ Nonetheless, in 1999 the U.S. government approved the merger of the two largest international grain trading corporations, Cargill and Continental Grain, in spite of explicit concerns that the merger might result in monopoly power⁴ (over 80% of international trade is controlled by only ten firms⁵). Ironically, the U.S. assistant attorney general in charge of re-

¹L. Wallach and M. Sforza, *Whose Trade Organization?: Corporate Globalization and the Erosion of Democracy*, Washington, DC: Public Citizen, 1999.

²W. Heffernan, Report to the National Farmers Union: Consolidation in the Food and Agriculture System, Columbia: University of Missouri, 1999. Note that it can be very difficult to assess market concentration.

³Action Group on Erosion, Technology and Concentration, Concentration in Corporate Power: The Unmentioned Agenda, ETC communique #71, 2001. Online: http://www.rafi.org/documents/com_globlization.pdf.

⁴Monopoly is a single seller, and monopsony is a single buyer. In this case, the U.S. Department of Justice was more concerned about prices to farmers than prices to consumers.

⁵G. van Empel and M. Timmermans, "Risk Management in the International Grain Industry." In *Commodities Now*, December 2000. Online <http://www.commodities-now.com/enonline/dec2000/article3/a3-pl.shtml>.

viewing the merger reportedly suggested that even more consolidation would be required to maintain the competitiveness of American agriculture in global markets.⁶

If mergers are indeed required for firms to remain “competitive” in a global economy, will this lead to a more efficient allocation of resources? As Chicago School economist and Nobel laureate Ronald Coase pointed out, firms are islands of central planning in a sea of market relationships.⁷ The islands of central planning become larger and larger relative to the remaining sea of market relationships as a result of mergers. More and more resources are allocated by within-firm central planning, and less by between-firm market relationships. Of the 100 largest economic organizations, more than half are corporations. One-third of the commerce that crosses national boundaries does not cross a corporate boundary; it is an intra-firm, nonmarket, transfer. Is there any reason that central planning should work better for a large corporation than it does for a nation?

Patents and Monopolies

At the global scale, intellectual property rights are tied to trade. Why? It is hard to trade property if the legal right to the property is in dispute. As a result, the Agreement on Trade Related Aspects of Intellectual Property (TRIPs) requires all WTO signatories to protect intellectual property rights (IPR) for 20 years,⁸ with violators subject to trade sanctions and fines.

In chapter 10, we discussed two major inefficiencies associated with patents. First, information is a nonrival good, and making it excludable leads to inefficiencies. Second, patents are nothing more than temporary monopolies, and monopolies are inherently inefficient. We also discussed the counterargument for patents, that unless we provide the economic incentive of monopoly ownership for a significant period of time (20 years, they suggest), little new knowledge and innovation will be forthcoming.

THINK ABOUT IT!

Do you remember why it is inefficient to make nonrival goods excludable? Do you remember why monopolies are inefficient?

Although patents have existed in England since the seventeenth century, in the United States and France since the 1790s, and in most of Europe since the 1880s, international patents are a fairly recent

⁶Reported in A. Cockburn, and J. St. Clair, “How Three Firms Came to Rule the World.” In *Counterpunch*, November 20, 1999. Online: www.counterpunch.org. A competitive market is defined as one with enough firms that all firms are price takers and none are price makers. Mergers in highly concentrated markets by definition make those markets less competitive, not more.

⁷R. Coase, The Nature of the Firm, *Economica* 4(16): 386–405 (1937).

⁸Wallach and Sforza, op. cit.

phenomenon. They have been practical only since the International Patent Institute was established at the Hague in 1947. Until the 1980s, patents played a relatively unimportant role in international commerce. Empirically it seems to be the burst of inventions that has stimulated demand for greater patent protection, more than vice versa.

In 1790 Samuel Slater, the “Father of American Industry,” essentially stole the design for the first American textile factory from Richard Arkwright, the English industrialist.⁹ Currently corporations and individuals from developed countries own 97% of all patents, and the WTO provides mechanisms for enforcing these patents globally. Is this likely to encourage or discourage a new Samuel Slater, a father of industry in a country that truly needs it? At the very least, it is difficult to argue that technology has foundered so much since the 1970s that we need a substantial expansion of patent protection under the WTO to stimulate its advance.

THINK ABOUT IT!

Make a list of the most important contributions, discoveries, or inventions in your major field of study. How many of them were motivated by patentability of intellectual property? There might be interesting differences between fields of study.

As economist Joseph Schumpeter emphasized, being the first with an innovation already gives one a temporary monopoly by virtue of novelty. In his view, these recurring temporary monopolies were the source of profit in a competitive economy whose theoretical tendency is to compete profits down to zero. This is the very condition of economic efficiency—why thwart it?

This is not to say that we should abolish all intellectual property rights. Such an action would create more problems than it would solve. But we should certainly begin restricting the domain and length of patent monopolies rather than increasing them so rapidly and recklessly. And we should become much more willing to share knowledge. Shared knowledge increases the productivity of all labor, capital, and resources—things that are inherently scarce, rival, and excludable. Knowledge is not inherently scarce and is the quintessential public good—nonrival and nonexcludable—even though patents make it artificially excludable.

One important and practical policy implication of these considerations is that international development aid should consist far more of freely shared knowledge, and far less of foreign investment and interest-bearing loans. Let’s recall the following words from John Maynard Keynes, one of the founders of the Bretton Woods Institutions:

⁹The Story of Samuel Slater, Slater Mill Historic Site, Website, <http://www.slatermill.org/html/history.html>.

Box 18-1**DISCOVERIES NOT MOTIVATED BY PATENT MONOPOLY PROFITS**

No doubt many important inventions have been stimulated by patent rights. However, the heliocentric view of the universe, gravity, the periodic table of elements, electromagnetic theory, as well as the laws of optics, mechanics, thermodynamics, and heredity were all discovered without the benefit of intellectual property rights and the profit motive. Mathematics has been called the language of the universe. Where would our technology be without it? While no culture has ever allowed a patent on mathematical theorems, mathematicians keep producing new ones.^a Nor has anyone ever had intellectual property rights to the English language, or to fire, the wheel, or money. Yet all these things somehow came into being. The invention of the shipboard chronometer, necessary for navigational calculation of longitude, was stimulated by a one-time prize, not a 20-year patent monopoly. Even economists work long and hard to produce economic theories that are not patentable. Alfred Marshall got no royalties from users of supply and demand and elasticity. J. R. Hicks expected no royalties, and got none, for developing the IS-LM model, and the proper concept of income.

In fact, it is difficult to name a single modern invention that does not depend on ideas freely shared from their first conception. While patent rights have stimulated important inventions, that is less than half the story. In the words of Lawrence Lessig:

Free resources have always been central to innovation, creativity and democracy. The roads are free in the sense I mean; they give value to the businesses around them. Central Park is free in the sense I mean; it gives value to the city that it centers. A jazz musician draws freely upon the chord sequence of a popular song to create a new improvisation, which, if popular, will itself be used by others. Scientists plotting an orbit of a spacecraft draw freely upon the equations developed by Kepler and Newton and modified by Einstein. Inventor Mitch Kapor drew freely upon the idea of a spreadsheet—VisiCalc—to build the first killer application for the IBM PC—Lotus 1-2-3. In all of these cases, the availability of a resource that remains outside of the exclusive control of someone else—whether a government or a private individual—has been central to progress in science and the arts. It will also be central to progress in the future.^b

^aD. S. Evans, *Who Owns Ideas? The War Over Global Intellectual Property*, Foreign Affairs 81(6): 160–166 (November/December 2002).

^bL. Lessig, *The Future of Ideas: The Fate of the Commons in a Connected World*, New York: Random House, 2001. Online: <http://music.barrow.org/2002/Q3/free/page3.htm>.

*I sympathize therefore, with those who would minimize, rather than those who would maximize, economic entanglement between nations. Ideas, knowledge, art, hospitality, travel—these are the things which should of their nature be international. But let goods be homespun whenever it is reasonably and conveniently possible; and, above all, let finance be primarily national.*¹⁰

Externalizing Costs

As we also discussed in Chapter 10, a necessity for efficient markets is that producers pay the costs of production, and they produce to the point where marginal costs are just equal to marginal benefits. This condition is not met when externalities exist, and there are several ways in which economic globalization increases the quantity and severity of externalities.

The goal of the WTO is to increase economic growth and the transport of goods between countries, and both growth and fossil fuel-based transportation are accompanied by significant externalities. At the national level, laws exist to reduce externalities, but the WTO has the power to challenge these laws, and the ability to enforce its decisions. While technically countries are allowed to pass environmental legislation, the WTO frequently declares such laws barriers to trade, and even the threat of a WTO ruling can deter lawmakers. For example:

1. Challenged by Venezuela, the United States was forced to allow the import of gasoline that does not comply with U.S. Clean Air Act regulations.
2. The WTO ruled against the U.S. Endangered Species Act, which prohibits the import of shrimp from countries that do not mandate turtle excluder devices.
3. Under GATT, Mexico won a decision against the U.S. Marine Mammal Protection Act's dolphin-safe tuna provision. Under threats by Mexico of WTO enforcement action, President Clinton and Vice President Gore took the lead in getting Congress to weaken the offending law.
4. Australia's law strictly limiting the import of raw salmon, designed to prevent domestic stocks from contamination with foreign bacteria, was declared a barrier to trade. Scientific studies showed that the risk of infection existed, but the WTO ruled that the probability of infection also had to be shown to justify import restrictions.¹¹

¹⁰J. M. Keynes, "National Self-Sufficiency." In Donald Muggeridge (ed.), *The Collected Writings of John Maynard Keynes*, vol. 21. London: Macmillan and Cambridge University Press, 1933.

¹¹Unfortunately, many of the ecosystem goods and services threatened by economic growth and free trade are characterized by uncertainty and ignorance, in which case by definition probabilities of possible outcomes cannot be determined.

In each of these cases, the overturned regulations had been put in place by relatively democratic governments to reduce negative externalities affecting nonmarket goods and services. A number of other environmental laws are currently threatened by the WTO.¹²

Standards Lowering Competition

At the same time that the WTO makes it difficult for countries to legislate against externalities, the need to compete for market share reduces national incentives to legislate against externalities in what is known as standards-lowering competition (a race to the bottom). The country that does the poorest job of internalizing all social and environmental costs of production into its prices gets a competitive advantage in international trade. More of world production shifts to countries that do the poorest job of counting costs—a sure recipe for reducing the efficiency of global production. As uncounted, externalized costs increase, the positive correlation between GDP growth and welfare disappears, or even becomes negative. Recall the prescient words of John Ruskin: “That which seems to be wealth” becomes in verity the “gilded index of far reaching ruin.” The first rule of efficiency is “count all the costs,” not “specialize according to comparative advantage.”¹³

One way to confront the race-to-the-bottom tendency is to argue for harmonization of cost-accounting standards across countries. This is certainly logical and in line with global integration. If all countries internalized external social and environmental costs to the same degree, there would be no incentive for mobile capital to move to the country that did not internalize these costs because such countries would not exist. It would be hard to negotiate such a global harmonization agreement. There are, in fact, good reasons why different countries have different cost-accounting practices. In any case, it might be argued that countries should measure costs according to their own values, not “international standards.” The traditional comparative advantage argument is compatible with each country’s measuring costs as it pleases. As we saw in Table 17.1, a-units and b-units, which might reflect totally different theories of value, need never be compared in the comparative advantage system. But with capital mobility and absolute advantage comes the necessity to compare a-units to b-units, and the problem of standards-lowering competition to attract mobile capital.

¹²All examples are from Wallach and Sforza, op. cit.

¹³J. Ruskin, “Unto This Last.” Online: <http://www.nalanda.nitc.ac.in/resources/english/etext-project/economics/Ruskin.pdf>.

Box 18-2 WEALTH, POWER, AND EFFICIENCY

Another potential problem with economic globalization is an increase in rent-seeking behavior in the form of lobbying by large corporations and wealthy individuals to influence policy.^a If large corporations are “islands of central planning,” showing less growth than smaller corporations probably as a result of the inefficiencies of central planning, why do they continue to thrive? One possibility is that the concentrated wealth of large corporations readily translates into political power, and large corporations can use this power to promote policies that allow them to thrive in spite of any inefficiencies inherent to centrally planned economies.

Large corporations routinely help politicians set not only domestic rules of the game but also international rules. The trade advisor to President Nixon was a vice president of Cargill, the world’s largest grain exporter. President Reagan relied on a Cargill employee to draft the U.S. agricultural proposal for GATT.^b President Clinton appointed Monsanto CEO Robert Shapiro as a trade representative to the WTO. President George W. Bush relied on Enron CEO Kenneth Lay when designing energy policies. The WTO meetings in Seattle in 1999 were primarily sponsored by large corporations. Can we be sure that this advice and assistance come with no strings attached?

^aRecall that rent is profit over and above the normal profits of operation.

^bK. Lehman and A. Krebs, “Control of the World’s Food Supply.” In Mander and Goldsmith, eds. *The Case Against the Global Economy*, San Francisco: Sierra Club Books, J. Mander and E. Goldsmith (1996).

Specialization and Diminished Well-Being

If an important goal of economic systems is to increase human well-being, then the specialization accompanying globalization is another source of inefficiency. Free trade and free capital mobility increase pressures for specialization according to competitive (absolute) advantage. Therefore, as noted earlier, the range of choice of ways to earn a livelihood becomes greatly narrowed. In Uruguay, for example, everyone would have to be either a shepherd or a cowboy in conformity with the dictates of competitive advantage in the global market. Everything else should be imported in exchange for beef, mutton, wool, and leather. Any Uruguayan who wants to play in a symphony orchestra or be an airline pilot should emigrate.

Most people derive as much satisfaction from how they earn their income as from how they spend it. Narrowing that range of choice is a welfare loss uncounted by trade theorists. Globalization assumes either that

emigration and immigration are costless or that narrowing the range of occupational choice within a nation is costless. Both assumptions are false.

While the range of choice in earning one's income is ignored by trade theorists, the range of choice in spending one's income receives exaggerated emphasis. For example, the United States imports Danish butter cookies and Denmark imports U.S. butter cookies. The cookies cross each other somewhere over the North Atlantic. Although the gains from trading such similar commodities cannot be great, trade theorists insist that the welfare of cookie connoisseurs is increased by expanding the range of consumer choice to the limit.

Perhaps, but could not those gains be had more cheaply by simply trading recipes? One might think so, but recipes (trade-related intellectual property rights) are the one thing that free traders really want to protect.

■ SUSTAINABLE SCALE

While globalization advocates laud efficiency, their goal is not simply more efficient production of what we now produce, but rather ever greater production. If the sole purpose or even the major priority of international trade is to promote growth in GDP with little or no attention paid to scale, in the long run, a “successful” trade regime will lead us beyond the sustainable scale for the global economy. This is true no matter how efficient the allocation of resources between countries. It should already be clear that greater externalities and standards lowering competition pose threats to sustainable scale.

Two other issues bear mentioning. First, the integration into one global system gives us only one chance to see if the system works—we cannot learn from our mistakes. Second, the negative environmental impacts of our consumption may occur in another country, where they are that much more likely to be ignored.

Learning from Our Mistakes

In the past, numerous civilizations have crumbled as they have surpassed ecological barriers. Examples are the civilization on Easter Island, the Mayan empire, and the early civilizations of the Fertile Crescent. Fortunately, these were isolated incidents in which only the local carrying capacity was overwhelmed, and today they can serve as examples of mistakes we cannot afford to make. However, as trade expands, local limits to scale become less relevant and global limits more so. While trade may decrease the chances of surpassing sustainable scale in any one area, it also means that if we do surpass it, we are more likely to do so for the planet as a whole. Consequently, it becomes more difficult to learn from

our mistakes as we go. Thus, globalization requires us to get it right the first time.

Out of Sight, Out of Mind

Even if globalization did not lead countries with high environmental standards to lower them, international trade can make it easier to ignore the costs of economic growth. In recent decades, as the most developed nations saw their environments deteriorating, they passed laws to control some types of pollution and resource depletion. To some extent this led to greater efficiency, decreased consumption of polluting products, and improved technologies for controlling pollution, but in many cases it seems to have led to the relocation of polluting and resource extracting industries to countries without such laws.¹⁴ The environments in the developed countries improved at the expense of the poorer countries. With the spatial connection between economic growth and environmental damage severed, many people seem to believe that the causal connection has been severed as well. Indeed, many economists now claim that environments in the developed countries improved precisely because of economic growth.

In reality, the net impact of relocation of environmentally damaging industries on scale can be highly negative. For example, when Australia's wet tropical rainforests were declared a World Heritage site (largely due to pressure from environmentalists) and the region's reasonably well-managed logging operations were shut down, total timber consumption in Australia did not decrease. Instead, Australia has substituted its own tropical timber supply with timber from tropical countries with worse logging practices. The net outcome is likely a greater loss of ecosystem services worldwide.

Similar, and perhaps more threatening, is the relocation of waste from toxic industries. Russia is currently discussing the establishment of nuclear and toxic waste processing facilities. These facilities will allow the overdeveloped countries to reduce further degradation of their environments.¹⁵ However, they may well result in less careful waste disposal than would have occurred in the country where the waste originated. In this case, simply transporting the wastes would increase the danger of negative environmental impacts.

We already know that markets fail to signal many environmental costs. In a democracy, when people are exposed to environmental externalities,

¹⁴D. Rothman, Environmental Kuznets Curves—Real Progress or Passing the Buck? A Case for Consumption-Based Approaches, *Ecological Economics* 25: 177–194 (1998).

¹⁵Overdeveloped countries are defined as those whose level of per-capita resource consumption is such that if generalized to all countries could not be sustained indefinitely. See H. Daly, *Beyond Growth: The Economics of Sustainable Development*, Boston: Beacon Press, 1996.

they can signal their preferences through political institutions. If people in the developed democracies export their wastes or environmentally damaging industries to less democratic countries, we lose this signal of environmental scarcity. The first rule of cost internalization is to internalize costs to the firm that generates them. If we fail to do this, then we must at least internalize costs to the country under whose laws the firm was operating when it generated the externalities. The second rule could be enforced by prohibiting the export of toxic waste.

Positive Aspects of Trade with Respect to Scale

We have thus far presented an incomplete picture of the impacts of globalization on scale. In the absence of international trade, appropriate scale (in terms of economic activity and human populations) would be determined at the national level and by the most limiting factor. For example, one country might have abundant agricultural land but inadequate mineral resources. In other countries, scale might be limited by land area, mineral resources, or fuel supplies, in yet others by waste absorption capacity, rainfall, or agricultural productivity. International trade can help alleviate the most limiting constraints on scale within each country. If international trade suddenly ceased, some countries would find themselves well beyond sustainable scale, and even beyond desirable scale in the short term.¹⁶

Efforts to sustain a high standard of living for too large a population with limited resources would no doubt force some countries to liquidate natural capital—for example, extend the agricultural frontier to lands that cannot sustain it, or burn their forests to meet energy needs. Other countries might be forced to mine low-quality, highly polluting fossil fuels.

International trade can help sustain larger populations with higher levels of material consumption than isolated national economies alone could sustain. Unfortunately, this happy outcome is likely only if sustainable scale and equitable distribution are explicit and are the principal goals of international trade. It is more likely to occur under internationalization than globalization.

■ JUST DISTRIBUTION

Finally, we turn to the impact of globalization on distribution. Proponents of globalization claim that it will bring about “a world free of poverty” (the professed goal of the World Bank). Will globalization achieve this goal?

¹⁶Surpassing sustainable scale means that the sustaining ecological system must eventually collapse, and surpassing short-term desirable scale means that the costs of additional growth outweigh the benefits for the current generation.

Next we examine the empirical evidence and explore some theoretical reasons that trade liberalization might not reduce poverty. We conclude with a brief look at the most important of commodities—food.¹⁷

Absolute Disadvantage

In the presence of capital mobility, money will logically flow to wherever there is an absolute advantage of production. The world's poorest countries may be poor precisely because they are inefficient at producing nearly everything. If this is true, then the countries most likely to suffer from globalization are in fact the very poorest.

Does this conclusion have any empirical support? According to the IMF, most developing countries have failed to raise their per-capita incomes toward those of industrial countries. United Nations Development Programme statistics show that in the three decades prior to 1996, the share of income received by the world's poorest 20% fell from 2.3% to 1.4%, while the share going to the world's richest 20% increased from 70% to 85%. Still, these statistics refer only to relative income, not absolute income. The world's poorest 20% have seen some gains in income over the past 40 years.¹⁸ Globalization on a significant scale, however, is a fairly recent phenomenon. What has happened to the poorest of the poor more recently?

Using the World Bank's Global Development Network Growth Database,¹⁹ we calculated the average increase in per-capita income from 1989 to 1999 for the 15 poorest countries at the start of that period²⁰ for which data are available.²¹ All of these countries are sub-Saharan African nations, and with the exception of natural resource endowments in some of them, they have few absolute advantages in global competition. We found that real per-capita income in these countries (as measured by GDP) ac-

¹⁷In spite of the recent trend toward privatization of water, we do not consider water a commodity. Water is not produced for sale; it is produced by nature. Privatization is basically an enclosure of the commons, and a very inefficient one since it invariably creates a monopoly. (How many different water companies can you buy water from? How many water lines lead into your house?)

¹⁸E. Kapstein, "Distributive Justice as an International Public Good: A Historical Perspective." In I. Kaul, I. Grunberg, and M. Stern, eds. *Global Public Goods: International Cooperation in the 21st Century*. New York: Oxford University Press, 1999.

¹⁹W. Easterly and M. Sewadeh. Online: <http://www.worldbank.org/research/growth/GDNdata.htm>.

²⁰Per-capita incomes are measured in real GDP in 1989. We conservatively used 1989 instead of 1999, as the later date would automatically select for countries that were less likely to have experienced growth.

²¹Ethiopia, Chad, Democratic Republic of the Congo, Niger, Burkina Faso, Malawi, Burundi, Tanzania, Uganda, Mali, Comoros, Central African Republic, Myanmar, Togo, Madagascar, and Guinea-Bissau.

tually *decreased* by an average of 5.5% over that period. Concerned that war in two of these countries might bias results, we looked at the 15 poorest countries that had not experienced war.²² We still found that these countries suffered on average a 3.2% *decrease* in income. In contrast, the wealthiest 15 countries in 1989 experienced an average increase in per-capita income of 15.5% by 1999. In absolute terms, the same amount of money required to boost per-capita income in the U.S. by 1% would double the per-capita income of the 24 poorest countries (for which we have data) in the world today.

If we lived on an infinite planet in which one person's consumption had no impact on anyone else, and if human nature did not lead us to measure our wealth in comparison to others, it would make no difference to the poor countries what happened in the wealthy ones. The fact is, however, that we do live on a finite planet. The increase in income in the wealthy countries is fueled by nonrenewable resource consumption (including unsustainable depletion of potentially renewable resources), which means that these resources are not available for future improvements in the well-being of the poorest. And resource use generates a corresponding amount of waste and accompanying damage to public good ecosystem services that would otherwise benefit these poorest countries.

This observation draws our attention to a fact otherwise obscured by the data. Most of these poorest countries were involved in international trade in the one area where they might have an absolute advantage: the extraction and export of natural capital. The revenue they received from both export and domestic sales of these resources counted as part of their income. Without this revenue, income as measured by GDP would have fallen even more. Yet as you will recall, we earlier defined income as the amount you can consume in one period without affecting your ability to consume in subsequent periods. Thus, revenue from nonrenewable natural resource extraction cannot be counted entirely as income, and the situation of these poorest countries is even worse than it appears.

Of course, the evidence presented here says nothing conclusive about globalization. One could argue, as economists often do, that the problem was insufficient liberalization. Perhaps things would have been even worse in the poorest countries without globalization. What happened in the countries that most avidly pursued economic liberalization?

Many countries have shown periods of economic growth as their economies have liberalized. For example, Argentina was considered one of the most avid converts to economic liberalization in Latin America in recent years, and in real terms, per-capita income increased by over 40%

²²We dropped Ethiopia and the Democratic Republic of the Congo and replaced them with Zambia and Guinea.

from 1990 to 1998.²³ Yet between 1987 and 1997, the income gap between the richest 10% and the poorest 10% increased from 15 to 1 to 24.8 to 1. The proportion of people in urban areas living below poverty fell for the first few years of the 1990s, coinciding with the control of inflation, but from 1993 to 1998, poverty rates increased from 27.3% to 35.8%. With the collapse of the economy in December 2001, poverty rates soared to over 50%. For Latin America as whole, trade liberalization has been accompanied by an increasing income gap between the richest 10% and the poorest from 37 to 1, already the worst in the world, to 48 to 1. Poverty rates in Mexico, Argentina, and Venezuela, among others, have soared over the last 7 years.²⁴ Just as importantly, the growth that has occurred there has been marked by intense instability resulting from the volatility of international capital flows, as we will discuss below.

While there is no proof that globalization has been the culprit in this decline, there is even less evidence that globalization is a cure for poverty.

Standards Lowering Competition and Labor

We discussed above how countries pressured by global competition may ignore external costs to the environment in a race to the bottom. To remain competitive, countries may similarly need to accept or even promote lower labor costs.

In the United States and Europe, an implicit social contract has been established to ameliorate industrial strife between labor and capital. Specifically, a just distribution of income between labor and capital has been taken to be one that is more equal within these countries than it is for the world as a whole. Global integration of markets necessarily abrogates that social contract. There is pressure on American and European wages to fall because labor is relatively much more abundant globally than nationally. By the same logic, returns to capital in these countries should increase because capital is relatively more scarce globally than nationally. This could lead U.S. income distribution, already approaching Third World ranges of inequality, to become even more unequal. Theoretically, one might argue that wages would be bid up in the rest of the world. But the relative numbers make this a bit like saying that, theoretically, when I jump off a ladder, gravity not only pulls me to the ground, it also moves the ground toward me.

In general, if a country pursues economic growth by developing the ex-

²³World Bank's Global Development Network Growth Database. GDP in nominal dollars increased by 260%.

²⁴A. Faiola, "Argentina's Lost World: Rush into the New Global Economy Leaves the Working Class Behind," *Washington Post*, December 8, 1999. Argentina's poverty rate has skyrocketed since the collapse of its economy in December 2001.

port sector, it must be able to sell what it produces on the highly competitive global market, and costs must be kept low. In a world of mobile capital, absolute advantage in production is what counts. Most less-developed countries (LDCs) do not have advanced technologies that can lower production costs, a well-developed infrastructure that can lower transportation costs, or institutions that make investments particularly safe. Instead, they have two sources of absolute advantage: abundant labor and (in some cases) abundant natural resources.

For export-oriented industrial production, the only absolute advantage LDCs generally have is the low wages they can offer. Until these countries develop some other source of absolute advantage, wages and benefits must be kept down to remain competitive on the world market. This does little to help alleviate poverty, and it frequently requires the suppression of labor rights.

What happens if instead a country seeks to industrialize to serve the needs of the domestic market? Obviously the prerequisite for this to occur is that a domestic market actually exists. A market can only exist if there is purchasing power, and purchasing power requires wages. This is why Henry Ford (no friend of labor) chose to pay his workers \$5.00/day (at the time an exceptionally high wage)—he wanted them to be able to afford the cars they were producing. Thus, for LDCs, a focus on liberal international trade will tend to push wages down, while a focus on production for the domestic economy may tend to push wages up.

When confronted with this argument, people might point to the Asian Tigers (Taiwan, Korea, Singapore, Hong Kong, and more recently, Thailand and Malaysia). These countries, like Japan, pursued export-oriented industrialization and saw dramatic increases in their standards of living. Yet on closer examination, the historical record of the Asian Tigers actually bolsters the argument for developing the domestic market. First, these nations are characterized by highly protectionist policies and a high degree of government intervention, not by open markets. More to the point, their initial successes were greatly facilitated by strong domestic markets. In Korea, Taiwan, and Japan, agrarian reform preceded industrialization. In these predominantly agrarian societies, the transfer of land to small farmers allowed the farmers to accumulate and spend the surplus they generated. As Dr. Sun Yat-Sen stated in the *Son Mm Chu-I* (the Three Principles of the People), “industrialization should follow, not precede, the building up of the internal capacity to consume.” Successful land reform in Taiwan doubled the purchasing power of the farmer at a time when Taiwan was an agricultural economy.²⁵ In their early

²⁵Quoted in F. Harrison, *Five Lessons for Land Reformers: The Case of Taiwan*. Reprinted from *Land & Liberty* May–June (1980). Online: http://www.cooperativeindividualism.org/harrison_taiwan_land_reform.html.

industrialization efforts, the Asian Tigers focused on import-substituting industrialization (ISI). Captive markets allowed them to develop the skills necessary for global competition.²⁶

On the other hand, production for the domestic economy can have a negative impact on real wages as well. If a country currently importing industrial goods decides instead to replace them, the new industries will have a hard time competing against established producers. Thus, ISI generally requires tariffs and quotas on imports. Such tariffs will stimulate demand for domestic goods (benefiting the producer) but drive up the price of imports in order to make the import substitute more competitive, and lower real wages.

In terms of distribution relative to export-oriented economies, however, this impact may not be as dire as it seems. First, export-oriented countries typically undervalue their currencies to make exports cheap and imports expensive, so there may be little difference between ISI and export promotion with respect to consumer prices. In addition, the easiest goods to produce are often the cheapest, those purchased by the working masses (soap, aspirin, matches, etc.). With simple production technologies, the LDCs are not at a serious disadvantage when producing these goods, and tariffs can be kept quite low. Luxury goods, on the other hand, tend to be more technologically sophisticated and may require higher tariffs. However, it is the wealthier classes that purchase these goods, and they can better afford to pay the tariffs.

As the industry develops, two things happen. First, efficiency should improve, and tariffs can be reduced—especially if the producers know that tariffs will be reduced and therefore have an imperative to improve efficiency. Second, any market based solely on a wealthy minority will quickly become saturated. Further industrialization will depend on larger markets, which can be created through higher wages for the labor force. Once industrial capacity is well developed and production techniques have been refined, a country may be able to compete on the export market without forcing wages down to the global minimum. This is essentially the strategy that was pursued by the newly industrialized Asian economies.²⁷

It is also worth remembering that in the 1960s and 1970s, the countries engaged in ISI often showed the highest growth rates, and they were touted as examples for others to follow. Brazil's economic growth under this strategy was considered miraculous. Economic wisdom is surprisingly fickle.

Probably the most important lesson to take from this discussion is that

²⁶E. Vogel, *The Four Little Dragons: The Spread of Industrialization in East Asia*, Cambridge, MA: Harvard University Press, 1991.

²⁷*Ibid.*

one size never fits all. Different cultures are likely to require different approaches to development, and what works under one set of global economic conditions might fail miserably under another. Economists would do well to keep this in mind.

Food Security and Free Trade in Agriculture

Potentially the most serious source of unfairness in the liberalization of trade is the threat it poses to food security. “Free” trade in food threatens security in two important ways. First, the market system provides goods and services to those who have the money to purchase them. If in the future the WTO or other international agreements succeed in liberalizing trade in agriculture, poor citizens of LDCs will be competing with the rich citizens of ODCs (overdeveloped countries) for food. In his groundbreaking study of famines, Amartya Sen has shown that famines are generally the result of a lack of entitlements to food rather than a lack of food itself.²⁸ In the market economy, this simply means a lack of money to purchase food, even when actual supplies are abundant. The situation can occur because of unemployment or a decrease in the value of the goods some group produces relative to food. In the presence of international trade, the domestic sector must bid against the rest of the world for food purchases. If the economy suffers a recession or there is a currency devaluation, local ability to purchase food decreases relative to global ability, and food may be exported even as the local population starves. Clearly, international trade can be critical in addressing famines that are caused by food availability decline, but only if countries have the resources to purchase that food. If agricultural markets are completely liberalized, it is easy to imagine Western nations importing food for their cattle from nations suffering famine.

The fact that farmers are typically the most disadvantaged group in many LDCs brings a second source of unfairness if agriculture is liberalized. LDC costs of agricultural production tend to be higher than those of large agro-industrial farms in ODCs (in part because so many negative externalities of industrial agriculture are not internalized). This means that trade liberalization will often decrease prices for food in LDCs. While lower food prices may help the urban poor and wage earners, it can also cause lower incomes and declines in welfare for the poorest group, namely farmers. Low food prices reduce incentives for domestic agriculture. Theoretically, under trade liberalization, these poor farmers should be able to grow cash crops for export. Unfortunately, cash crops often

²⁸A. Sen, *Poverty and Famines: An Essay on Entitlement and Deprivation*, 6th ed. New York: Oxford University Press, 1992.

require higher levels of inputs and are far riskier than traditional food crops that have been bred for millennia to minimize the risk of failure. While average returns over several years may be higher with cash crops even with more frequent failures, people do not eat “on average”—they eat every day.

Box 18-3 PUBLIC LAW 480 AND FOOD SECURITY

If one country becomes dependent on others for food supply, they run serious risks to their autonomy. For example, the U.S. Public Law 480 provides food at subsidized cost to LDCs. While nominally a gesture of benevolence, American politician Hubert Humphrey once said in reference to this law: “I have heard . . . that people may become dependent on us for food. I know that was not supposed to be good news. To me that was good news, because before people can do anything they have got to eat. And if you are looking for a way to get people to lean on you and to be dependent on you, in terms of their cooperation with you, it seems to me that food dependence would be terrific.”^a

Secretary of Agriculture Earl Butz also referred to food as a weapon and “one of the principal tools in our negotiating kit.”^b Understandably, countries that come to depend on food imports may not share Humphrey’s enthusiasm for that dependence.

^a*Sen. Hubert H. Humphrey, in naming P. L. 480 the “Food for Peace” program, Wall Street Journal, May 7, 1982.*

^b*USDA Secretary Earl Butz, 1974 World Food Conference in Rome.*

SUMMARY POINTS

We are better served by a process of internationalization in which countries are free to act on their own information to address local problems of scale and distribution (areas where the market manifestly fails) in a culturally sensitive manner. We must also carefully analyze the actual and potential impacts of globalization on scale, distribution, and efficiency.

Evidence suggests that globalization may be undermining the conditions required for efficient market allocation, by creating fewer, bigger firms, more negative externalities, and more monopolies on nonrival information. More negative externalities and increased economic growth, coupled with a limit on the national ability to regulate externalities, is a threat to sustainable scale. Empirical evidence also suggests that globalization under the principle of absolute advantage may simply reinforce existing patterns of winning and losing, leading to even greater concentration of wealth, both within and between countries.

BIG IDEAS to remember

- Perfect competition vs. mergers, patents, and transnational corporations
 - Trade vs. environment in the WTO
 - Standards-lowering competition
 - Wealth and power
 - Specialization and welfare
 - Trade and scale
 - Distribution and globalization
 - Food security and international trade
-



International Flows and Macroeconomic Policy

National policies require national boundaries. Yet, as globalization eradicates boundaries, even its advocates agree that “the world will move from closed, nationally controlled systems toward one open, global system under no one’s control.”¹ If no one is in control, no one makes policies. We have already discussed some of the ways in which globalization undermines the ability of nations to determine their own policies concerning the environment. We will now turn our attention to the impact of globalization of financial capital on macroeconomic policy.

Most financial capital flows are in the forms of electrons flitting from one computer to the next, but such flows can have serious impacts not only on the output of real goods and services, but also on the ability of governments to enact economic policies. According to the IMF, “globalization may be expected to increasingly constrain governments’ choices of tax structures and tax rates.”² The IMF also claims that it is the nation-state that must address issues of distribution and social welfare. Yet the pursuit of global free markets weakens the national policy levers necessary to achieve these goals. We will explain in greater detail how this happens, but an adequate explanation first requires a certain amount of background material on the balance of payments and exchange rate regimes.

¹L. Bryan and D. Farrell, *Market Unbound: Unleashing Global Capitalism*, New York: Wiley, 1996.

²International Monetary Fund, *World Economic Outlook*, Washington, DC: IMF, 1997.

■ BALANCE OF PAYMENTS

There are two basic types of economic transactions between residents of different countries: the exchange of real goods and services and the exchange of assets. The net outcome of these transactions is measured by the **balance of payments (BOP)**. The BOP has two components, current accounts and capital accounts, corresponding to the two types of transactions.

The **current account** measures the exchange of real goods and services as well as transfer payments. Goods and services are generally consumed in the current period, which is why it is called the current account. Real goods, of course, are market goods. Services include interest payments on loans, royalties on intellectual property, profits earned on investments abroad, and similar transactions. Transfer payments include money workers abroad send home to families, grants to foreign countries, and similar transactions. **Capital accounts** include stocks, bonds, and property abroad. These items are not consumed. They are a stock of capital yielding a flow of revenue.

Money flowing into a country increases the balance on current account or capital account, and money flowing out of a country decreases it. If a country imports more goods and services than it exports, it runs a deficit in its current account, and if it exports more than it imports, it runs a surplus. Similarly, if foreigners purchase more assets in the home country than the home country purchases abroad, the home country runs a capital account surplus, and if the opposite is true, it runs a deficit. Note that a surplus in the current account can balance out a deficit in the capital account, and vice versa, to keep the balance of payments neutral.

In general, countries try to keep their balance of payments neutral, running neither a surplus nor a deficit. In an accounting sense, formal balance (capital plus current account) is guaranteed by a residual balancing item, formerly gold under the gold standard, but is now guaranteed by changes in a nation's foreign exchange reserves, short-term IMF credits (special drawing rights or SDRs), and IOUs from deficit countries called short-term capital flows.

■ EXCHANGE RATE REGIMES

An **exchange rate** is the amount of one currency you would have to pay to receive a given amount of another. For example, in May 2003 one U.S. dollar could purchase about 70 Angolan kwanza. Any international exchange of goods, services, or assets requires an exchange of currencies, as those who sell generally want to be paid in their own currency. How the exchange of currency takes place depends on the exchange rate regime

The *balance of payments* (BOP) is the sum of the current account and capital account.

The *current account* measures the exchange of real goods and services as well as transfer payments—things consumed in the current period.

The *capital account* measures stocks, bonds, and capital abroad—stocks of capital yielding flows of revenue.

used by a particular country, and this regime also determines how economic globalization affects domestic policy levers. There are two basic types of regime: fixed and flexible. When discussing exchange rates, “floating” is synonymous with “flexible.”

In a **fixed exchange rate regime**, the value of one country’s currency is pegged to that of another country. For example, from 1991 to 2002, the Argentine peso was pegged to the U.S. dollar at a one-to-one ratio; that is, the Central Bank of Argentina agreed to sell as many dollars as anyone wanted for one peso each. Obviously, the central bank needed to have foreign currency on hand to make this exchange. Ideally, this currency came from demand for Argentine products. When an American wanted to buy beef from Argentina or Argentine bonds, the central bank would also sell the required pesos for \$1 each.

A problem can occur when nationals of a fixed rate country consistently want to buy more (or fewer) goods, services, and assets from other countries than other countries want to purchase from them, that is, when the fixed rate country is running a BOP deficit (or surplus). If the fixed rate country runs a BOP surplus, it will accumulate excess reserves of foreign currency for which there is insufficient demand. This will put pressure on the country to revalue its currency—for example, offer more dollars per Argentine peso. If the fixed rate country runs a BOP deficit for too long, it will run out of foreign reserves. In this case, there will be pressure to borrow money from abroad and/or devalue its currency—offer fewer dollars per Argentine peso. This is what happened to Argentina when it was forced to devalue its currency in January 2002.

In a **flexible exchange rate regime**, exchange rates are determined by global supply and demand for currencies, and central banks play no direct role. The only foreign currency available for purchase for foreign goods, services, and assets is that which foreigners spend on the purchase of domestic goods, services, and assets. If there is more demand for foreign currency than national currency, the national currency depreciates, and when the opposite is true, the national currency appreciates.³ Market forces determine the exchange rate. This means that the BOP must always be zero—a current account deficit must be financed by a capital account surplus, and vice versa.

Most of the developed countries have flexible exchange rate regimes, but in reality, they are not perfectly flexible. In many cases, countries will

³Under a fixed exchange rate regime, when the domestic currency loses value it is known as devaluation, and when it gains value it is known as revaluation. In either case, this happens directly as a result of central bank policy. Under a flexible exchange rate regime, when market forces cause a currency to lose value, it is known as depreciation, and a gain in value is known as appreciation.

grow concerned that their currencies are overvalued or undervalued, and will buy or sell currency to correct the perceived imbalance. For example, in September 2000, the central banks of the G7 countries⁴ coordinated policies to prop up the euro, which had fallen 30% against other currencies in the year since its release.⁵ Some countries try to keep their exchange rate within a certain range, which is known as a managed float, somewhat of a hybrid between a fixed and flexible exchange rate.

Potentially, countries can manipulate the supply and demand for their own currencies. For example, high tariffs on imports will reduce the demand for imports, and hence the demand for foreign currencies. Alternatively, a country may simply control capital flows, directly determining the supply of its own currency and the demand for foreign currencies. In this case, national policies can determine the BOP under a fixed rate regime or affect the exchange rate under a flexible rate regime.

■ CAPITAL MOBILITY AND NATIONAL POLICY LEVERS

What happens to countries with these two types of exchange rate regimes under economic globalization, when there are no barriers to international exchange of goods, services, and assets (perfect capital mobility)?

Fixed Rate Regimes

Let's take the case of a country in recession. As we discussed in Chapter 16, a country that wants to stimulate economic growth can use either fiscal policy or monetary policy. This is not true, however, in an open economy, where perfect capital mobility means that interest rates between nations cannot vary by much, as money will simply flow to the country offering the highest returns.⁶

For example, what happens when a country with a fixed exchange rate regime wants to stimulate the economy by increasing the money supply? Normally, this approach would lower interest rates, stimulate consumption and investment, and hopefully revive the economy. However, in the absence of restrictions on international capital flows, people can respond to lower interest rates by moving their savings to a country that offers higher rates. To do so, they will sell domestic currency to the central bank in exchange for foreign currency. This reduces the money supply in a process that will theoretically continue until domestic interest rates are ap-

⁴G7 stands for the Group of 7, the seven most powerful industrialized nations in the world—the U.S., Canada, Germany, Japan, Italy, England, and France.

⁵BBC News, September 23, 2000. G7 Ready for Further Euro Action. Online: http://news.bbc.co.uk/1/hi/english/business/newsid_936000/936917.stm.

⁶In reality, investors are also concerned with risk and look to risk-adjusted returns.

proximately equal to interest rates available elsewhere, thereby reducing foreign reserves in the process. Similarly, if the monetary authority in a fixed exchange rate country tries to reduce the money supply to curb inflationary pressures, it cannot do so. A higher interest rate attracts foreign investors, who purchase domestic currency, driving the money supply back up and the interest rate back down.

What about fiscal policy? What happens if the fixed rate country instead decides to pursue an expansionary fiscal policy to stimulate economic growth and increase employment? Theoretically, an expansionary fiscal policy (i.e., the government spends more than it taxes) will increase domestic interest rates. If high capital mobility means that global interest rates must be approximately equal, an increase in interest rates will lead to an increase in foreign investment in domestic assets. This will increase the surplus in the capital accounts, thus leading to a balance of payments surplus. Moreover, the central bank will be forced to sell domestic currency, increasing the money supply, and driving interest rates back down, further stimulating the economy. By purchasing international currency, the central bank accumulates foreign reserves. It seems then that fiscal policy is particularly effective in the presence of fixed exchange rates and open markets.

Unfortunately, global investors may punish countries that run fiscal deficits. Such countries may be perceived as pursuing unsound macroeconomic policies, increasing the risk of economic instability. Investors dislike risk, and may withdraw money in spite of the higher interest rates. “Risky” countries generally must offer very high interest rates to attract foreign currencies. The higher interest rates, of course, dramatically increase the debt burden on the government, forcing even greater fiscal deficits and subsequent instability (increasing the risk premium once again, in a vicious cycle), or else dramatic cuts in other government programs, which are of course likely to have recessionary impacts. In addition, if the fixed rate country has previously borrowed money from the IMF to maintain its balance of payments, conditions on the loan probably forbid the borrower from increasing its fiscal deficit.

Flexible Rate Regimes

Continuing with our assumption of perfect capital mobility and hence a common international interest rate, what is the impact of a perfectly flexible exchange rate regime on fiscal and monetary policy levers? To understand this scenario, we must make explicit an intuitively obvious point. Under open market regimes, demand for domestically produced goods is determined by domestic demand plus foreign demand for domestic goods. When the domestic currency is highly valued relative to other currencies, foreign goods and services will be cheap relative to domestic ones.

Thus, a strong national currency (a high exchange rate relative to other currencies) increases the demand for imports and reduces the demand for exports, thus reducing national income, and a weak currency decreases the demand for imports and increases the demand for exports, increasing national income. When a currency appreciates or is revalued, it will be similar to a contractionary fiscal policy (i.e., a decrease in government spending, or a shift to the left of the IS curve), and when a currency depreciates or is devalued, it will be similar to an expansionary fiscal policy (i.e., an increase in government spending, or a shift to the right of the IS curve).

Now let's take a look at monetary policy in a country with perfectly flexible exchange rates that wants to stimulate the economy by increasing the money supply and lowering interest rates. Investors then want to transfer domestic currency out of the domestic economy to seek higher returns elsewhere, and they must buy foreign currency to do so. The increased demand for foreign currency drives up its price relative to the domestic currency, which then depreciates. Depreciation of the domestic currency makes domestic production cheaper relative to imports and increases global demand for domestic production. The net result should be greater national income at the same interest rate. It appears, then, that under a flexible exchange rate regime, monetary policy should be highly effective. We should note, however, that global demand has really not changed. Depreciation of the domestic currency implies an appreciation of foreign currencies and reduced income in those countries.

Fiscal policy under a flexible exchange rate regime is a different story, however. If a country in recession attempts to stimulate its economy through government expenditure or tax cuts, this will drive up interest rates. As soon as interest rates increase above the international level, foreign investment in domestic assets will increase. Flexible exchange rates immediately lead to appreciation of the currency, driving up the price of exports and reducing the price of imports. People stop consuming domestic products and purchase imports instead. Firms reduce investment, lay off workers, and/or reduce wages to compensate, lowering the increased interest rates and counteracting the effects of increased government expenditure. In this simplified world, fiscal policy is completely ineffective in the long run.

■ ECONOMIC STABILITY

We now turn to the question of economic stability. Much of the above discussion implicitly assumed perfectly informed, rational actors, who would respond to changes in investment opportunities (such as interest rates) instantaneously and appropriately. Yet global financial capital markets are

characterized by extreme volatility. We frequently see sudden, unpredicted changes in international capital flows in and out of countries that have dramatic impacts on real economic variables in those countries. These impacts tend to spread from country to country and can even lead to global economic crises. Here are some examples:

1. The Latin American debt crisis, which started in 1982 when Mexico found itself unable to make payments on its debt, and quickly spread to 39 other countries.
2. The Tequila crisis, which began in Mexico in December 1994 and required emergency loans of \$52 billion from the U.S. Treasury and the IMF.
3. The Asian flu, starting in Thailand in 1997, quickly spreading throughout South East Asia, then on to Africa, Russia, Poland, and Argentina.

All of these crises had certain elements in common. They caught the global markets by surprise, lowered economic output in the affected countries without any change in physical productive capacity, spread from country to country, and had triggers beyond the control of the affected countries. Such crises have occurred at regular intervals for literally hundreds of years. While this means that crises are not solely a consequence

Box 19-1**PETRO-DOLLARS AND THE LATIN AMERICAN DEBT CRISIS**

One of the causes of the Latin American debt crisis was almost certainly the recycling of “petro-dollars,” the high profits from the 1973 and 1979 oil price increases invested in banks by the OPEC countries. These petro-dollars were loaned at very low interest rates to LDCs. Unfortunately, the interest rates were floating; that is, they moved up and down with changes in the global interest rate.

In 1981, the U.S. Federal Reserve Bank implemented a tight monetary policy in the U.S. to curb inflation induced by increased oil prices. Simultaneously, the Reagan government engaged in record levels of deficit spending. Both actions drove up interest rates on the dollar, from about 3% in the early 1970s to more than 16% a decade later. At the same time, these high interest rates in the U.S. increased demand for dollars, driving up the value of the dollar relative to other currencies by 11% in 1981 and 17% in 1982, further increasing the dollar denominated debt burden.^a Not surprisingly, the debtor countries had considerable difficulty repaying their loans under these terms.

^aFederal Deposit Insurance Corporation, *History of the Eighties—Lessons for the Future*, 2001. Online: <http://www.fdic.gov/bank/historical/history/index.html>.

of the current push for globalization, it is likely that the increase in the speed and quantity of global financial transactions can dramatically increase their frequency and impacts. Unfortunately, no one understands the dynamics sufficiently to predict the next occurrence.⁷ Not only does such volatility reduce the ability of national policy makers to manage the economy, but it can have seriously negative impacts on human well-being.

Financial Crises

Why do such financial crises occur? There are a number of theories. While it is beyond the scope of this text to present them in great detail, we'll mention them briefly. Some crisis may occur as a result of speculative bubbles. Some asset is increasing in value, which attracts investors. The increased demand drives the price of the asset up further, bringing in yet more investors. Investors may know the asset is overvalued, but in any period there is a probability the price will increase and a probability it will crash, so many rational investors risk continued investment. Most economists consider the eventual crash a "market correction."

Other crises result from **moral hazard**. Investors believe they will be bailed out by someone (e.g., by the IMF, as happens frequently) and hence make risky investments. It is a case of a gamble in which the investor keeps the winnings if it succeeds, yet risks little loss if it fails, so investment in the gamble is far higher than it should be.

Yet other crises occur when the "fundamentals" are wrong. Countries allow their exchange rates to become overvalued, run current account deficits that are too high, or print too much money. Speculators see these signs and bet money that corrective action will be taken. Yet again, many economists view the resulting crash as a "market correction"—though many also recognize that the market may take the "correction" too far.

None of these theories adequately explains what happened in Mexico in 1994 and in Asia in 1997, or to a lesser extent in Brazil in 1997–1998.⁸ In all of these countries, the evidence points to a self-fulfilling panic. In such a panic, some investors begin to fear that a government will be unable to service its debt because of rising interest rates, falling exchange rates, and/or economic recession. If a government is unable to service its debt, investors will lose on holdings of dollar denominated assets (which are all essentially a form of debt). If the country is unable to maintain its

⁷D. M. Roodman, *Still Waiting for the Jubilee: Pragmatic Solutions for the Third World Debt Crisis*, World Watch paper 155, Washington, DC: World Watch, 2001.

⁸Brazil reacted quickly to prevent the spread of the flu, but raised the question if the cure was worse than the disease. The country spent some \$35 billion in foreign reserves to defend its currency, increased interest rates to over 40%, and introduced budget cuts and tax increases. Economic growth slowed and unemployment increased, but a nearly 70% depreciation of the Brazilian currency since December 1998 was not prevented.

exchange rate, foreign investors will lose on assets denominated in local currency. As some investors withdraw, others become more jittery.

A chain reaction occurs. Capital flees the country *en masse*. With fixed or managed exchange rates, capital flight forces governments to buy local currency and sell dollars, depleting foreign reserves, depriving them of the resources required to pay the foreign debt and maintain exchange rates. While the initial decision to flee was irrational before it was taken, it now becomes a rational decision for any remaining investors to join the panic. Investment in the affected country is now very risky, and national bonds are rated as “junk bonds.” Desperate to attain capital, governments are forced to offer higher interest rates to attract the dollars needed to meet short-term obligations, increasing the likelihood that they will be unable to repay this new debt. With higher interest rates and no foreign capital available, local businesses collapse, and the domestic economy spirals downward. Now governments lose the tax revenue necessary to meet debt obligations. Governments are forced to turn to a floating exchange rate, which is almost inevitably accompanied by massive devaluations. Everything the speculators feared comes to pass, but only because investors acted on their fears.

Self-fulfilling panics are basically instances of multiple equilibriums. If speculators withdraw their capital, the rational thing for others is to withdraw capital also, leading to one equilibrium. If speculators leave their capital in place or invest more, then this also becomes the rational act for others, leading to a different equilibrium.⁹ The point is that the decision to remain or withdraw is based on imperfect knowledge.

Compounding the problem, when one country devalues its currency, it becomes more competitive in the export market against similar countries. Speculators are then justified in believing that those similar countries will lose access to the resources required to meet their debt obligations. As more economies succumb, the demand for exports decreases. Financial crises clearly generate negative externalities affecting other countries, and economic stability is thus a public good. The instabilities associated with large foreign debt is one reason we like Keynes' rejected idea of the International Clearing Union (discussed in Chapter 17) that penalized both surplus and deficit countries, giving all countries an incentive to avoid large balance of payments deficits that cumulate into foreign debt.

Economic stability is a public good and an important objective of economic policy. If having a larger economy in the distant future is the only goal, then instability does not matter a great deal as long as the long-run

⁹J. A. Sachs, A. Tornell, and A. Velasco, The Mexican Peso Crisis: Sudden Death or Death Foretold? *Journal of International Economics* 41: 265–283 (1996); P. Krugman, Are Currency Crises Self-Fulfilling?, *NBER Macroeconomics Annual*, 345–378 (1996).

Box 19-2 SELLING SHORT

Another factor contributing to financial instability is when speculators sell the local currency short. To sell short, a speculator essentially borrows the currency from someone else and sells it at the going price, betting that the price will fall (i.e., the government will be forced to devalue) and the local currency loan can be paid back at a lower dollar cost.^a Simplifying greatly, selling short increases the supply of domestic currency, putting downward pressure on the price. Governments are forced to sell dollars to cover these short positions. If the banks lack the resources, they cannot cover the positions and are forced to devalue. If enough people sell short, devaluation is inevitable, and the speculators profit. If someone highly respected for their financial acumen makes a big speculative investment to sell short another currency, other speculators will take notice and put their money alongside his. Currency speculators can often outspend national governments, forcing even developed countries to devalue their currencies, as happened to England in 1992. This type of herd behavior can turn selling short into a self-fulfilling prophecy.

The perversity of this type of profit from speculation is that the speculators increase their entitlements to real goods and services by controlling more financial resources, yet the production of real goods and services actually declines. When greater wealth goes to those who produce, we call it earned income. What should we call it when greater wealth goes to those who destroy productive capacity?

^aFor example, you borrow pesos and use them to buy dollars. You then hold the dollars waiting for a peso devaluation. Then you use the dollars to buy pesos. You get a lot more pesos than you originally borrowed, thanks to the devaluation. You pay back your loan and have a lot of pesos left over!

trend is toward growth. But a richer future cannot be the only goal for LDCs. When a country's income falls, that means the incomes of people within that country fall. For many people, no income means no access to the basic necessities of life. Governments will find it difficult to defend an economic policy that lets people starve this year by claiming that those starving people will be able to eat three times as much next year. Indeed, it is a fairly well established fact that people react differently to gains than losses, even when they are of the same magnitude. This, in fact, is a basic implication of the theory of diminishing marginal utility. Yet too many economists still seem to miss this point.

THINK ABOUT IT!

Can you explain why the law of diminishing marginal utility suggests we should give more weight to the loss of something than to a gain of equal magnitude?

Box 19-3 IS RISK AVERSION IRRATIONAL?

It is telling that advocates of globalization complain that “U.S. households show an irrational preference for ‘safe’ bonds, bank deposits, and money market funds over ‘risky’ equities, especially where their retirement savings are concerned.”^a The same advocates would have us believe that a desire for stability and security for countries is as irrational as for retirees, as they go on to assert that

We are about to enter into an era of explosive growth and change as a result [of globalization], but the journey to prosperity will not be entirely smooth. The capital market will force governments to cut unsustainable entitlement policies including pensions, subsidized healthcare or welfare, and force businesses dependent on government protection to restructure or perish. There are likely to be massive job losses and great social and personal dislocations as these changes take place. The turmoil will be the greatest in Europe and in the developing countries. . . .^b

In spite of assertions to the contrary, it is not irrational to favor a small but secure income over a gamble between starvation and fabulous wealth, even if the risk of starvation is small. Nor is it evident that the benefits of explosive growth will compensate for massive job losses, social and personal dislocations, and general turmoil—to say nothing of potential ecological impacts.

^aL. L. Bryan and D. Farrel, *Buy Stocks, Shun Bonds (Opportunities in the Global Capital Market)*, The McKinsey Quarterly 167 (1996).

^b*Op. cit.*

Along with the public good/externality problem, two other market failures in financial markets deserve mentioning. Moral hazard, as briefly described above, is widely recognized as a market failure. Yet financial crises are contagious not just between countries, but also between financial institutions. When some banks started to fail during the great depression, this triggered panic that caused other banks to fail. When a financial institution is “too large to fail,” governments feel obliged to bail them out, so moral hazard is difficult to avoid if large financial institutions are unregulated. Yet regulated financial sectors have fewer investment opportunities than unregulated sectors, and are likely to be less profitable and attract fewer investors. After all, if the added safety of regulation compensated for the diminished profits, then the financial sector would regulate itself, yet this is clearly not the case. Thus, countries may pursue a “race to the bottom” in terms of financial regulation.

Another serious market failure is the information asymmetry that always exists between lenders and borrowers, which is aggravated when capital crosses international boundaries. Lenders price loans according to

the risks involved, but the borrowers generally have a much clearer understanding of the risks than the lenders. This can lead to adverse selection. If lenders raise interest rates to compensate for unknown risks, then the only borrowers will be those engaged in the riskiest activities.

What has been the global policy response? In both Asian and Mexican crises, the IMF response was to impose tight fiscal and monetary policies on the affected countries. The IMF saw this as imposing discipline on those economies; however, the effect of such policies is to plunge them even further into recession. The IMF is also pursuing further deregulation of global financial movements. In an official communiqué of the IMF Interim Committee issued in April 1998, the IMF declared that it will amend its articles to “[make] the liberalization of capital movements one of the purposes of the Fund and [extend], as needed, the Fund’s jurisdiction for this purpose.”¹⁰ This in spite of the fact that Michel Camdessus, then president of the IMF, predicted that “a number of developing countries may come under speculative attacks after opening their capital account” as a result of their unsound macroeconomic policies.¹¹ This, of course, ignores the possibility of self-fulfilling panics and contradicts the original IMF charter to protect the stability of national economies.

Why does the IMF stubbornly pursue such policies, in the face of so much evidence that they are wrong? IMF economists are almost all trained in neoclassical economics. They believe well-informed people make rational choices based on rational expectations. As a result, they believe market prices are rational and lead markets to general rational equilibrium. Where effective markets don’t exist, they must be created in order to achieve these harmonious results. The IMF routinely bets billions of other people’s dollars that this is the case. In contrast, George Soros, a renowned international financier, routinely bets billions of his own dollars that this is not the case. In his own words, he believes that “people act on the basis of imperfect understanding and equilibrium is beyond reach,” and as a result, “market prices are always wrong, in the sense that they present a biased view of the future.”¹² Soros frequently wins his bets, while the IMF record speaks for itself. Finally, Soros further claims that

¹⁰International Monetary Fund, Communiqué of the Interim Committee of the Board of Governors of the International Monetary Fund, Press Release No. 98/14, Washington, D.C., April 16, 1998. Cited in M. Chossudovsky, *Financial Warfare*. Online: <http://www.corpwatch.org/trac/globalization/financial/warfare.html>.

¹¹Communiqué of the IMF Interim Committee, Hong Kong, September 21, 1997. Cited in *ibid*.

¹²Quoted in W. Greider, *One World, Ready or Not: The Manic Logic of Global Capitalism*, New York: Simon & Schuster, 1997, p. 242.

“extending the market mechanism to all domains has the potential of destroying society,”¹³ a belief with which we strongly concur.

BIG IDEAS to remember

-
- | | |
|---|---|
| ■ Balance of payments (BOP) | ■ Fiscal policy under different exchange rate regimes |
| ■ Current account | ■ Petro-dollars and debt |
| ■ Capital account | ■ Financial crises |
| ■ Fixed vs. flexible exchange rates | ■ Speculation |
| ■ Monetary policy under different exchange rate regimes | ■ Economic stability as a public good |
| | ■ Moral hazard |
-

¹³See Society Under Threat—Soros, *The Guardian*, London, October 31, 1997. Cited in Chosudovsky, op. cit.



Conclusions to Part V

Globalization has been proposed as a panacea for economic problems, because it is expected to bring economic growth and wealth to all. This conclusion is based on flawed assumptions. With free international capital flows, globalization favors absolute advantage and not comparative advantage. While globalization may bring greater economic growth, the growth will not benefit all countries, and the costs of growth could outweigh the benefits at the margin. Empirical evidence shows that the benefits of growth go to the already wealthy, while the costs, in terms of lost ecosystem services, are borne by all—and perhaps by future generations most of all. Globalization also weakens the policy levers governments have to address issues of scale and distribution on their own, and unregulated flows of financial capital can create serious economic instability. While trade can be desirable, it is more likely to be so if it takes place between independent nations with the ability to make their own policies, and these nations recognize that economic growth brings costs as well as benefits.

We now turn our attention to some of the policy tools such independent nations could utilize to attain the goals of sustainable scale, just distribution, and efficient allocation.

PART VI

Policy





General Policy Design Principles

Listening to many economists and even policy makers, one might often get the impression that the only role of government is to create conditions that allow the market to function. The argument is that not only are markets the best institution available for allocating scarce resources to attain our desired goals, but they are also the best way to determine what our goals are. After all, if the mere act of buying and selling reveals people's preferences, the job of both economist and policy maker is simply to allow the market to satisfy those preferences. But we know that this is not the case. Markets by definition can only reveal preferences for market goods, yet many of the goods and services that enhance human welfare are nonmarket goods. Thus, not only do markets fail to reveal preferences for these resources, they also fail to allocate them effectively.¹ Markets also fail to address the issues of scale and distribution. Therefore, the first point we must make in this chapter is that the market cannot tell us how much clean air, clean water, healthy wetland, or healthy forest we should have, or what level of risk is acceptable when the welfare of future generations is at stake. Nor can it tell us what is a desirable initial distribution of resource ownership.

¹Also, it is not strictly true that markets reveal preferences even for market goods; they reveal choices, which are, to be sure, an expression of preferences, but a very conditioned expression under the constraint of existing prices and incomes.

■ THE SIX DESIGN PRINCIPLES

We have already seen the even more fundamental point in Chapter 3 that, at a minimum, policy requires two philosophical presuppositions: first, that there are real alternatives (nondeterminism); second, that some states of the world really are better than others (nonnihilism). We now examine six general design principles for policies, followed by a consideration of their proper sequencing and of how and where policy interventions should first impinge on the market—that is, on quantity or price—at the input or output end of the throughput? We end the chapter with further reflections on property rights. Then, in Chapters 21–23, within the guidelines of the design principles, we will offer some particular policies for promoting a steady-state economy—one that is sustainable, just, and efficient.

1. Economic policy always has more than one goal, and each independent policy goal requires an independent policy instrument.

If there is only one goal, the problem is technical, not economic. For example, building the most powerful engine possible is purely a technical problem. But building the most powerful engine that is not too heavy to power an airplane involves two objectives, power and lightness, that have to be optimized in terms of a higher goal—making an airplane fly. This is already an economic problem of optimizing the combination of conflicting goals, even though we encounter it at an engineering level.

In ecological economic policy, we have three basic goals: sustainable scale, just distribution, and efficient allocation. Nobel laureate Dutch economist Jan Tinbergen set forth the principle that *for every independent policy goal we must have an independent policy instrument*.² You have to be very lucky to hit two birds with one stone—it nearly always takes two stones to hit two birds flying independently. For example, should we tax energy and raise its price for the sake of inducing more efficient use, or should we subsidize energy and lower its price to help the poor? This question is endlessly and uselessly debated. One instrument (price of energy) cannot serve two independent goals (increase efficiency, reduce poverty). We need a second instrument, say an income policy. Then we can tax energy for the sake of efficiency, and distribute income (perhaps from the tax proceeds) to the poor for the sake of alleviating poverty. With two policy instruments, we can serve both efficiency and equity. With only one instrument, we are forced to choose either efficiency or equity.

Since ecological economics insists on three basic goals, it is already

²J. Tinbergen, *The Theory of Economic Policy*, Amsterdam: North Holland Publishing, 1952.

clear that we will need three basic policy instruments. The three goals are “independent” in the sense that attaining one will not bring about attainment of the others. Of course the goals are not “isolated” in the sense of having nothing to do with each other—they are, after all, parts of a single economic system. Now that we know how many instruments, what other principles will help answer the question: What kind of instruments? The remaining principles attempt to give guidance in this regard.

2. Policies should strive to attain the necessary degree of macro-control with the minimum sacrifice of micro-level freedom and variability.

Consider this example. If what is limited is the capacity of the atmosphere to absorb CO₂, it is important to limit the total CO₂ emissions. Average per-capita emissions times population will have to equal the limited total. But it is not necessary that each and every person emit exactly the per-capita average. There is room for micro-variation around the average in the light of particular conditions, as long as the total is fixed.

Let's take another example. Population stability requires the average of 2.1 children per couple. But it is not necessary (or even possible in this case) for each family to have the required average number of children corresponding to generational replacement. Macro-control is compatible with varying degrees of micro-variability around the average. In general we should opt for the least micro-restrictive way of attaining the macro-goal. Markets are useful in providing micro-variability, but by themselves they do not provide macro-control.

3. Policies should leave a margin of error when dealing with the biophysical environment.

Since we are dealing often with staying within biophysical limits, and since those limits are subject to much uncertainty and at times irreversibility, we should leave a considerable safety margin, or slack, between our demands on the system and our best estimate of its capacity. If we go right up to capacity, we cannot afford mistakes because they are too costly. The inability to tolerate mistakes, or sabotage, exacts a large price in reduced individual freedom and civil liberties.

Security issues surrounding the nuclear fuel cycle and the safeguarding of plutonium have already given us a foretaste of the problems involved in living too close to the edge. Our historical experience with small life-support systems operating close to capacity—namely, spaceships, or even ordinary ships or submarines—has thus far not permitted democracy. Military levels of order and discipline are required on fragile vessels operating near carrying capacity. Only our large spaceship Earth, with lots of

slack, can be sufficiently forgiving of error to tolerate democracy. There are political as well as economic costs to excessive scale, a fact that has received too little attention.

4. Policies must recognize that we always start from historically given initial conditions.

Even though our goal may be far from the present state of the world, the latter remains our starting point. We never start from a blank slate. Present institutions must be reshaped and transformed, not abolished. This imposes a certain gradualism. Even though gradualism is often a euphemism for doing nothing, it is nevertheless a principle that must be respected.

What are our present institutions? Basically, the market system and private property, but also public property and government regulation. The World Bank and the IMF may be with us for a while, even though they are not nearly as basic an institution as private property or the market. We have neither the wisdom nor the time to start over again without our most fundamental institutions, even if we could imagine alternatives. The considerable stretching and bending of these institutions that we will recommend will be thought radical by some, so it is important to emphasize the conservative principle of starting from where we are, even if the basic idea is not to remain there.

5. Policies must be able to adapt to changed conditions.

Change is an ever-present reality. Human impacts on the ecosystem are enormous, and are likely to cause new problems over time. Ecosystems themselves naturally show considerable variation over time—where time can be measured in seasons, years, or eons. Human knowledge is increasing, leading to a new awareness of previously unrecognized problems, as well as new solutions to old ones. The economic system is also continually evolving, and policies that work well now may not work as the system changes.

In addition, we may find that some policies that seem ideal in theory may not be ideal when implemented, and they may even have seriously negative unforeseen side effects. As we apply policies, we will learn how they work in the real world, and thus learn how to improve them. The process of developing and implementing policy solutions must respond to this feedback, and real-life outcomes must carry far more weight than stylized theories. **Adaptive management**—changing our policies as conditions change and as we learn more—must be a guiding principle. Indeed, we believe that ecological economics itself is an example of adaptive management to the problems arising from the transition from an empty to a full planet.

6. The domain of the policy-making unit must be congruent with the domain of the causes and effects of the problem with which the policy deals.

This is often called the **principle of subsidiarity**. The idea is to deal with problems at the smallest domain in which they can be solved; problems should be addressed by institutions on the same scale as the problem. Don't seek global solutions for local problems, and don't try to solve global problems with purely local measures.

Consider the example of garbage collection. Garbage collection is largely a municipal problem. Aggregating all municipal garbage collections into a "global garbage problem" is not helpful. Deal with it at the municipal level, at least in the first instance. If local garbage has to be disposed of farther and farther away, or if it contaminates air or water and is thus transported far away, then it becomes a correspondingly larger problem—county, state, region, and so on. By contrast, global warming is fundamentally a global problem, because emissions anywhere affect the climate everywhere. Here we really do need global policy.

■ WHICH POLICY COMES FIRST?

In ecological economics we have three basic goals, so we need three basic policy instruments. The goal of efficient allocation requires the instrument of the market, at least for goods that are private (excludable and rival). For public goods the market will not work. The goal of sustainable scale requires a social or collective limit on aggregate throughput to keep it within the absorptive and regenerative capacities of the ecosystem. The goal of distributive fairness requires some socially limited range of inequality imposed on the market. As we have seen, the market cannot achieve distributive equity or sustainable scale. Furthermore, the market cannot even attain allocative efficiency unless the distribution and scale questions have already been answered. So now we know that in the sequencing of policy instruments the market comes third, after its preconditions have been established.

But what about the sequencing of scale and distribution? Here it is reasonable to put scale first, because limiting scale usually means that previously free natural resources and services have to be declared scarce economic goods.

THINK ABOUT IT!

Why is this so? Why would limiting scale make previously "free" resources scarce economic goods?

Once they are scarce, they become valuable assets, and the question of who owns them must be answered. That is an issue of distribution. The

logic is that of the “cap and trade” policies. For example, the total scale of SO_2 emitted into an airshed is capped at a scale deemed sustainable. The right to emit SO_2 is no longer a free good. Who owns that right? Previous users? All citizens equally? The state collectively? Some answer to the distributive question must be given before trading in a market can solve the allocative problem. People cannot trade what is not theirs. Some goods that were outside the market can be made marketable goods (i.e., excludable). By limiting the scale of resource use and distributing the ownership of the resource, we can convert nonmarket into marketable goods. But as we saw earlier, not all goods can be converted into market goods. Many good things are inherently nonrival or nonexcludable. We will return to this issue.

The set of prices that corresponds to a Pareto optimal allocation will be different if we set the cap differently, or if we distribute ownership differently. This means that we cannot set the cap or distribution according to computations of their social costs and benefits based on existing prices. To do so would be to engage in circular reasoning because the prices depend on the scale or distribution. The ideal scale or distribution, calculated on the basis of existing prices, would, if attained, result in a different set of prices that would invalidate the original calculation. Thus, we cannot set the scale according to the criterion of efficient allocation, nor can we set distributive limits by the criterion of efficient allocation.

What, then, is the criterion for scale? Sustainability is the criterion for scale.

And what is the criterion for distribution? Justice is the criterion for distribution.

These, obviously, are not matters of market economics; rather, they are biophysical and cultural. They must be socially and politically determined, and thus, as a matter of policy, these decisions can be made more or less simultaneously. In strict logic, scale comes before distribution, because if there were no cap on total use the resource would be a free good, and the distribution of ownership of a free good would make no sense. Given these prior social decisions on scale and distribution, the market will determine allocatively efficient prices. Indirectly these prices will reflect the scale and distributive limits and therefore may be thought of as “internalizing” the values of sustainability and justice that have been previously decided politically, independently of prices.

Economists sometimes argue that scale is not an independent consideration—that if we had perfect information and could internalize all external costs and benefits into prices, then the market would automatically stop growth at the optimal scale. In other words, scale would have been subsumed under allocation. This has a certain plausibility if we accept the assumption of “perfect” information. However, if in the name

of perfect internalization we insist that prices should incorporate the costs and benefits of different scales, we would also have to insist that prices reflect the costs and benefits of different distributions. But if we tried to use prices based on a given distribution as the means of measuring the costs and benefits of a change in distribution, we are again being circular.

Economics has clearly recognized the circularity and insisted that just distribution is one thing, efficient allocation another. Economists would not, for example, appeal to perfect information and advocate raising the price of things poor people sell or lowering the price of things poor people buy in order to internalize the external cost of poverty into prices. Instead they might advise us to redistribute income directly to attain a more just distribution, and let prices adjust. This also makes sense for questions of scale.

The way to get prices to reflect the values of just distribution and sustainable scale is to impose quantitative restrictions on the market that limit the degree of inequality in distribution of income and wealth to a just range, and that limit the scale of physical throughput from and back to nature to a sustainable volume. These imposed macro-level distribution and scale limits reflect the social values of justice and sustainability, which are not personal tastes and cannot be reflected in the market by individualistic actions. The market then recalculates allocative prices that are consistent with the *imposed* scale and distribution constraints, thereby in a sense “internalizing” these social values into prices.

Since it is circular to use prices to calculate optimal scale and optimal distribution, we need some metric of benefit and cost other than price (exchange value). As already suggested, this metric is the value of justice in the case of distribution; it is ecological sustainability, including intergenerational justice, in the case of scale. These are collective values, not individual marginal utilities per dollar equated among different goods in order to maximize satisfaction of individual tastes. If we reduce all dimensions of value to the level of subjective personal taste, then we cannot capture or bring to bear on the market the real weight of objective social values, such as distributive justice and ecological sustainability.

■ CONTROLLING THROUGHPUT

If we are going to impose macro-level constraints on the market to control scale, we must ask: At which end of the throughput flow should we impose these constraints? We could impose restrictions at the output end (pollution), as in the SO_2 example. Or we could limit the input flow from nature (depletion). Since there are fewer mines and wells than there are tailpipes, smokestacks, garbage dumps, and outflow pipes into rivers, lakes, and oceans, it would be easier to put the control on depletion than

on pollution. By the law of conservation of matter-energy, if we limit the inflow we will also automatically limit the outflow. Even if it is the outflow that is causing the immediate problem, even if sinks are more limiting than sources, the outflow may be more easily controlled by limiting the flow at its narrowest point, the inflow.

As a general rule it makes sense to control depletion directly, thereby indirectly controlling pollution. As with all general rules, there are exceptions. Although depletion limits provide a quantitative limit on pollution in a gross sense, many qualities of pollution could result from the same quantity of depletion. Depending on how resources are used, the same inputs could be converted into very toxic or very benign pollutants. Therefore, we cannot entirely concentrate on inflows and expect the outflows to take care of themselves. But inputs (depletion) should be our first control point.

■ PRICE VS. QUANTITY AS THE POLICY VARIABLE

In Chapter 21, we will look in detail at two basic approaches to limiting throughput: raising prices through taxation to reduce demand, or limiting quantity directly through quotas and letting prices adjust. Before we examine the specifics of each approach in detail, we investigate the effectiveness of the two approaches.

Given that we should mainly intervene on the input side, how should we do it? Should we try to control quantity and let the market determine the resulting price, or try to control price and let the market determine the resulting quantity?

■ THINK ABOUT IT!

Trying to control both price and quantity would be a bad idea. Can you explain why? Think of a demand curve.

If we can, by taxes, set the price where we want it, that will, via the demand curve, determine a corresponding quantity. Alternatively, if by quotas we set the quantity where we want it, that will, via the demand curve, determine a corresponding price. Theoretically, given a demand curve, we could get the same result either by fixing price or by fixing quantity (see Figure 20.1). How do we choose between them?

One point to bear in mind is that demand curves are uncertain and shift. So if we set the price, errors and omissions will result in quantity changes. If we set the quantity, errors and omissions work themselves out in terms of price fluctuations. Therefore, an important criterion by which to choose is: Where is it least painful to experience errors, as price changes or as quantity changes?

Because the ecosystem cares about quantities extracted and absorbed

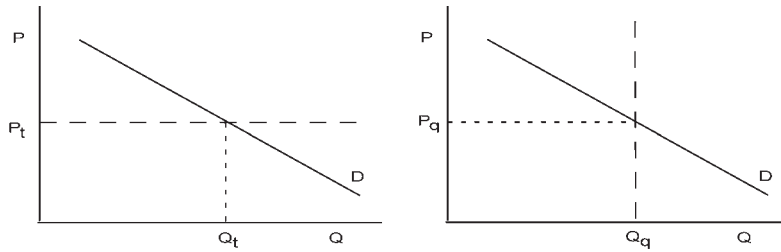


Figure 20.1 • Setting quantity with taxes and quotas. If taxes are set so that the price is equal to P_t , consumers will demand quantity Q_t . Alternatively, quotas could limit output to quantity Q_q , and for this quantity consumers are willing to pay P_q . It is possible, of course, to set Q_q so that $P_t = P_q$, or set P_t so that $Q_t = Q_q$.

and not the prices people pay, ecological economists have a preference for fixing quantity and bearing the adjustment cost of errors in terms of price fluctuations. That is ecologically safer, and more in accord with our design principle of leaving a big safety margin. It also follows more strictly the sequencing logic just discussed of setting the scale first outside the market and then allowing prices to be determined by the market.

The superiority of quotas for achieving desirable scale is even more evident when both global population and per-capita levels of resource use continue to grow. Such growth implies increasing demand for both sources and sinks. In the presence of quotas, the increased demand is reflected entirely by increased prices, creating ever-greater incentives to use resources more efficiently. If resource consumption and/or waste emissions are limited by use of taxes, greater demand will lead to both higher prices and greater consumption, unless taxes are continually raised. In a world of ecological thresholds and irreversible outcomes, taxes lead to a greater risk that we will eventually use up our slack.

We must also recognize that markets do not work as perfectly as we economists might like. In Chapter 19, we discussed financial panics and the tendency for people to follow herd behavior. In Indonesia, the Asian flu crisis led to an 85% depreciation of the currency over the space of weeks. Indonesia has a significant percentage of the planet's rainforests, which supply an abundance of important, nonmarket ecosystem services. Imagine Indonesia sought to internalize the value of these ecosystem services by taxing deforestation at the rate of 10 rupiah per board foot of timber extracted. Most of Indonesia's timber is sold on the global market, so the 85% depreciation in its currency would have implied an 85% decrease in the tax in relation to the trade currency. Even if the tax had been in U.S. dollars, the currency of international trade, most other costs of harvest are in rupiah. The 85% drop in all other costs would still reduce the price and

increase demand, leading to greater deforestation. As we know from Chapter 6, greater deforestation can reduce the sustainable yield of the forest, and possibly even drive forest stocks below the point of critical depensation. A quota would not be subject to such irrational fluctuations in economic variables.

■ SOURCE VS. SINK

As we have seen, before the market can operate, ownership of the newly scarce asset must be distributed. This raises many difficult issues, and it may cause a retreat from some of our earlier design principles in order to respect the last one—namely, that we start from historically given initial conditions. At the input end, most resources are already owned. At the output end, the atmosphere is not privately owned. A **source** is that part of the environment that supplies usable raw materials that constitute the throughput by which the economy produces and that ultimately returns as waste to environmental sinks. A **sink** is that part of the environment that receives the waste flow of the throughput and may, if not overwhelmed, be able to regenerate the waste through biogeochemical cycles back to usable sources. Sources are generally owned, and sinks are generally not owned. Directly controlling sources (depletion) involves more interference with existing property rights than controlling sink access. To socialize all resource ownership after resources are privately owned is revolutionary. To socialize the unowned sink, the atmosphere, and then charge a dumping fee seems less threatening to private property than a direct control over the amount extracted. But to control emissions is to dam the river at its widest point, contrary to the principle that it is easier to dam it at its narrowest point. Should we advocate revolution then? Most of us would not, but let's try to stick to our principle for a while at least.

What might be done to reconcile the principle of intervening at the depletion and with the difficulty that sources are private property? We might recognize that property is a “bundle of rights.” The resource owner has to give up one stick out of that bundle—namely, the right to decide independently the rate of extraction of his resource. He still owns the resource and receives payment for whatever amount he extracts. But the scale of extraction is no longer a free good. It is socially limited to a national quota, and resource owners must bid at auction for the right to extract a share of the limited total extraction permitted. The scale limit may ultimately be set according to sink limits if they are more binding, but still enforced at the source end.

Alternatively, we could have a market in sink permits, capped at an aggregate scale. Suppose, in the case of fossil fuels, all users had to purchase emissions permits to burn whatever fossil fuel they purchased. This would

indirectly limit demand at the source, and source owners would feel the pinch of scarce sink capacity—the scarcity of one complementary factor reduces the value of the other. This may appear less an infringement on the property rights of the source owners, since they have no claim to the sink, and it is the sink that is being directly limited. But the sink limit on the throughput is surely translated back to the input end. And since that sink limit will be experienced by source owners indirectly, even if we put the limit directly on the sink, why not go ahead and put the limit on the source in the first place? That would be the more efficient place to put it, even if it is the sink that is most scarce.

The other possibility noted is to fix prices through taxes and allow the market to set the corresponding quantity. Once again this is more efficiently done at the depletion end rather than the pollution end, but both are possible. The tax may be levied at the input end even though its basic motive is to limit the output. The advantage of taxes is administrative simplicity—we already have a tax system, and altering it is less disruptive than setting up a quota system with auctions. This is a significant advantage. On the other hand, taxes really do not limit quantities very strictly, and they maintain the false perception that there are no quantitative limits as long as one pays the price. As long as we pay the price plus a corrective tax, the message conveyed is that we can get as much as we want, individually and collectively. The quota, by contrast, makes it clear that the total quantity will not increase, and that all the price is doing is to ration the fixed quantity among competing users. The latter seems a more honest and truthful perception, since we are dealing with a scale-limited physical throughput, not income or welfare.

■ POLICY AND PROPERTY RIGHTS

Before we begin to examine specific policies for achieving a more sustainable, just, and efficient world, we must discuss one of the core features of any policy: property rights. Concern over scale is concern over sustainability, and what is sustainability but the right to resources for future generations? If we believe there is a need for improved distribution, we are basically questioning the existing endowments of property rights. Finally, markets cannot efficiently allocate nonexcludable resources, and excludability is nothing more than a property right. Policy is largely concerned with creating, redefining, and redistributing property rights.

Property rights and excludability are not inherent properties of goods or services. No good is excludable, and no one has property rights unless a social institution exists that makes it excludable and assigns property rights (though we also know that it is not possible to make all goods excludable). A property right for one individual simultaneously imposes a

duty or obligation on other individuals to respect those rights. For example, if person A has the right to breathe clean air, then person B has the corresponding duty not to pollute that air. The state ensures that person B will fulfill her duty. Property rights are therefore a three-way relationship between one individual, other individuals, and the state.³

In the absence of property rights we have privilege, or presumptive rights. If one person has privilege, he is entitled to behave as he pleases, and others have no rights. If a factory owner has privilege with respect to the atmosphere, he can pollute the air as much as he pleases. If others suffer from this pollution, then they must seek to change the prevailing lack of property rights.

When human populations and impacts were small relative to the sustaining ecosystem, the use of natural capital was appropriately characterized by privilege. Why not allow industries or individuals to pollute if few people lived nearby to be affected by that pollution? Why not allow industries or individuals to harvest fish or harvest trees if they existed in abundance? Why not give away rights to minerals to those who discovered them, as long as seemingly limitless virgin lands remain for future exploration? It makes little sense to establish property rights to super-abundant resources.

However, as we know, the world is no longer so empty. The privilege to extract and pollute now imposes cost on others. This creates pressure to develop environmental policies that assign or modify property rights. Those who have privilege to extract or pollute are likely to defend the status quo, claiming that privilege as a right, when in reality it is an absence of defined rights. As we pointed out in Chapter 10, many economists have argued that it does not matter to whom rights are assigned; as long as rights are assigned, the market can efficiently allocate resources. We maintain, in contrast, that while the distribution of rights may not matter in terms of Pareto efficiency (i.e., Pareto efficient outcomes are possible for any distribution of property rights, though it will be a different outcome for different distributions), it matters profoundly for equity. We take the position that property rights belong to the people, as represented by the state, until otherwise assigned, and their distribution should be decided by a democratic process that respects future generations.

There are three important types of property rights, or entitlement rules, and rights to a specific piece of property may be affected by any combination of these.

1. An entitlement known as a **property rule** holds if one person is free to interfere with another, or free to prevent interference. For exam-

³D. Bromley, *Environment and Economy: Property Rights and Public Policy*, Oxford, England: Blackwell, 1991.

ple, an individual may own a piece of land. If he has the right to build a landfill that destroys the neighbors view, or to prevent the neighbor from walking across the land, he is free to “interfere” with the neighbor. Nor can the neighbor interfere with the landowner’s landfill operation. If the neighbor wants to walk across the property or prevent the landfill from being built, the landowner’s consent is required.

2. An entitlement known as a **liability rule** holds one person is free to interfere with another or prevent interference, but must pay compensation. For example, the landowner might be free to build the landfill, but by law he is then forced to compensate his neighbor for the smell, loss of view, and other disamenities. At the same time, the state could call on its right of eminent domain to take away the land from the landowner to build a highway and pay compensation at fair market price.
3. An entitlement, known as an **inalienability rule**, holds if a person is entitled to either the presence or absence of something, then no one is allowed to take away that right for any reason. There may be certain types of chemicals or products that are absolutely not permitted in the landfill, regardless of compensation. The negative impacts of these products are so severe that present and future generations have an inalienable right not to be exposed to them. Dioxins and radioactive waste would fall into this category.

Finally, we must remember that property rights need not be private property rights. Property rights can belong to individuals, communities, the state, the global community, or no one. While many conventional economists favor private property rights, we have already learned that private property rights are not possible in all circumstances (e.g., the ozone layer). In addition, many cultures have successfully managed common property resources for millennia. Almost all nations have certain resources owned by the state; recently, international agreements, such as the Montreal and Kyoto protocols, have recognized the need for ownership and management of some resources by the global community. The search for suitable policies neither can nor should be limited only to those that require private property rights.

Now that we have discussed basic principles of policy, appropriate policy sequence, high-leverage points of intervention, and the relationship between property rights and policies, we turn our attention in the next three chapters to some specific policies. We will more or less follow the policy sequence suggested above: scale, distribution, then allocation.

However, while in most cases we cannot hit two birds with one stone, some of the policies we look at are really a bundle of policies affecting all

three goals. Other policies may hit one goal squarely while having an impact on another policy goal as well. Our division of the discussion is therefore not exclusive: All three goals will be discussed in each chapter, and policies are grouped only according to their dominant impact. Our three independent goals require three independent policy instruments in the same sense that solving three simultaneous equations for three different variables requires three independent equations; that is, one equation must not be derivable from the other two, and one variable cannot be the same as another, just expressed differently. Three simultaneous equations in three unknowns form a system, so clearly all three variables are interrelated—they are not independent in the sense that a change in one has no effect on the others (i.e., isolated) because they are clearly parts of an interdependent system. But they are independent in the sense in which each independent variable in a set of simultaneous equations requires an independent equation if the system is to be solvable.

BIG IDEAS to remember

-
- | | |
|---|--|
| ■ Six policy design principles | ■ Price vs. quantity as control instrument |
| ■ Proper sequence of policies | ■ Circularity of internalizing scale or distribution in prices |
| ■ Source vs. sink as throughput control point | ■ Property rights |
-



Sustainable Scale

Environmental policies are inherently related to scale. In an empty world, environmental goods and services are not scarce resources, and hence not the focus of policies. The issue is whether environmental policies address scale directly or only tangentially. Just as importantly, to be effective, the policies should square with the six design principles outlined in Chapter 20. We will discuss four different types of policy that affect scale: direct regulation, Pigouvian taxes, Pigouvian subsidies, and tradeable permits. We examine how each is applied in the real world.

■ DIRECT REGULATION

The dominant form of environmental policy affecting scale in most of the world is the regulatory instrument, which can take a variety of forms. Sometimes an activity or substance is considered to have unacceptable costs and is simply banned. For example, many countries no longer allow lead additives to gasoline or the production of DDT, and currently negotiations are under way for banning the production of 12 different persistent organic pollutants (POPs) at a global level. When a substance is sufficiently dangerous, such bans are appropriate.

In other instances, regulation will limit the quantity of a pollutant that can be produced, and set emissions levels for the firms or individuals responsible for producing it. For example, individual paper factories may have legal limits to the amount of waste they can discharge into a river, and in many countries, vehicles have to pass emission tests. In yet other instances, regulations will force all firms or individuals to use the best available control technology (BACT) to limit pollution. BACTs may be imposed on all firms or individuals, or only on new entrants to an industry. BACT regulations play an important role in the U.S. clean air laws.

For fisheries, a common regulation has been to limit the fishing season or regulate the type of equipment that may be used in order to reduce the annual catch.¹ Failure to comply with regulations generally involves fines or other penalties. These policies are therefore known as **command-and-control regulations**.

What are the advantages and disadvantages of such policies? Most of them limit the amount of pollution/resource harvest to an acceptable level, thus contributing to the goal of desirable scale. With renewable resources, regulations may be the best way to address biological requirements. Examples are banning harvests during mating seasons, imposing minimum mesh size on fishing nets, forbidding harvest of gravid females, leaving the best and largest of a species as seed stock, or banning certain harvest methods that are particularly destructive of habitat. Regulations can be applied to everyone equally, or tailored to meet alternative distributional goals. Finally, policy makers are generally familiar with this approach. It is reasonably easy to understand and can be fairly cheap to monitor and enforce—for example, it is very easy to check whether a given firm is using a mandated technology.

The disadvantage is that in general, regulations fail to meet the criteria for allocative efficiency and thus are often not the most cost-effective way to reach a desired goal. Moreover, they fail to provide incentives for surpassing a goal, bringing pollution below the regulated level, for example. These points deserve elaboration.

As shown in Chapter 7, the basic requirement for economic efficiency is that marginal costs equal marginal benefits, at both the individual level and the social level. Ideally, environmental policies should achieve this goal. In practice, however, for our pollution example this would require that we know the marginal social costs of pollution, the marginal net benefits of activities that pollute, and the marginal abatement costs of pollution. Of course, there are really no benefits to pollution *per se*, but all production causes pollution, and we could not exist with no production at all. In reality, it is virtually impossible to know all the marginal costs of pollution, and very difficult for policy makers to know marginal abatement costs. Perfect allocative efficiency is therefore something of a pipe dream.

While we cannot hope for a perfectly efficient solution, we can hope for a cost-effective one. A cost-effective solution achieves a given goal at the lowest cost, even if marginal costs do not exactly equal marginal benefits. It is, therefore, a very desirable goal, but one that is unlikely to be attained by simple regulations. The reason is that command-and-control

¹Regulating inputs to reduce catch is an entirely different issue than regulating inputs to reduce negative externalities, such as excessive by-catch or habitat destruction.

regulations ignore the second general design principle described in Chapter 20, that policies should sacrifice the minimum of micro-freedom to attain macro-control.

This point is perhaps most readily illustrated with a specific example. Imagine there are three firms polluting a waterway upstream of the drinking water intake valve for a city. A regulatory agency determines that for health reasons, pollution loads should be cut by 40% and demands that each firm cut its emissions correspondingly. The problem is that different firms may have different marginal abatement costs (MACs) and/or different operating costs due to a variety of factors, such as manufacturing process or age of manufacturing equipment. It may be very expensive for one firm to cut its emissions by 40%, and very cheap for another firm to do so.

Table 21.1 shows hypothetical abatement costs for three firms, each discharging equal amounts of pollution. A 40% reduction per firm would cost a total of \$360,000. The most cost-effective way for society to reduce pollution would be to assess for each additional unit of pollution to be eliminated which firm has the lowest MAC and have that firm make the reduction. This would cost a total of \$210,000. If we assume increasing MACs, this rule would lead to identical MACs for all firms undertaking abatement.

THINK ABOUT IT!

Why are MACs identical for all firms under this rule? You may want to review the material on the equimarginal principle of maximization in Chapter 8 to figure this out. The principle is the same.

Unfortunately, the regulators do not really know the firms' abatement costs. While each firm presumably knows its own abatement costs, gathering this information would be costly, and if the goal were to force reductions in this manner, each firm would have an incentive to misinform the regulator.² Also, it would hardly seem fair to make some firms reduce pollution far more than others.

Another problem that arises with regulation is that once regulatory goals have been achieved, there is no incentive to reduce pollution any further and few incentives for new pollution reduction technologies. Similarly, if regulations apply to specific areas (such as U.S. clean air laws), there is no incentive not to increase pollution in areas below the

²When the EPA first proposed tradeable emission permits in SO₂, industry estimates of MACs were as high as \$10,000 per ton. EPA estimates were in the neighborhood of \$1000 per ton. Permits currently trade for under \$100 per ton. Carol Browner, speech, "Public Health and Environmental Protection in the 21st Century," University of Vermont's 2002 Environmental Literacy Seminar Series. March 25, 2002.

■ **Table 21.1**

MARGINAL POLLUTION ABATEMENT COSTS FOR THREE FIRMS EMITTING 5000 TONS OF POLLUTION EACH			
MAC	Firm 1	Firm 2	Firm 3
First 20%	\$50,000	\$10,000	\$25,000
2 nd 20%	\$200,000	\$25,000	\$50,000
3 rd 20%	\$400,000	\$50,000	\$100,000
4 th 20%	\$800,000	\$100,000	\$175,000
Final 20%	\$1600,000	\$500,000	\$200,000

maximum allowed level. Yet we have already seen that pollution has marginal external costs even at very low levels (see Figure 12.7).

What we seek, then, are policies that take advantage of the equimarginal principle of maximization by equalizing MACs across firms, provide incentives to develop new technologies for reducing environmental costs, and keep costs low by allowing firms to act on their private knowledge of their own abatement costs. The ideal policy would also set the marginal benefits of production equal to the marginal environmental costs it imposes, but as we stated earlier, environmental costs are largely unknown.

We will now examine three policies that can theoretically achieve these goals: taxes, subsidies, and tradeable permits.

■ **PIGOUVIAN TAXES**

Early in the last century, economist A. C. Pigou began grappling with the problem of internalizing environmental externalities. As we discussed in Chapter 10, externalities occur when one economic agent causes an unintended loss or gain to another agent, and no compensation occurs. In the case of a negative externality, the basic problem is that the economic agent is able to ignore a cost of production (or consumption). Under such circumstances, the market equilibrium of marginal costs equal to marginal benefits does not emerge, and some of the wonderful benefits of markets fail to appear. Pigou came upon the simple solution of imposing a tax equal to the marginal external cost. This would force the economic agent to account for all economic costs, creating an equilibrium in which marginal social costs were equal to marginal social benefits.³

³A. C. Pigou, *The Economics of Welfare*, 4th ed., London: Macmillan, 1932 (originally published in 1920).

Note that this policy requires a change in property rights. When a firm is free to pollute, it has privilege, and those who suffer from the pollution have no rights. A Pigouvian tax essentially creates a property right to the environment for the state, using a liability rule. Firms can still pollute, but they must now pay for the damages from their pollution.

Of course, as we cannot accurately measure marginal environmental costs, the **Pigouvian tax** cannot be set precisely at that level. Even if we did know marginal environmental costs, these costs change with the amount of pollution, and the ideal tax would presumably also have to change. While Pigouvian taxes will not lead to perfectly efficient outcomes, they will reduce environmental costs, and do so cost-effectively. How do they accomplish this?

When abatement costs are less than the tax, it is cheaper for the firm to abate, so that's what they will do. On the other hand, when abatement costs are more than the tax, paying the tax minimizes costs and maximizes profits. This means that after implementation of the tax, the MAC for all firms will be equal to the tax. Firms for which it is cheap to reduce pollution will therefore make large reductions, and firms for which it is expensive will reduce much less. The latter firms will, of course, pay correspondingly more in taxes than the former. Note that no one but the firm needs to know the firm's marginal abatement costs. Each firm acting on its own preferences and own knowledge with a maximum of micro-freedom generates the cost-effective outcome desirable to society.

THINK ABOUT IT!

Using the figures in Table 21.1, can you figure out what range of taxes would lead to a 40% reduction in emissions?

Firms continue to pay tax on every unit of pollution that they produce. This means that there is always an incentive for achieving further reductions in pollution and doing so more cost-effectively. Such incentives are perhaps the most important reason that taxes are superior to command-and-control regulations.

In addition to taxes, firms must also pay abatement costs. It is therefore quite possible that total costs to the firm and industry (all related firms together) under the tax will be higher than they would have been under command-and-control regulations (e.g., forcing each firm to cut emissions by 40%). However, relative to society, the tax is a transfer payment and does not count as a cost. And by ensuring that the firms with the lowest abatement costs make the largest reductions in pollution, the tax ensures that actual costs to society are less under the tax than they would have been under regulations. Nor can the tax really be considered unfair to the firm, as it is simply a payment for the costs the firm is imposing on society. It is possible that the tax would even drive some firms

out of business, but as long as the tax were no greater than the marginal external environmental cost, it would simply mean that the costs the firm was imposing on society were greater than the benefits it was providing.

THINK ABOUT IT!

Can you figure out what the total costs to each firm would be under command and control regulations vs. a Pigouvian tax of \$50,001/ton?

It would be very difficult to predict the decrease in negative externalities that would result from any given tax, and a trial-and-error process might be required. Yet changing taxes every year, or even every few years, creates a burden for firms, who lose the ability to plan for the future. Perhaps the best approach would be to begin with a fairly low tax, but let firms know that the tax will increase over time. This approach would let firms gradually change their practices, reducing overall cost by allowing new technologies to come online before the tax reaches its ultimately desired level.

As long as human populations and the economy are still growing, the demand for activities that impose environmental costs is also likely to grow. This means that to maintain the desired level of environmental amenities or resource depletion, the tax would need to increase over time. As with all environmental policies, the principle of adaptive management is appropriate.

■ PIGOUVIAN SUBSIDIES

A **subsidy** is a bonus or payment for doing something, the opposite of a tax. A **Pigouvian subsidy** is a payment to each firm for each unit by which it reduces environmental costs; it has many of the same attributes as the tax. Ideally, the subsidy will equal the marginal benefit to society of abating pollution. As long as abatement costs are lower than the subsidy, the firm will reduce pollution. Again this will equalize MACs across the industry, the precondition for a cost-effective outcome. While a tax follows the “polluter pays principle,” a subsidy, in contrast, basically assumes that the polluter has the privilege to pollute, and society must pay him not to.

THINK ABOUT IT!

Can you figure out what would be the total cost to society of a \$50,001/ton subsidy not to pollute (assuming no new entrants to the industry), and the total benefits to each firm net of abatement costs?

One serious problem with subsidies is that they can perversely lead to an increase in pollution. A subsidy increases the profit margin for the polluting industry, possibly attracting new entrants. While each firm pollutes

less than in the absence of the subsidy, more firms could still lead to more total pollution. While many people might justifiably resent the notion of paying people not to impose costs on the rest of society, and the potential outcome of greater pollution is entirely undesirable, this does not mean that Pigouvian subsidies are entirely irrelevant. Pigouvian subsidies can be desirable as an incentive to ecosystem restoration. For example, paying farmers to reforest their riparian zones might reduce nutrient runoff and provide a host of other ecological services. In addition, under international law, sovereign nations have the right to do as they choose with their resources, and there is no global government that could impose a Pigouvian tax on the negative environmental costs of deforestation, for example. Under such circumstances, something like a Pigouvian subsidy may be the best option. We will return to this issue in some detail later.

One final point bears mentioning. While Pigouvian taxes or subsidies may lead to the welfare-maximizing outcome of marginal social costs equal to marginal social benefits, the same does not hold true at the level of the individual. This is a result of the fact that many environmental costs are public bads. Every individual suffers from the same amount of environmental cost, yet each individual has different preferences concerning those costs. A perfect market solution would have to distribute the tax among the afflicted population to exactly compensate for the marginal damage they suffer from the environmental cost. Of course, it would be impossible to determine the marginal cost curve for every individual on the planet, and individuals would have incentives to misinform the agency collecting this information if it would influence how much of the tax they were to receive. Also, if individuals were compensated for the externalities they suffer, they might do less to avoid externalities, and this too could reduce efficiency.⁴

■ TRADEABLE PERMITS

Tradeable permits are another cost-effective mechanism for achieving a specific goal. Rather than increasing prices through a tax to reduce demand, tradeable permits require society to set a **quota**, a maximum amount of pollution or resource depletion that it will allow. This approach is currently used in the United States to regulate SO₂ emissions, and in several countries to regulate fisheries.

What factors should determine the allowable quota? From the economist's perspective, the ideal quota should be set so that the marginal

⁴E. T. Verhoef, "Externalities." In J. C. J. M. van den Bergh, *Handbook of Environmental and Resource Economics*, Northampton, MA: Edward Elgar, 1999. Again we reiterate that "efficiency" should not be the sole criterion for decision making in any case.

benefit from one more unit of pollution or harvest is exactly equal to its marginal social and private cost. Uncertainty and ignorance probably make this ideal unattainable. Even if we could accurately estimate marginal costs at the existing scale, as we learned in Chapter 19, the very act of setting the quotas (scale) changes the prices used in calculating the costs and benefits.

In the case of renewable resource harvests, quotas should recognize that renewable resource stocks can provide a flow of harvests but are simultaneously funds that provide a stream of services over time. The quota should recognize this dual nature of renewable resources. Where quotas are currently used in fisheries, managers focus almost entirely on the stock-flow aspect of the resource, virtually ignoring the fund-service aspect. While we are admittedly far more ignorant concerning the exact role of ecological fund-services, we do know their value is not zero, and they should not be ignored. At the very least, quotas should be determined with substantial input from ecologists to ensure that they provide sufficient slack with respect to ecological thresholds (remember general design principle #3 in Chapter 20). The quota process should also respect the principle of adaptive management, allowing adjustment as new information becomes available.

THINK ABOUT IT!

If fisheries managers decide to take an ecological economic approach to establishing harvest quotas and explicitly incorporate the fund-service benefits of fish stocks into their decisions, what impact will this have on the quota? Draw a graph showing the sustainable yield curve and total private cost curve for a fishery, and label the following points: (1) the open access equilibrium and (2) the annual profit-maximizing quota. Relative to points 1 and 2, in what region of the graph would you find the quota that maximizes net present value? In what region would you find the quota that maximizes benefits from both the stock-flow and fund-service nature of the resource? In determining optimal quotas, do you think policy makers should discount benefits to future generations? Why or why not?

Once established, the quota is then distributed among polluters and resource users in the form of permits or individual quotas. Permits may be auctioned off or distributed for free, and the distribution process can be designed to achieve other social goals, such as greater income equality. Permits can be issued annually, or once and for all. They can be for set quantities or for a proportion of an adjustable quota. Many variations are possible. A permanent permit for annual emissions or annual resource depletion provides firms with the most certainty for future plans, but annu-

ally adjustable permits may be necessary in the presence of imperfect information, natural variation, and ecosystem change.

Once firms own the permits, they are free to buy and sell them, as is the case with any other market good. If permits are for pollution, the firm will abate pollution as long as that is cheaper than the cost of a permit, and will use permits when abatement costs are more expensive. Again this leads to equimarginal abatement costs, a precondition for cost-effective outcomes, and again allows maximum micro-freedom. If permits are for resource depletion, the firms with the most profitable use of the resource will be able to pay the highest price for the permits. This theoretically ensures that the resource will be allocated toward the most desirable ends, but, we reiterate, only if we believe individual “votes” concerning the desirable ends should be weighted by individual wealth.

THINK ABOUT IT!

Again using Table 21.1, if each firm is given quotas to emit 6000 tons of pollution, who will actually end up emitting the pollution, what is the possible price range for the quotas/ton, and what is the range of final net costs of pollution abatement to each firm?

Quotas also require a change in property rights, but whereas taxes impose a liability rule, quotas impose a property rule. The owner of the quota essentially owns the waste absorption capacity (a rival good made excludable by quotas) of the medium into which they are emitting wastes. The right may initially be awarded to the government, members of the community affected by the pollution, or to the polluters. The same principle is true when quotas are used to end privilege in an open access regime.

One issue with quotas is that there may be little incentive to reduce total pollution or resource extraction below the quota. If the quota is carefully chosen, this need not be a problem. Within a tradeable quota system, any profit-maximizing firm still has the incentive to reduce emissions or resource harvest so that it may sell a portion of its quota. Thus, while quotas will not drive undesirable activities below the quota level, they provide incentives for reaching quotas ever more cost-effectively. Also, if the economy or population is growing, quotas ensure that resource use will not grow.

Many economists have pointed out that if environmentalists think a quota is too high, they are free to purchase shares of the quota and discard them. Unfortunately, we again run into the problem of public good provision here. The environmentalist would incur the entire cost of purchasing the permit but share the benefits with everyone. In addition, if permits are issued annually in variable amounts, then the government could potentially issue more permits in response to those being purchased

and not used. Alternatively, if the regulatory authority decides too many permanent permits were issued to begin with, or new information changes the assessment of how many permits are desirable, the government can readily purchase some back and not use them, as we illustrate with the following case study.

Tradeable Permits vs. Shorter Seasons

As we have previously discussed, oceanic fisheries have been heavily overfished, and policies are urgently needed to address this problem. A number have been tried, providing good evidence for the superiority of solutions that maximize micro-freedom. Within this context, we will compare efforts in the United States to reduce unsustainable fish harvests by shortening the season with efforts in New Zealand to implement quotas and tradeable permits.

The halibut fishery is one of the oldest on the Northwest coast of the Americas, and by 1960, open access conditions had led it to be fished almost to extinction. In 1960, the International Pacific Halibut Commission (IPHC) was created to regulate the annual harvest and restore the catch to the maximum sustainable yield. Harvests were limited by imposing a season, which was then gradually reduced as required. This method proved highly effective at restoring the population and increasing the annual harvest. However, by the early 1990s, the season was as short as one or two 24-hour periods per year (depending on how long it took to reach the annual quota established by the IPHC), during which fishermen engaged in a mad race to maximize their share of the catch.

What are the implications for efficiency and cost-effectiveness of such a short season? First, fishing is already one of the most dangerous industries, and engaging in a mad race just makes it that much more dangerous, especially if the “season” happened to coincide with bad weather. Loss of life was frequent. Boats often captured so much fish they were in danger of sinking, and sometimes did. Fishermen were forced to cast as many lines as possible, ensuring that some would be lost. The situation was made worse when the large number of boats caused lines from different boats to get tangled and cut, perhaps leaving already hooked halibut to die. The 2-day open access fishery led fishermen to invest in more equipment to take more fish in a shorter period. In spite of increasing stocks, the season was continually shortened, and the equipment (and labor force) then went unused out of season. Almost all halibut fishermen also take other fish with the same equipment, but the net result was still excess capacity. Demands for rapid harvest led to poorer treatment of the fish, and a lower quality product.

Once landed, all of the fish arrived at the market at the same time. There was therefore a very limited market for fresh halibut, and most had

to be frozen. Again, large capital investments were required to create the infrastructure for freezing all the halibut in such a short period, and there was excess capacity for the remainder of the year. Processing fish by freezing is capital-intensive, and it therefore increased the barriers to market entry, threatening to limit competition. Processing fresh fish, in contrast, is labor-intensive and has far lower capital costs.

In 1990, when Canada modified its system by establishing individual quotas for ships and extending the season, it created a market for fresh halibut. In fact, because Canadian halibut was mostly sold fresh, Canadian fishermen enjoyed a 70% price premium over their Alaskan counterparts. The failures of the U.S. system were so pronounced that in 1995, the U.S. instituted an individual fishing quota system as well. The quotas were assigned to currently active fishermen based on their recent harvests. They were intended to allow fishermen to extend their harvest efforts over the entire season, thereby paying more attention to quality than to speed. Some leasing and trading of the quotas were allowed, but with strict limits on concentration of shares.⁵

New Zealand's fisheries went through the cycles typical of most countries. At first the resource was scarcely exploited, with the exception of in-shore fisheries. In the 1970s, seeking to exploit a new source of foreign exchange, the government began a program of subsidies to develop the industry. The result was overcapitalization (basically too many boats chasing too few fish) and dramatic declines in fish populations. In 1982, the government forced fishermen who earned less than 80% of their income from fishing out of the market. This had a highly negative impact on Maori fishermen, who traditionally earned their living from a variety of activities, but it did little to reduce pressure on the fishery. Then, in 1986, New Zealand decided to follow the economists' advice and implement a system of transferable fish quotas. Similar systems are also in place in Iceland and the Philippines.

The process is simple. Scientists determine the total allowable catch (TAC) for each species from each of several geographic areas, typically with the goal of achieving maximum sustainable yield. From this number they subtract the expected take by the sport fishery and set aside 20% for the Maori. (An 1840 treaty awarded rights to all New Zealand fisheries to the Maori, but New Zealand chooses not to honor this treaty.) The remainder is the **total allowable commercial catch (TACC)**, which is divided up into **individual transferable quotas (ITQs)**, which may be

⁵K. Casey, C. Dewees, et al. The Effects of Individual Vessel Quotas in the British Columbia Halibut Fishery, *Marine Resource Economics* 10(3): 211–230 (1995); C. Pautzke and C. Oliver, Development of the Individual Fishing Quota Program for Sablefish and Halibut Longline Fisheries off Alaska. Anchorage, Alaska: North Pacific Management Council, 1997. Online: <http://www.fakr.noaa.gov/npfmc/Reports/ifqpaper.htm>.

bought, sold, or leased on the market. The initial ITQs in terms of tons of fish were awarded to fishermen in proportion to their catch history. To make the ITQs more attractive to fishermen, initial awards were close to historic catches and exceeded the TACC. The government then purchased back sufficient ITQs to reach the TACC. Fish populations fluctuate naturally, and so did the TACC. Initially, the government was forced to buy or sell ITQs whenever the TACC changed. Then, in 1990, ITQs were changed to represent a proportion of TACC.

In terms of scale and allocation, the policy has been very effective. Fish populations have recovered, though there have been problems with the introduction of new fish species into the market, because typically little is known about their life cycles. For example, considerable evidence suggests that orange roughy has been overexploited in spite of the TAC. In terms of efficiency, fishermen now need only invest enough to capture their share, lowering their capital costs. Harvests are spread out over a longer period, increasing the market for fresh fish. Fishermen can purchase ITQs for different species when they have a large by-catch⁶ and sell or lease ITQs when they fail to meet their quota. Less efficient fishermen can sell their quotas to more efficient ones. The value of fisheries in New Zealand has apparently doubled in recent years.

The impact on distribution, however, is far less desirable. ITQs tend to concentrate in the hands of the larger firms, leading to concentration of the wealth in a lucrative industry. Maoris—in spite of the treaty awarding them rights to all the fisheries—have disproportionately been forced out of the market. Part of the problem lies in access to credit. ITQs do not count as collateral for bank loans. When the TAC decreases, small-scale fishermen lack collateral for bank loans to purchase more ITQs, while large firms have other assets they can use as collateral. In part, this problem stems from the initial allocation of ITQs based on catch histories. Those firms that played the largest role in overexploiting the fisheries initially were rewarded with more ITQs.⁷

The case of New Zealand fisheries shows the importance of separate policies for achieving separate goals. The TACC (one policy instrument) set the scale, and the ITQs (a separate policy instrument) achieved effi-

⁶By-catch is the harvest of species different from the target species. Depending on the species and the existing laws, by-catch may be kept or thrown back. By-catch is often killed in the harvest but is nonetheless thrown back. Dolphins as by-catch for some types of tuna fisheries, and sea turtles as by-catch for some types of shrimp fisheries, have received considerable attention. For some fisheries such as shrimp, by-catch may be more than ten times the mass of the target species.

⁷P. Memon and R. Cullen, *Fishery Policies and their Impact on the New Zealand Maori*, *Marine Resource Economics* VII(3): 153–167 (1992); New Zealand Minister of Fisheries, *The Quota Management System*, no date. Online: <http://www.fish.govt.nz/commercial/quotams.html>. R. Bate, “The Common Fisheries Policy: A Sinking Ship” *Wall Street Journal*, June 2000. Online: http://www.environmentprobe.org/enviroprobe/evpress/0700_wsjs.html.

cient allocation. But the policies did not address distribution issues, which turned out to be problematic. They require a third instrument, perhaps one that could limit the concentration of ITQs to help maintain market competitiveness and avoid forcing poorer fishermen out of the market.

■ POLICY IN PRACTICE

We see, then, that policies are available that meet environmental goals cost-effectively and that provide incentives for reducing pollution, resource depletion, and so on, even after those goals have been met. Most of these policies are widely accepted by economists as cost-effective solutions, yet regulatory agencies in general seem to prefer the less efficient command-and-control regulations. Why is this so? There are a number of reasons.

Environmental regulations often are administratively simple and may have low monitoring costs. Regulatory agencies have substantial experience with these options, and institutions can be slow to change. Conceptually, regulations are simple and widely perceived as fair, at least when they affect everyone equally. Many regulators pay little attention to cost, and may be more concerned with reducing their own transactions costs than with lowering the costs to polluters. Abundant other reasons also exist, but considerable evidence suggests that in many circumstances, the overall costs to society of reaching a given target are higher under regulation than under mechanisms that allow a maximum of micro-level freedom by relying on market allocation, subject to macro-level control.

In the United States, the cap and trade systems have had some success on a limited basis (e.g., SO_2), while in Europe the tax scheme, referred to as “ecological tax reform,” has been more popular. The idea is sold politically under the banner of “revenue neutrality”—the government taxes the same total amount from the public, just in a different way. Following the design principle of gradualism, European governments have sought to impose the most desirable resource tax first, and to couple it with the worst existing tax, eliminating the latter to the extent that revenue from the former permits. Thus, one may get a “double dividend”—the environmental benefit of taxing a resource whose price is too low, plus the fiscal benefit of getting rid of a distortionary or regressive tax.⁸ Subsequently, one seeks to couple the next most desirable resource tax with the next worse other tax, and so on.

The slogan of ecological tax reform is: “Tax bads, not goods.” The idea is to shift the tax burden from value added by labor and capital

⁸The existence of a “double-dividend” is a source of frequent dispute among environmental economists but is more accepted by ecological economists.

(something we want more of) to “that to which value is added”—namely, the throughput and its associated depletion and pollution (something we want less of). It seems a matter of common sense to tax what you want less of and stop taxing what you want more of. Ever suspicious of common sense, however, neoclassical economists have invented general equilibrium models with particular assumptions (such as the familiar production functions with no resource inputs) that lead to counterintuitive results. We find these models in general to be artificial and unconvincing. In any case the policy, at an incipient level, seems to be working in Europe.⁹ The main political dilemma European governments face in trying to implement ecological tax reform seems to be maintaining competitive advantage in international trade by keeping resource prices low, versus internalizing external costs in prices and thereby raising them to the detriment of competitive advantage—a problem we encountered in our look at globalization.

This latter problem is a severe and general policy difficulty. Our fifth general policy design principle stated that the domain of the policy-making authority should coincide with the domain of actions open to those who cause, or are affected by, the policy (see Chapter 20). If a policy is enacted to limit pollution and a firm can avoid compliance simply by moving across a boundary, then the extent of domains does not coincide. Globalization, as we saw in Chapter 18, expands the domain of actions to the entire world, while keeping the domain of public policy confined to the national level. Because national policies are easily evaded in such a situation, we have a general weakening of public policy along with an increase in the relative power of private individuals and corporations. Public efforts at the national level to deal with poverty, environmental degradation, public health, education, and even macroeconomic goals of full employment without inflation are all automatically sacrificed to the overriding goal of growth in the global production of market goods, as stimulated by free trade and free capital mobility.

This is why people are demonstrating in the streets of Seattle, Prague, Genoa, Washington, D.C., and anywhere else the WTO, the IMF, and the World Bank meet. Shortening the length of the meetings and changing the venue to places like Qatar do not address the issues raised by critics. Is it too much to hope that the concepts of ecological economics can provide a framework in which the legitimate claims of both growth and limits can be recognized?

⁹B. Bosquet, *Environmental Tax Reform: Does It Work? A Survey of the Empirical Evidence*, *Ecological Economics* 34(1): 19–32 (2000).

BIG IDEAS to remember

- | | |
|-----------------------------------|---|
| ■ Direct regulation | ■ Abatement costs |
| ■ Command-and-control regulations | ■ Total allowable commercial catch (TACC) |
| ■ Pigouvian taxes and subsidies | ■ Individual transferrable quotas (ITQs) |
| ■ Tradeable permits (quotas) | ■ Ecological tax reform |
-



Just Distribution

The distribution of wealth and income is always a contentious issue. But it is also crucially important. Why?

First, people who are too poor will not care about sustainability. Why should they worry about the welfare of the future when they are not even able to provide for their own basic needs? Throughout the world, the excessively poor are forced to mine soils, clear-cut forests, overgraze grasslands, and tolerate excessive pollution just to survive. And as we have seen, the impacts of these activities are not merely local; they have global consequences.

Second, people who are excessively rich consume large amounts of finite resources, possibly depriving future generations of the basic means of survival. Even the economists most reluctant to make interpersonal comparisons cannot deny that the marginal utility of consumption for those below subsistence is far higher than for those buying increasingly frivolous luxury goods.

Third, if we care about sustainability, we care about intergenerational distribution. We do not want to force the future to live in poverty simply so we can consume more luxuries. Yet what ethical system can justify a concern for the well-being of those yet to be born, while not caring for the well-being of those alive today?

Finally, we know that the economic system cannot grow forever on a finite planet. We must limit growth to ensure the well-being of the future, but one cannot ethically tell poor people they must continue to suffer deprivation to ensure that the future does not suffer. If the pie must cease to grow, then we are ethically obliged to redistribute it.

If distribution is so important, then why is it so contentious? Many people believe that in a free market society, people have wealth because they have earned it, and it is unjust to take from people what they have

earned with the sweat of their brows. We agree that in general distribution policies should not take away from people what they have earned through their own efforts and abilities. However, people should not be able to capture for themselves values created by nature, by society, or by the work of others. And they should pay a fair price for what they receive from others, including the services provided by government, and for the costs they impose on others. In addition, we must recognize that a less unequal distribution of resources may generate public goods such as economic stability, lower crime rates, stronger communities, and better health (as discussed in Chapter 15), and society should pay for public goods. If we follow these principles, the resulting distribution should be both just and sustainable.¹

Distribution must focus on both income and wealth, and on market goods as well as nonmarket goods. Policies that provide more money for government from higher-income and wealthier individuals can further improve distribution by allowing governments to cut taxes for the less well-off, or by funding public goods projects that benefit everyone. Policy makers have devised many plans to achieve distributional goals, both within and between nations. Some have proven successful, some not. We now review some policies designed to achieve a more just distribution.

■ CAPS ON INCOME AND WEALTH

Must we set a maximum individual income? At first glance, many people consider this type of policy an unwarranted intrusion on individual liberties. What right does the state have to take what someone has earned with the sweat of her brow? Income and wealth are the “just deserts” of hard work. From this viewpoint, income caps are unjust.

However, on a finite planet subject to the laws of thermodynamics, if some people consume too much this generation, they will reduce the resources available to future generations. This means that in the future, society may be worse off than it is today, and people may have to work harder than the current generation to consume even less. In this case, a sense of obligation toward future generations demands that society as a whole reduce consumption so that future generations have the same opportunities to be rewarded for their work as the present, the same opportunities to receive their “just deserts.” However, to demand that society as a whole reduce consumption yet not to demand that the wealthiest members of society also do so is a difficult position to defend.

¹An unsustainable outcome would be an unacceptable cost for future generations.

Box 22-1 WEALTH AND POWER

Many wealthy people earn far more than they could conceivably consume. If Bill Gates invested all his wealth in inflation-indexed government bonds with real yields of 3%, which is probably as close as one can get to a risk-free investment, he would be earning over \$3 million per day.^a Many of the world's richest people earn more than they or their offspring could conceivably spend. Why would anyone accumulate wealth if they do not intend to consume it? The only reasonable answer is to amass power and status.

Certainly, it is difficult to argue that wealth does not bring power in existing political systems. While many people argue that the inequitable distribution of wealth is acceptable, in democratic countries, far fewer say the same about inequitable distribution of power.^b And the power that rewards the accumulation of wealth is readily used to generate yet more wealth and hence more power, in a vicious cycle. For example, it is painfully clear that corporate donations to the political parties in most countries are not made to strengthen democracy, but rather to promote legislation that provides greater economic advantage for the contributors. How else can we explain the fact that so many major corporations contribute money simultaneously to two politicians running against each other for the same office? By seeking economic advantage through political influence, wealth undermines market forces and the beneficial outcomes they are capable of generating.

Strangely enough, most Americans remain opposed to income caps.^c Americans and citizens of many other capitalist democracies seem to have two completely incompatible core beliefs: We have the right to a democratic government, and the right to become richer than Midas. However, as Supreme Court Justice Louis Brandeis reportedly said, "We can have a democratic society, or we can have the concentration of great wealth in the hands of the few. We cannot have both."

^aAt the time of this writing (November 2001), Bill Gates topped *Forbes* list of the world's richest people for 2003, with a net worth of \$40.7 billion.

^bR. Lane, Market Justice, Political Justice, *American Political Science Review* 80(2): 383–402 (1986).

^c*Ibid.*

Is there any harm in accumulating wealth simply for status? There is, for two reasons. First, people generally exhibit their status through conspicuous consumption, which increases scale. Second, status is measured relative to others' positions and is thus a zero sum game. Everyone's status in society cannot increase. Therefore, if I work hard to accumulate wealth and increase my status, I am reducing the status of others relative to me. In order to maintain their status, they will have to work harder as

well, sacrificing leisure time, time for community, and time for family. If we all worked twice as hard to increase our status, no one's status would change, we would all have less time to pursue other goals, and we would consume more natural capital. Status through wealth accumulation can turn into a kind of arms race in which we all work harder and become worse off.

Conspicuous consumption is therefore a negative externality, and people should pay for the negative impacts it imposes on others. A progressive consumption tax would help redistribute resources, and by taxing a negative externality, it would lead to a more efficient allocation of resources as well.² Empirically, in the wealthier countries, there is evidence that people are growing less satisfied with life instead of more satisfied, in spite of continuing dramatic increases in national wealth.³

THINK ABOUT IT!

Can you explain how a progressive consumption tax could even make the wealthy better off?

Policies for capping income might also include a highly progressive income tax that asymptotically approaches 100%, more direct limits on how much someone can earn, or relative limits that establish a legal ratio between the highest and lowest income allowed. Progressive income taxes are used worldwide. Many economists claim that such taxes are a deterrent to economic growth. However, economic growth in the U.S. was quite high during the 1950s, when the highest marginal federal tax bracket was 90%, compared to less than 40% today. If capping income is, in fact, growth-inhibiting, such a policy would support policies aimed at limiting the scale of the economy.

Policies for capping wealth could include a progressive wealth tax, as currently exists in a number of European nations. People already pay taxes on real estate, which is a form of wealth, so why not extend this to all wealth, particularly the forms that are highly concentrated among the wealthiest? Very high inheritance taxes would also help, as an estimated 46% of accumulated wealth is directly inherited.⁴

Many people would object that progressive taxes take a disproportionate amount from the rich, and therefore do not meet the criteria we discussed at the beginning of this chapter. However, governments generally provide most of the infrastructure and institutions that allow businesses

²R. Frank, *Luxury Fever: Why Money Fails to Satisfy in an Era of Excess*, New York: Free Press, 1999.

³R. Lane, *The Loss of Happiness in Market Democracies*, New Haven, CT: Yale University Press, 2000.

⁴G. Alperovitz, Distributing Our Technological Inheritance, *Technology Review* 97(7): 30–36 (October 1994).

to thrive and people to grow wealthy. Would Bill Gates, Warren Buffet, and other billionaires be so wealthy if they had been born in sub-Saharan Africa? In addition, political philosophers have long argued that one of the dominant roles of government is to protect private property. Clearly then, the more private property someone owns, the more they benefit from the services of government, and the more they should pay for those services.⁵

Another argument against income caps is that they are also harmful to the poor. From this viewpoint, allowing unlimited accumulation of wealth creates incentives that increase total production and employment opportunities and make the worst off better off than before. Capping income for the wealthiest reduces the opportunities for the poorest to escape poverty. If this is true, then how do we explain the productivity and relative absence of poverty in northern Europe, where taxes are very high?

■ MINIMUM INCOME

Many countries, including the U.S., have instituted policies intended to guarantee a minimum income. These policies can help achieve sustainability by ameliorating poverty, as well as by reducing the gap between a society's richest and poorest members. Moreover, minimum income policies are justified because they can help provide a number of other public goods. In Chapter 16 we showed how economic recessions can have positive-feedback loops. Something causes consumption to decline. People buy less, so firms produce less and lay off workers. Laid-off workers consume less, so firms again reduce production. In the presence of a minimum income, even when people are laid off, they will continue to consume. Indeed, those with the lowest incomes typically spend the highest percentage of those incomes on consumption. A minimum income helps break the positive-feedback loop that causes economic recessions, and a more stable economy can benefit everyone. In addition, abundant evidence links income disparity to crime, violence, and other public bads (see Chapter 15). A minimum income may not eliminate these problems, but it can help reduce them, and therefore we favor such a policy.

Neoclassical welfare economics, whose foundations are utilitarian philosophy and diminishing marginal utility, as we saw in Chapter 8, implicitly calls for the elimination of poverty. If the goal of society is to maximize utility summed over individuals, and wealth and income offer diminishing marginal utility, then clearly an additional unit of wealth for a poor

⁵We are glad to report that Warren Buffet and Bill Gates, Sr., are on record as vigorously agreeing with the proposition that society contributes the conditions in which individuals can earn great wealth and that such individuals should be willing to pay significant taxes, especially estate taxes. See W. H. Gates and C. Collins, *Wealth and Our Commonwealth: Why Americans Should Tax Accumulated Fortunes*, Boston: Beacon Press (2003).

person provides more utility than the same unit would provide for a wealthy person. Economists reluctant to accept this conclusion have asserted that different people have immeasurably different capacities to enjoy (or suffer?) and hence we cannot make interpersonal comparisons of utility. Thus, many economists have focused on maximizing production rather than utility, which effectively skirts the distribution issue.⁶

On the other hand, it is clear that, on average, a unit of additional income would benefit someone living in absolute poverty more than the same amount would benefit a millionaire. People may have different capacities for enjoyment at some level, but we are very alike in our suffering—we are all poisoned by the same toxins, made ill by the same germs—and our biological subsistence needs are the same. The additional utility when one moves from below subsistence needs to above them is obviously immense.

Curiously, most Americans profess to believe that the current distribution of income in the United States is unjust, yet they remain reluctant to provide income to those who have not “earned” it. However, the “just deserts” argument is based on the assumption that people are paid according to their contribution to society. Yet the last two centuries have seen a fairly steady upward trend in real incomes. This is not so much because people make more substantial contributions to society on their own, but because they benefit from past contributions to productivity. In other words, many people are awarded more than their just deserts already, so why not do the same for the worst off?

The specific policy approaches to ensuring a minimum income are more debatable than the need for some policy. The most commonly employed policies are:

1. Welfare programs, in which the government provides direct monetary or material aid to the poor.
2. Unemployment insurance for the unemployed.
3. Minimum wages and negative income taxes for the employed.

These approaches can play a role in ensuring minimum incomes, but such simple transfers are probably not the best approach to ending poverty for either society or the recipients of such transfers.

Among traditional approaches to a minimum income, many ecological economists would argue first for equal opportunity in education, job access, and job advancement, followed by guaranteed jobs at a living wage, and direct transfer payments playing a role only when necessary. In addition, we believe that people have equal entitlements to wealth created by nature and by society, independent of the entrepreneurial ability of the in-

⁶J. Robinson, *Economic Philosophy*, Garden City, NY: Doubleday, 1964.

dividual. Distributing this wealth equally would provide a minimum income. This involves other, less conventional approaches, which we now explore by looking at income as returns from factors of production.

■ DISTRIBUTING RETURNS FROM THE FACTORS OF PRODUCTION

In order to take a systematic look at income distribution, we recall from Chapter 15 that there are four sources of income: wages, profits, interest, and rent. Wages are the returns to labor, profits are the returns to entrepreneurship, interest is the return to capital, and rent is the return to land and other natural resources. Most efforts at distributing income focus on returns to labor, while the greatest disparities in income are actually the result of the other factors of production. We now turn our attention to distributing the returns to capital and the returns to natural capital.

Distributing the Returns to Capital

Financial capital, including equity in productive assets, is highly concentrated both within and between nations. The United States probably offers the most egregious example among the developed countries. As reported in Chapter 15, by the late 1990s, the richest 1% of Americans controlled 95% of the country's financial wealth,⁷ up from 48% in 1989. Thus, even though returns to capital (productive and financial) are responsible for less than 30% of income in most developed countries, almost all of that income flows to a small sliver of the population. Between 1997 and 1999, the wealth of the Forbes 400 richest Americans grew by an average of \$1,287,671 per day per person. In contrast, between 1985 and 1997, the net worth of the bottom 40% of households declined 80%.⁸ Returns to financial wealth, profits, and interest are a major factor in the income disparities seen in the U.S. and many other countries.

Capitalist systems are presumed to be populated by capitalists, and capitalists are the individuals who own the capital. Yet in most so-called capitalist nations, very few people are actually capitalists. Market economies are based on ownership, which is responsible for the impressive productive efficiency of such systems. A broader distribution of capital ownership could enhance the efficiencies of the market economy, and if done correctly, it could actually increase the ability of the system to

⁷J. Gates, *Democracy at Risk: Rescuing Main Street from Wall Street—A Populist Vision for the 21st Century*, New York: Perseus Books, 2000. Gates cites E. N. Wolff, "Recent Trends in Wealth Ownership," paper for Conference on Benefits and mechanisms for Spreading Asset Ownership in the United States, New York University, December 10–12, 1998.

⁸Gates, *op. cit.* Since the net worth of the bottom 40% of households is small, it does not take much of an absolute decline to reduce it by 80%. Nevertheless, the figure is dramatic.

provide important nonmarket goods and services. These claims demand some justification.

A good place to start might be an analogy with land ownership. Numerous studies have shown that land worked by an owner with secure title is more productive than land worked by sharecroppers or wage laborers.⁹ This makes sense. Making land productive requires investments in its productive capacity. A sharecropper or squatter will have little incentive to invest in the productive capacity of his land, and a wage laborer even less. In the case of the sharecropper, returns on the investment must be shared with the landowner, who at any time is able to evict the sharecropper. A squatter also cannot be certain that he will retain control of his land a year hence, and will not risk investing resources in the presence of such uncertainty. These points are widely accepted by economists and are considered ample justification for land ownership by the individual.

Yet the labor force in industry may have even fewer incentives to increase productivity than sharecroppers. What incentive do wage laborers have to do any more than the minimum required to keep the job, especially in jobs where managers have little chance to distinguish between the productive capacity of different workers? Workers have the most familiarity with the work they do, and in many cases may therefore have the best insights into how to do it faster, better, and cheaper. However, if there are no immediate benefits for the worker from more efficient production, why should she waste her time thinking about how to achieve it?

In addition, as we have pointed out, work is where many of us spend most of our waking hours. Economists typically consider work a disutility to be endured only to gain access to the material goods that provide us with utility, but there is no reason this should be the case. An economic system should not be devoted to the most efficient means of producing material goods, but rather to the most efficient means of producing human well-being. Most owners of capital concentrate on maximizing profits. It is rarely the case that profit maximization alone will create working conditions that generate the greatest worker well-being.

Imagine a company in which the workers own significant shares of stock. Such programs, known as Employee Shareholder Ownership Programs (ESOPs), are already widespread throughout the world. In ESOPs, workers do not manage the company, but they do have the same influence over management decisions that shareholders enjoy. Under ESOPs, worker income is composed of wages plus profits on stocks. Workers have much more of an interest in the profitability of the company. If there are

⁹E.g., A. Brandão, P. Salazar, and F. G. Feder, *Regulatory Policies and Reform: The Case of Land Markets*. In C. Frischtak, ed. *Regulatory Policies and Reform: A Comparative Perspective*, Washington, DC: World Bank, 1995, pp. 191–209.

mechanisms through which workers can make suggestions, it is in the workers' self-interest to think about ways to improve production. Workers are also concerned about how other workers perform, as it now affects their income as well. The net effect is usually an increase in productivity. If the worker-owners control a large enough share of the stocks, they are likely to work toward making the workplace a more desirable place to be, a place that satisfies a variety of human needs. Rather than an adversarial relationship between workers who want only benefits for themselves and capitalists who desire only profits, worker-owners will strive for a balance between the two. If measured in terms of the ability to satisfy human needs, efficiency is likely to increase under worker-ownership, though material production may not.

It is worth mentioning that corporations often offer substantial stock options to CEOs, with a rationale similar to ESOPs. Because CEOs do not own the companies they manage, they may try to manage a company to maximize personal benefit rather than corporate benefits. If stock options form a substantial portion of CEO salaries, then what's good for the corporation (at least in the short term, and as measured by stock value) is also good for the CEO. The problem is, as we have recently seen with the slew of corporate accounting scandals (Enron, WorldCom, and others), some CEOs may focus too hard on the short run and only on the value of stock. Stock values can be inflated through accounting fraud, and CEOs often have enough information to bail out before the crash. In addition, this type of "ESOP" generally aggravates the existing gross inequalities in income and wealth distribution.

Expanded ownership opportunities can also help address externalities. Many industries generate considerable pollution with highly negative impacts on the local population. If owners live far away, they will seek to maximize profits and in so doing, may ignore these negative externalities to the extent allowed by law. What happens if instead sufficient ownership of the industry resides with the local population to give them influence in management? The local population will strive for a balance between the negative externalities and the profits. In effect, the negative externalities have been internalized, a necessary condition for an efficient solution. Transaction costs will be reduced to those of coordination among shareholders, which exist in any publicly owned firm. Such outcomes can be achieved through Community Shareholder Ownership Programs (CSOPs).

An additional problem with ESOPs and CSOPs is concentration. Economies evolve, and as part of that process, firms go bankrupt, no matter how well they are run. If individuals have all their capital invested in one firm, a bankruptcy can have devastating impacts on their well-being.

Mechanisms for Distributing the Ownership of Capital. Broad-based ownership of capital may be an effective tool for improving distribution, increasing the efficiency of the economy in satisfying human needs, and in internalizing externalities. The question is: What policies will help achieve this goal? Simply taking ownership rights away from current owners and handing them over to workers is likely to be unfair and is in any case too radical, departing from our principle that we must pay attention to initial conditions. More feasible alternatives abound, but we will only touch upon a few of them here. First, productive assets wear out and must continually be replaced. Working toward broader ownership does not require that we directly redistribute existing property, but rather that we change ownership patterns for new capital. Second, not only do mechanisms for achieving this exist, but they have been tested in numerous countries and have received support from across the political spectrum.¹⁰

ESOPs are perhaps the most widely used system for broadening ownership patterns in capitalist countries. In the U.S., by 1996 some 9 million employees were participating in ESOPs, who controlled about 9% of corporate equity in the country. In the case of United Airlines, pilots led a worker coalition to purchase 55% of existing stock to increase their say in managing the company. (The company is currently in bankruptcy proceedings—a good warning against putting all your eggs in one basket!) In other companies, workers are awarded stock as a benefit, perhaps in place of a profit-sharing plan. In yet others, workers may be awarded stock to defend against a hostile takeover by other corporations (workers often lose their jobs in such takeovers, and may therefore be reluctant to sell their shares). Some corporations sell stock to workers to fund expansion, or even takeovers of their own. A number of tax incentives and other subsidies have been used to encourage ESOPs.

Given the advantages of more broad-based capitalism, there are a number of other feasible strategies governments could use to encourage this phenomenon. Government contracts, purchases, licenses (e.g., for public airwaves), and privatization programs could all show preference for companies that promote broad-based ownership.¹¹ Existing loan subsidies, such as loan guarantees many governments offer to purchasers of national exports, could be reconfigured to benefit only companies promoting broad-based ownership. National and international development banks could offer preferential loans to such companies. Innumerable other examples of corporate welfare exist and could similarly be channeled toward

¹⁰E.g., right-wing U.S. President Reagan supported ESOPs, as did Robert Reich, the left-wing Secretary of Labor under President Clinton. J. Gates, *The Ownership Solution: Towards a Shared Capitalism for the 21st Century*, New York: Perseus Books, 1998.

¹¹Privatization, the sale of public (government) assets to the private sector, has been occurring at a breakneck pace throughout the world, often as part of IMF structural adjustment programs.

creating capitalists. Another requirement would be to train people how to become capitalists and manage capital. Public schools in most countries train people to become workers, but not owners.¹² In sum, capitalist societies need more capitalists!

Distributing the Returns to Natural Capital

Ownership of land and other natural capital is also quite concentrated throughout the world. While the factor share of rent is generally calculated to be only 2% of income, this calculation ignores two major sources of returns to natural capital. First, returns from the extraction of natural resources are often classified as profit, when in reality most of the returns are actually rent. (Recall that rent is the profit above and beyond what is required to supply the resource. The supply of nonrenewable resources is fixed, and the sales price of many renewable resources is often higher than would be required to supply the market.) Second, many of the returns to natural resources are in the form of hidden subsidies. For example, when an industry pollutes water or air and is not required to pay for the costs this imposes, the industry is capturing the returns to the waste absorption capacity of the environment.

Ending Public Subsidies. When the state owns the resources in question, extractive industries are typically required to pay royalties on those resources. In many cases, these royalties are quite small. The state should be able to charge a royalty equal to the scarcity rent.¹³ By spending the royalty on public goods or using it to reduce taxes, the state can use rents to improve the distribution. In some primary industries, government subsidies to natural resource extracting corporations can be quite blatant. A number of examples from the United States illustrate this point.

Under the Mining Law of 1872, corporations can purchase the surface and mineral rights to federal land for \$2.50–\$5.00 per acre, depending on the nature of the mineral deposit.¹⁴ This law was originally designed to provide incentives for people of European descent to settle the American West, but now it is little more than a giveaway to large corporations, many of which are not even from the U.S. Publicly owned rangeland is frequently leased to big ranchers at a fraction of the fair market value.¹⁵

¹²These and other policies are described in greater detail in Gates, *The Ownership Solution*, op. cit.

¹³D. M. Roodman, *The Natural Wealth of Nations: Harnessing the Market for the Environment*, New York: Norton, 1998.

¹⁴M. Humphries and C. Vincent, CRS Issue Brief for Congress: IB89130: Mining on Federal Lands, May 3, 2001. National Council for Science and the Environment. Online: <http://www.cnle.org/nle/mine-1.html>.

¹⁵B. Cody, *Grazing Fees: An Overview*. CRS Report for Congress, 1996. Online: <http://cnle.org/NLE/CRSreports/Agriculture/ag-5.cfm>.

Rights to timber in national forests are often sold for less than it costs the government to build the access roads to the resource, or at times even for less than it costs to prepare the bids.¹⁶ As a result, many publicly owned forests in the U.S. are logged for timber, even when in private hands logging would not be economically viable.

One of the more controversial pieces of environmental legislation in the 1990s was the timber salvage rider attached to the bill that provided federal assistance to the victims of the Oklahoma City bombing.¹⁷ This legislation sought to “salvage” all trees on national forests that were threatened by insects or fires by chopping them down. The bill suspended environmental regulations for salvage operations and explicitly stated that timber should be sold at a financial loss to the government, if necessary. As most forests are at some risk from insects and fire, the bill was little more than a massive giveaway of public resources.

The federal government is currently pursuing a similar policy, known as the Healthy Forests Plan. While all of these examples are from the United States, similar policies are in place worldwide. Getting rid of all of these subsidies would reduce the loss of ecosystems and their services, save taxpayers money, and generate abundant new government revenue.

Alaska Permanent Fund and Sky Trust. The state of Alaska has taken a direct step toward distributing the income from natural capital. Alaska charges royalties for extraction of its abundant oil reserves. These royalties go into the Alaska Permanent Fund. Interest on this fund is distributed to all residents of Alaska. The idea is that the natural capital of Alaska, or at least its oil supplies, do not belong to corporations, but rather are the common property of all Alaska residents. Putting the money in a trust fund helps ensure that even when the oil is exhausted, future Alaska residents can share in the bounty.

Peter Barnes, an eco-entrepreneur, has proposed a “sky trust” similar to the Alaska Permanent Fund to distribute income from nonmarketed natural capital. He begins by asking: Who owns the sky? His answer is that the sky is a common property resource, owned by all citizens of a country. Yet some people use the waste absorption capacity of the sky more than others. Industries pollute the sky without paying, and some individuals pollute far more than others. Because there are few institutions defending our common property rights to the sky, it is treated as an open access resource, with the well-known results of poor air quality, acid rain, climate change, and other ill effects.

¹⁶R. Gorte, *Below-Cost Timber Sales: Overview*. CRS Report for Congress, 1994. Online: <http://www.cnie.org/nle/for-1.html>.

¹⁷A rider is a piece of legislation attached to another, unrelated bill. The bill cannot be passed without the rider.

The sky trust is a bundle of policies designed to address scale, distribution, and allocation. The scale and allocation components are achieved by establishing quotas, then auctioning them off in the form of individual tradeable quotas. We have already discussed how these mechanisms work. All the returns from the sales of these quotas would go into a trust fund, the returns from which would be distributed equally to all citizens in the form of cash. All citizens would receive equal shares, but those who pollute more would pay more, so redistribution would occur.¹⁸ The same basic idea could be used with all natural capital on the assumption that it is a gift of nature to all humans, and not just to a select few.

While this is a very promising policy for a number of reasons, the redistribution mechanism also raises some concerns. The sky is a public good, as are many of the other ecosystem services suitable for sky trust-type policies. Cash payments, in contrast, entitle the recipients only to market goods. In effect, this policy would channel the receipts from rationing public goods mainly toward the purchase of private goods. When people spend the money they receive from the trust fund, it will stimulate the consumption of other natural resources and the creation of waste.

We believe another option should be considered—that of using the trust to fund much needed expenditures on public goods. However, under the current system, with hundreds of billions of dollars spent annually, to convince people that the consumption of market goods is the only path to happiness, accompanied by a pervasive public distrust of government, direct cash disbursements might be more politically feasible than spending the trust on public goods, or even on tax reductions. If direct cash disbursements are necessary to make the sky trust politically feasible, the approach has enough to recommend it that it is still well worth pursuing. As people in the future come to better understand and value the importance of nonmarket goods, perhaps the trust could be turned over to the creation of public goods.

Land Tax. Land is another part of the commonwealth, an asset provided by nature that originally belonged to all citizens of a nation. One can do as one likes with many assets, but nations almost always try to retain sovereignty over their territory. Yet land ownership in most countries is highly concentrated, as are the returns to ownership.¹⁹ In addition, as we pointed out in Chapter 11, the value of Ricardian land is almost entirely the result of positive externalities—land is more valuable as a result of proximity to others. In other words, land values are created by society, not

¹⁸P. Barnes, *Who Owns the Sky? Our Common Assets and the Future of Capitalism*. Washington, DC: Island Press, 2001.

¹⁹In the U.S., the richest 10% own 60–65% of land by value, and in Brazil, the richest 1% own 50% of the countryside. Roodman, *op. cit.*

by the landowner. Land supply is also fixed. With extremely limited exceptions, no matter how high the price, more land will not be created, and no matter how low the price, the same quantity of land will exist. Therefore, all returns to land are economic rent, payment above the minimum necessary supply price. With supply fixed and demand increasing, both the price of land and the rent on that land will increase, leading to increasing concentration of wealth and income.

A line of thinkers following economist Henry George (whose name has cropped up in previous chapters) argue that as society creates the value in land, society should share in the returns to land. While redistributing land itself would be a difficult, disruptive, and politically infeasible policy in many countries, it is much simpler to simply redistribute the rent via land taxes. In the extreme, some proponents of this approach call for a single tax on land, though many interpret this as a tax on all resources that are a gift from nature. In most countries, property taxes fall on both land and the infrastructure on it, and these are two very different types of resources. We should tax that to which value is added, and not the value added. Such a policy has many desirable features.

In terms of distribution, a higher tax on land will drive down the value of land because it drives up the cost of owning it. Theoretically, the price of land should equal the net present value of all future income streams from that land. As a higher tax reduces the income stream, it reduces the price. A land tax also makes land speculation much less profitable. It simply becomes too expensive to pay taxes every year while waiting for land prices to rise. Removing the speculation demand for land reduces land prices even further. Lower prices make land and home ownership more broadly accessible, especially if higher taxes on land are accompanied by reducing or eliminating taxes on buildings on that land. And the entire tax on land will be paid by the landowner—it will not be shifted onto renters because the supply of land is perfectly inelastic. This point is explained in greater detail in Figure 22.1. It is also worth noting that speculative bubbles are a source of economic instability, so a land tax can help stabilize an economy. George argued that almost all business cycles were driven by land speculation.

THINK ABOUT IT!

Can you explain why eliminating profits from speculation will reduce the cost of land? Think in terms of supply and demand.

Land taxes can also help reduce urban sprawl and all of its negative impacts. Those who own land will have greater incentive to either invest in its productive aspects or sell it to someone who will invest. The higher the land value, the higher the pressure to invest or sell. The net result is not necessarily more investment, but rather more investment on the most

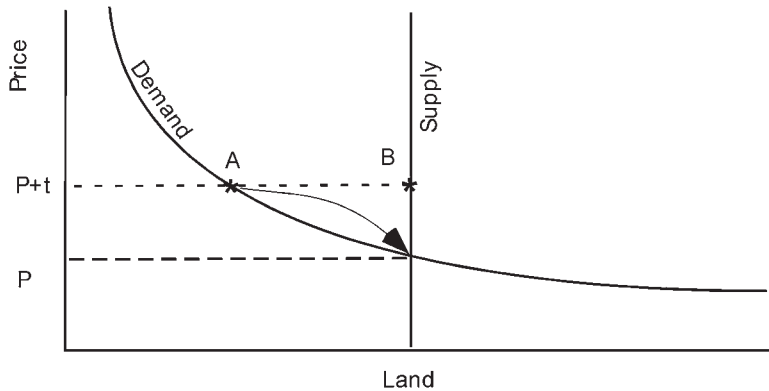


Figure 22.1 • Who pays the land tax? The supply of land is perfectly inelastic with respect to price. This means the same amount will be supplied at any positive price. Demand, in contrast, is sensitive to price, and there is one market price at which supply and demand clear. If land is taxed by amount t , and the land seller tries to pass the tax on to the buyer or renter, raising the price temporarily to $P + t$, the market will not clear; amount AB will remain unsold. Landowners who were unable to sell or rent at the higher price will therefore be forced to lower their price, driving the price back down to P . This means the government can tax away the entire rent on land, without reducing supply.

valuable land. Land values are higher the denser the population. This means that land in cities will be more intensively developed, reducing the pressures for urban sprawl.

A land tax accompanied by eliminating the property tax on buildings would reduce the cost of supplying buildings, and more buildings would be built. This approach would improve distribution, because poor people often spend the majority of their income on rent, and a greater supply of buildings would mean lower rent. However, to achieve desirable distribution and scale effects, the land tax (or quotas) should extend to all free goods from nature. In this case, new buildings would have to pay taxes (or quota costs) on new materials consumed, while the renovation of older buildings would not, thereby shifting investments into restoration. Currently, restoring older buildings increases their value and hence the tax burden, and the tax deters the activity. Greater investments in restoration relative to new construction will further favor urban centers over new suburban construction.

In most cases, governments provide the infrastructure for suburban developments, which is basically just a subsidy for the people who move to the suburbs. Cities that have pursued high land taxes accompanied by low or zero taxes on infrastructure on that land include Melbourne, Pittsburgh, Harrisburg, and other cities, and it is well established that the taxes have indeed limited sprawl and led to urban renewal. Land taxes are most

beneficial in urban settings. In rural areas, land left in its natural state is already providing ecosystem services and often should not be taxed.

THINK ABOUT IT!

Can you explain why eliminating the tax on buildings will increase the supply and reduce rents?

■ ADDITIONAL POLICIES

Two other distribution policies merit a brief mention. First, if governments recaptured the sole right to seigniorage (the right to print money), they could use that money to improve distribution. Second, we can look at “demand-side” policies. Where do poor people spend most of their income? On rent and health care.²⁰ We have already seen how shifting taxes from buildings to land could drive down the cost of housing.

Universal, government-sponsored health care, as most citizens of developed countries already enjoy, would also dramatically increase real income for the poor. It would be likely to decrease national expenditures on health care as well, as the U.S. currently spends a higher percentage of its national income on health care than any other country, and still fails to insure over 40 million Americans. Though it is a highly contentious topic, health care is probably as poorly suited to market provision as ecological services.

BIG IDEAS to remember

- | | |
|-----------------------------|-------------------------|
| ■ Minimum income | ■ Rent |
| ■ Caps on income and wealth | ■ Public subsidies |
| ■ Wealth, status, and power | ■ Alaska Permanent Fund |
| ■ ESOPs and CSOPs | ■ Land Tax |

²⁰G. Flomenhoft, “The Triumph of Pareto: Does Equity Matter?” Unpublished working paper, presented at the US Society for Ecological Economics 2003 conference, Saratoga Springs, May 23, 2003.



Efficient Allocation

Allocation receives the lion's share of attention in conventional economic thinking, but it plays a tertiary role in the ecological economic approach to policy. This is not to deny its importance. Efficient or at least cost-effective allocation is, in fact, a vital component of good policy, and we have seen how the various policies discussed so far effectively allocate resources. In this chapter, we turn to some of the “big picture” issues in allocation.

1. We address the myth that if mechanisms can be developed for internalizing all external costs and valuing all nonmarket goods and services, the market alone will lead to efficient allocation.
2. We examine the asymmetric information flows that shape our preferences, and how these influence resource allocation. Allocating resources efficiently toward goods that do little to increase our well-being is not efficient.
3. We return to policies aimed at macro-allocation, the allocation of resources between private and public goods.
4. We examine some problems confronting the allocation of resources under local control and national sovereignty that supply global public goods.
5. We propose an expanded definition of efficiency more compatible with the goals of ecological economics.

■ PRICING AND VALUING NONMARKET GOODS AND SERVICES

Recall from Chapter 8 that markets lead to efficient allocation of market goods by using the price mechanism as a fulcrum on which to balance

supply and demand. Many economists argue that if we could determine monetary values for nonmarket goods, we could then use the market mechanism to efficiently allocate them. As a result, one of the most active research areas in environmental economics is the calculation of “prices” for nonmarket goods. Once prices are established, we need a mechanism to “internalize” these values into the market system. While it is critically important to establish the value of nonmarket goods and services, it is controversial whether establishing monetary values, which are identical to exchange values, is appropriate or meaningful. Internalizing those values is, in any event, not a simple task.

We already discussed in Chapter 20 the problem of circularity in using prices to determine the optimal scale when those prices were already based on the initial scale. A similar problem arises when we try to put market values on nonmarket ecosystem services, an exercise that is sometimes thought necessary to determine the costs and benefits of changes in scale. However, even if we could overcome that problem, we would still confront a number of serious difficulties.

Recalculating Marginal Values

You will recall the resolution of the “diamonds-water paradox” from Box 14.1. Water is far more valuable than diamonds, yet it is available at a very low price. Price is exchange value, or the marginal use value of the good or service in question. Use value is the total value, or the value of all units together. The use value of water is infinite, yet where this resource is very abundant, the value of an additional unit approaches zero. However, when water is extremely scarce, an additional unit may mean the difference between life and death, so its marginal value also becomes immeasurably large. The same is true for any essential good or service, such as the life-support functions of ecosystems: When an essential resource is scarce, the marginal value is extremely high, and it increases rapidly with growing scarcity.

Around 150 years ago, many ecosystem goods and services were so abundant an extra unit had no appreciable value. As a result, the economic system ignored the value of such goods. Over time, however, these goods and services have become increasingly scarce and their marginal values have soared, which is why economists now attempt to calculate their values. As we approach ecological thresholds, which we may already be doing, the marginal value and hence “price” of these goods and services will increase extremely rapidly.

To internalize these ecosystem values, we would need to continuously recalculate them, centralize the information, then feed it back into the market mechanism via taxes or subsidies. Yet calculating the value of such resources is very expensive, and centralizing the knowledge and feeding it back into the pricing mechanism would require an enormous and expen-

Box 23-1 PRICING NATURE AND LIFE

In response to a legal demand to assess the value of claims in lawsuits, economists have designed a number of *ad hoc* methods of estimation. Probably this began with “statistical value of a life” calculations to assess damages in cases involving legal liability for accidental death. The usual procedure is to calculate the present value of expected future lifetime earnings of the deceased. Most people would not give up their life voluntarily for such an amount, primarily because they would no longer be around to receive the payment, but even if the payment went to their heirs, there would be few volunteers. Clearly this sum is in no sense “the value of a life.” But as a practical procedure for settling legal cases involving accidental death, it is not unreasonable—as long as we always remember that we are valuing human capital as an object, not human beings as subjects. But the question remains: Should compensation be for lost human capital or the lost life of a person?

Question: If you contributed to a fund for the survivors of the victims of September 11, would you prefer that compensation be made according to the differing human capital values of the victims, or an equal amount for each victim?

Given this precedent of valuing human capital, it was a short step to price the loss of ecosystem services and natural capital caused by accidental oil spills such as the Exxon Valdez, industrial accidents, and so on. As a practical way of making after-the-fact compensation for accidents, this is not unreasonable. Problems arise when we move from after-the-fact to before-the-fact, from adjustments to an involuntary occurrence to an imagined voluntary tradeoff. This is what **contingent valuation** estimates do. They present people with imaginary alternatives and ask them, before the fact, and indeed totally hypothetically, how much they would be willing to pay, say, to save 100 grizzly bears, or how much they would accept for the loss of 100 grizzly bears. Interestingly, the two questions usually yield very different answers, although in theory they should be the same for a market transaction. Alternatively, the grizzlies could be valued as a public good. Citizens could be asked how much they would be willing to be taxed along with everyone else to save 100 grizzlies. This makes more sense, but it is still very artificial.

There are other methods besides contingent valuation for estimating hypothetical prices for nontraded goods. Environmental economics texts discuss them in detail. Why we have given little attention to this subject is, we trust, evident from our discussions of the separateness of allocation and scale, and the circularity of deciding scale questions by allocative prices.

sive bureaucracy. The paradox is that we love markets precisely because they constantly and almost costlessly recalculate prices on the basis of decentralized information with minimal government intervention. However, this approach to allocation would be expensive, centralized, and require large-scale government intervention.

Uncertainty, Ignorance, and Unfamiliarity

In addition, methodologies for valuing nonmarket goods are fraught with problems. Most rely on artificially constructed markets or ways of inferring nonmarket values through existing markets. Two problems in particular merit discussion: our lack of knowledge of ecosystem function and our lack of familiarity with valuing nonmarket goods.

As an example, the contingent valuation method constructs a hypothetical market basically by asking people what they would be willing to pay for a given nonmarketed good or service. The problem, as we have repeatedly discussed, is that even the experts are ignorant of all the goods and services healthy ecosystems provide, how they provide them, the impacts of human activities on their provision, where critical ecological thresholds lie, and the outcomes when these thresholds are passed.

If we emit a given stream of pollutants into a lake, what will the impact be? What ecosystem services will be lost? Will the waste flow accumulate, causing worse damage over time—perhaps irreversible loss? Will the loss of the system being polluted affect other systems? What is the time scale involved? Even if it were possible for the experts to resolve all these uncertainties (which it is not) and disseminate that information to the population at large, people have no experience with markets in such goods and services, and would still have a very difficult time assigning meaningful exchange values.

Time, Distribution, and Valuation

Yet another problem is the time factor. Most ecosystem goods and services are renewable, and hence will provide benefits into the indefinite future. A typical decision is whether to sacrifice a renewable flow from a natural fund-service for a nonrenewable (manmade) fund-service or for a one-time liquidation of stock. This demands that we compare present values with future values. As we discussed in Chapter 15, economists generally do this by discounting future values. The discount rate will typically be one of the most important variables in determining value, and there is no agreed upon objective rule for determining an appropriate rate.

We must also recognize that the question of what should be left to the future is inherently an ethical decision concerning intergenerational *distribution*. Conventional economists argue that the question is not one of

Box 23-2**METHODOLOGIES FOR MONETARY VALUATIONS OF ECOSYSTEMS**

Several methods are available for putting dollar values on the nonmarketed goods and services provided by ecosystems. Many of these are appropriate for valuing only a small subset of services. Most textbooks in environmental economics provide an adequate introduction to these methodologies. We recommend as a good starting point the Web site “Ecosystem Valuation” at <http://www.ecosystemvaluation.org>, where the following methods are listed:

- *Market Price Method*: Estimates economic values for ecosystem products or services that are bought and sold in commercial markets.
- *Productivity Method*: Estimates economic values for ecosystem products or services that contribute to the production of commercially marketed goods.
- *Hedonic Pricing Method*: Estimates economic values for ecosystem or environmental services that directly affect market prices of some other good.
- *Travel Cost Method*: Estimates economic values associated with ecosystems or sites that are used for recreation. Assumes that the value of a site is reflected in how much people are willing to pay to travel to visit the site.
- *Damage Cost Avoided, Replacement Cost, and Substitute Cost Methods*: Estimate economic values based on the costs of avoided damages resulting from lost ecosystem services, the costs of replacing ecosystem services, or the costs of providing substitute services.
- *Contingent Valuation Method*: Estimates economic values for virtually any ecosystem or environmental service. The most widely used method for estimating nonuse or “passive use” values. Asks people to directly state their willingness to pay for specific environmental services, based on a hypothetical scenario.
- *Contingent Choice Method*: Estimates economic values for virtually any ecosystem or environmental service. Based on asking people to make tradeoffs among sets of ecosystem or environmental services or characteristics. Does not directly ask for willingness to pay; this is inferred from tradeoffs that include cost as an attribute.
- *Benefit Transfer Method*: Estimates economic values by transferring existing benefit estimates from studies already completed for another location or issue.

distribution, but rather efficient allocation. If a resource will be sufficiently more valuable in the future than in the present, it should be saved for the future. Therefore, maximizing the net present value (NPV) of resource use will lead to the optimal allocation. However, NPV is the value

of present and future resources to this generation. If you recall our discussion of property rights from Chapter 20, this corresponds to a property rule assigning property rights to the current generation, which is free to interfere with the future's access to resources. Under this approach, all that matters is the value of resources to the present generation.

As an alternative, we could assign some resource property rights to future generations. For example, assigning rights via a liability rule would leave this generation free to use resources as long as it compensated the future with an equivalent amount of other resources. Under an inalienability rule, the future would be entitled to a certain share of resources, and the present would be obliged to leave them. These three rules are simply different initial distributions of resources, and each would lead to a different set of prices for both market and nonmarket resources. Which rule to use is an ethical decision, not a matter of allocative efficiency.

Market vs. Nonmarket Values

Assuming that all these other issues are addressed, would monetary valuation then lead to efficient allocation? Price provides a feedback mechanism used by the market to maximize profit, which economists assume creates the appropriate conditions for maximizing human well-being. Is a single feedback mechanism sufficient for allocating all the resources that contribute to human well-being? Could natural ecosystems in all their incomprehensible complexity function with only one feedback mechanism? Some ecologists might argue that ecological systems do indeed function this way, where maximizing energy consumption is the ultimate feedback system.

Although energy consumption is a useful simplification in some ecological models, it is not the only driving force in nature. And even if it were, as a principle of competitive exclusion, it does not translate well to the human economy, in which maximizing a cost is surely uneconomic. Moreover, human ethical beliefs make the interaction between the human economy and the ecosystem more complex than the functioning of the ecosystem alone. We therefore cannot support the reductionist approach of assuming that the profit motive alone is sufficient to maximize human well-being, much less to guide us in any quest toward an ultimate end.

As a concrete example of the problems with monetary valuation of everything, it would be quite simple to develop a methodology for calculating a dollar value for democracy. Certainly people in general have a better understanding of democracy than of ecosystem services, and we could readily devise a survey that would tell us how much a voter would be willing to pay for the right to vote (or alternatively the minimum amount for

which a voter would sell their vote). We could do the same for human rights, and many people consider the right to live in a healthy environment such a right. But most people would probably agree that politics and human rights are in a different moral sphere than economics, and power in the sphere of economics should not translate to power in these other spheres.¹ (While this, of course, does happen, people do not generally consider it desirable.) Political rights, human rights, and other ethical values are not individual values, but social values. Attempting to estimate social values by aggregating individual tastes suffers from the fallacy of composition and is a categorical mistake.

THINK ABOUT IT!

Would you sell your right to vote? If there were such a market for votes, would you expect the price to be high or very low?

It is true that we are constantly forced to make decisions between mutually exclusive alternatives, such as more forests or more strip malls, which require a comparison between market and nonmarket values. However, many nonmarket goods are fundamentally different from market goods in ways that make “scientific” comparison not only impossible but also undesirable. Putting dollar values on everything does not make the necessary decisions more objective; it simply obscures the ethical decisions required to make those “objective” valuations.

Most textbooks in environmental economics devote considerable space to discussing methodologies for valuing nonmarket goods and services. Valuing ecosystems can play an important role in capturing the attention of the public and policy makers, and can offer insights into appropriate economic policies. But attempting to calculate an exchange value for all nonmarket goods, then use that value to decide what we will preserve and what we will destroy, is an example of economic imperialism, as discussed in Chapter 3.

Ecological economics takes the broader perspective that such methodologies are inadequate to capture the range of human values and physical needs we have for nonmarket goods. Instead of spending time trying to calculate the “correct” price for nonmarket goods, ecological economics stresses that we should act on our knowledge that zero is the incorrect price, and spend our time trying to improve upon and implement policies that recognize they have significant, often infinite value, even if we cannot precisely quantify it.

¹M. Walzer, *Spheres of Justice*. New York: Basic Books, 1990.

■ MACRO-ALLOCATION

As we discussed in Chapter 16, macro-allocation is the problem of how to allocate resources between the provision of market and nonmarket goods. The government plays an important role in providing nonmarket goods and can also influence demand for market goods through the use of taxes and subsidies. Presumably, in democratic countries, citizens will elect politicians who will make the right choice regarding macro-allocation. One serious problem with this assumption is the discouraging lack of information people have regarding nonmarket goods and services. For people to make appropriate choices, they require appropriate amounts of information. In this section, we will first look at policies addressing unequal information flows concerning the attributes of market and nonmarket goods, then examine possibilities for the government to provide incentives to the private sector for providing public goods.

Asymmetric Information Flows²

Asymmetric information is where either the buyer or seller has information that the other does not have, and that information affects the value of the good or service exchanged. Economists have long known that asymmetric information is a market failure, generating serious inefficiencies. For example, if I am selling a car, I know how well it works, but the potential buyer does not. The buyer will adjust the price she is willing to pay based on the risk of purchasing a lemon, and this risk-adjusted price will be less than the value of a good used car. The rational seller will not be willing to sell a good car at the risk-adjusted price, and the market will provide only lemons (at least according to theory). Joseph Stiglitz won the Nobel Prize in economics for such basic insights.

We face a related problem with the asymmetry of information flows that form our preferences. While many economists argue that preferences are innate, businesses are betting an estimated \$652 billion per year that preferences are heavily influenced by advertising.³ Advertising costs money, and it can only pay for itself by advertising market goods. Most words we hear today are direct sales pitches for market goods and the programs sponsored by them.⁴ In stark contrast to advertising for market goods, very little money is spent convincing people to prefer nonmarket

²Much of this discussion is adapted from J. Farley, R. Costanza, P. Templet et al. Synthesis Paper: Quality of Life and the Distribution of Wealth and Resources. In R. Costanza and S. E. Jørgensen (eds.), *Understanding and Solving Environmental Problems in the 21st Century: Toward a New, Integrated Hard Problem Science*. Amsterdam: Elsevier, 2002.

³International Advertising Association, 2000. To place this figure in context, only seven countries in the world had a GNP higher than \$600 billion in 1997.

⁴A. T. Durning, *How Much Is Enough? The Consumer Society and the Fate of the Earth*, New York: Norton, 1992.

goods. To the extent that advertising alters preferences, it systematically does so in favor of market goods over nonmarket goods.

People have a finite amount of resources to allocate. If advertising convinces us as a society to allocate more resources toward market goods, correspondingly fewer are available to allocate toward nonmarket goods. And as we know, all resources allocated toward consumer goods are extracted from nature and return to nature as waste. Seen from this light, advertising convinces us to degrade or destroy public goods for private gain. It appears that current levels of consumption in the developed countries are incompatible with a sustainable future; yet reducing consumption levels will be exceedingly difficult in the presence of so much advertising.

Nor is this the only market failure associated with advertising. Arguably, human welfare is determined by our ability to satisfy our needs and wants. Advertising creates wants by making us believe we need some product or another, yet it gives us no greater ability to satisfy those wants. In this sense, advertising directly diminishes our welfare. We can make this point no better than B. Earl Puckett, former head of Allied Stores Corporation, who once declared that “it is our job to make women unhappy with what they have.”⁵ In this line of thinking, advertising is basically a “public bad.”

The problem is one of providing symmetric information flows for nonmarket goods. This is a very contentious issue. We briefly present several possibilities here for discussion, but welcome new and better ideas for addressing the market failures associated with advertising.

The first involves the recognition that advertising over the airwaves in many countries is subsidized. The airwaves are valuable public property, but are often given free of charge or at low cost to communications corporations. Since transmissions beamed over airwaves have properties of public goods in that they are nonexcludable (at least when the transmissions are not scrambled) and nonrival, there is a solid rationale for giving away airwaves to those who will beam such transmissions, in spite of the fact that the airwaves themselves are rival, excludable, and scarce. However, if the government charged corporations for the use of airwaves for advertising, it would target only that portion of the airwaves devoted to private profit. Advertising is currently considered a business cost and is tax-exempt. For the reasons listed above, however, it would be more appropriate to tax advertising as a public bad. At a minimum we should not allow advertising to be written off as a cost of production. The rationale is that production is supposed to meet existing demand, not create new demands for whatever happens to be produced.

While taxes would presumably reduce the quantity of ads for market

⁵Ibid., pp. 119–120.

goods, it would not help to generate concern for nonmarket satisfiers of human needs. There are several alternatives for helping to achieve this goal. One approach would be a law mandating “full disclosure” advertising. Just as medicines are labeled with all their potential adverse side effects, advertisements could list all the potential adverse side effects of the products they advertise. This would include, of course, all the negative impacts on the environment and their implications. Another alternative would be to provide free airtime for public service announcements that specifically seek to create demand for environmental services and other nonconsumptive satisfiers of human needs. The media is a phenomenally powerful tool for persuasion, and thus an effective approach to policy would be to make the information flows it provides more symmetric.

A problem with both of these restrictions on advertising is that people will complain that they infringe on the basic right of free speech. However, the right to free speech does have restrictions, it does not include the right to lie or misrepresent. Nor does it include the right to amplification by a powerful megaphone. For example, no one is allowed to shout “fire!” in a crowded theater if there is no fire, because it threatens the well-being of others. Shouting “fire!” may not be fundamentally different from encouraging people to consume when such consumption threatens the well-being of future generations. Many nations already curb advertising on alcohol and tobacco, and the Australian Consumers Association is currently attacking the right to advertise unhealthy foods on children’s TV shows.⁶ The same rationale also applies to curbing advertising that indirectly encourages destruction of the environment.

Subsidies for Nonmarket Goods

Even if people are well educated concerning the benefits of nonmarket goods and subsequently elect governments willing to provide them, there remains the problem of how best to do so. Often the best strategy will be for the government to simply supply them outright, or directly pay private sector contractors to do so. In many circumstances though, the nonmarket goods are positive externalities from the production of market or private goods. For example, when farmers terrace their land, use contour plowing, and retain buffer zones along streams, they may dramatically improve downstream water quality, thereby maintaining the productive capacity of their land for future generations. The problem is that the private sector will supply less of the positive externality than is socially desirable. When the positive externality is in the form of a public good, the best approach may be for the government to subsidize the portion of the private activity that generates the public good.

⁶Ibid.

Several types of subsidies are possible. A direct subsidy can simply compensate the private sector for its provision of the positive externality,⁷ at which point the externality is partially “internalized” and presumably supplied in more adequate amounts. Alternatively, tax relief can be used to subsidize positive externalities. Possibilities range from a decrease in land taxes for farmers who reduce erosion to tax breaks for business investments in training personnel (people these days regularly change jobs, and firms may offer less training than is socially optimal if the worker will move her newly acquired productivity to another firm). As another option, a subsidy can be in the form of subsidized credit. If people underinvest in activities with positive externalities, lower interest payments would stimulate greater investment.

Using Seigniorage

Where would governments find the money for subsidized interest rates? Again we suggest the option of restoring the sole right to seigniorage to the government, as discussed in Chapter 14. When banks create money, they do so through interest-bearing loans. On average, such loans must generate financial returns that can repay the loan plus interest, which means money is loaned for the production (and consumption) of market goods. Unless the economy is growing, it becomes impossible to pay back all loans with positive real interest rates. Governments, in contrast, could use their power to create money to make interest-free loans or even outright grants to activities that best promote the common good, including the provision of nonmarket goods. Not only would this help in the macro-allocation of resource toward nonmarket goods, it could lead to a financial system whose viability is not based on unending growth.

■ SPATIAL ASPECTS OF NONMARKET GOODS

In previous chapters we discussed the spatial characteristics of nonmarket goods. Most ecosystems provide services at the local level, the regional level, and the global level. For example, a forest can affect climate stability at each of these levels. Yet the principle of subsidiarity requires that the domain of the policy-making unit be congruent with the domain of the causes and effects of the problem with which the policy deals. The causes of ecosystem degradation are often at the local level; effects are felt at local, regional, and global levels; and policy-making institutions are primarily local and national. This poses serious problems for effective policy.

⁷In most cases, it is difficult to know the exact value of the nonmarket good and hence the optimal level of compensation. We do know that zero is the wrong value, and a reasonable compensation will improve allocation.

To make the problem more concrete, we will use the specific example of clearing tropical forests for farmland on the Atherton Tablelands of Australia. On private lands, the decision is local. The net private marginal benefits to farmers of clearing forests for agriculture decreases with area cleared. The first units cleared meet basic needs and are on the very best lands. Additional units cleared meet less important needs and utilize less adequate lands (e.g., steeper slopes, less fertile soils, and greater distances). Eventually, farmers unaware of the ecosystem services provided by forests cleared to the point that water quality was affected, and shade cover was inadequate, leading to reduced yields—that is, a negative marginal net private benefit (MNPB) to deforestation.

This scenario is depicted in Figure 23.1 by the curve MNPB. With the right to do as they pleased on their land, the rational, well-informed farmers should have cleared the forests until marginal benefits were zero, at point A. Instead, due to ignorance, they cleared to point B. It was not just the farmers that were ignorant—at the time this was occurring, few people, if any, knew of the negative impacts caused by deforestation.

Towns downstream of deforested farmlands suffer from irregular water flow and poor quality. The nature tourism industry in the region generates far more income than farming, and it also suffers from deforestation. These local marginal external costs of deforestation (MEC local) are also depicted on the graph. If the local governments had been aware of these negative externalities, they might have implemented policies designed to limit deforestation to point C, perhaps by imposing a local deforestation tax equal to OC' , or issuing tradeable deforestation permits in the quantity OC.

National MEC of deforestation includes local MEC. In addition, deforestation on the tablelands causes erosion, siltation, and nutrient runoff, all of which flows out to sea to be deposited on the coral reefs. This affects fisheries and tourism outside the shire boundaries. The state or national government should have implemented policies to limit deforestation to point D, perhaps through a national deforestation tax equal to $C'D'$, or by purchasing DC of the OC local level quotas and discarding them.

Global MEC similarly includes national MEC, in addition to deforestation's contribution to global climate change and biodiversity loss. Point E would therefore be the globally optimal level of deforestation, but there is of course no global authority that overrides national sovereignty on issues such as deforestation, so taxation would not be a policy option. Nor is there a global coalition that could purchase deforestation permits (in the event they happened to be internationally tradeable). While individual governments, multilateral institutions and international NGOs currently play some role in preserving ecosystems and reducing the rate of ecosystem degradation in some countries, it is almost certainly insufficient to

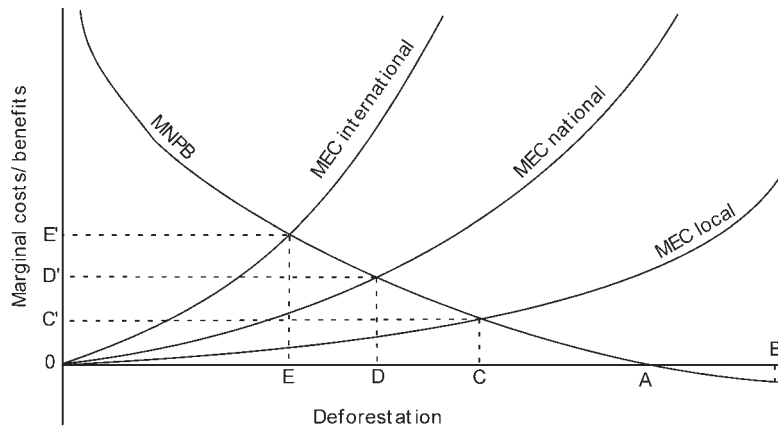


Figure 23.1 • Marginal costs and benefits of deforestation at different spatial levels. Curve MNPB shows the marginal net private benefits of deforestation to farmers on the Atherton Tablelands. This curve accounts for the cost of lost ecosystem services that directly benefit the farmer. Through ignorance, farmers initially cleared amount OB of forest, while had they been better informed, they would have cleared amount OA. Marginal external costs of deforestation are shown for local, national, and international society, along with the corresponding optimal levels of deforestation: C, D, and E, respectively.

achieve globally optimal levels of ecosystem preservation. The insufficiency of these efforts may result from a combination of ignorance and the free-rider effect.

The spatial distribution of ecosystem goods and services thus presents a serious problem. Control over ecosystems is usually at the level of the individual, who retains all the benefits of deforestation (e.g., timber sales and subsequent use of farm land) and shares the costs of lost ecosystem services with society. Where effective institutions and information exists, local and national governments can impose regulations that optimize ecosystem preservation from their points of view. Alternatively, civil society in the form of volunteer community organizations and NGOs can step in when governments fail to act. However, they can only rely on good will, altruism, or other behaviors inconsistent with the assumptions of rational self-interest, but thankfully important parts of the human psyche nonetheless. When governments do act, they are unlikely to pursue globally desirable levels of ecosystem preservation, as there are few incentives to do so, and national sovereignty allows them to ignore global benefits. Making optimal outcomes even more elusive, policy makers at all levels are probably unaware of the full range of benefits intact ecosystems actually supply.

While continued conversion of tropical forests may be uneconomic at the global level, in many of the world's remaining tropical forests (and

other healthy ecosystems), it may still make economic sense at the local level to continue the deforestation process. If we look at the Amazon rainforest, much of South East Asia, and Central Africa, there are still vast tracts of largely untouched forest along with high levels of poverty and landlessness. There is little question that for the individual, the best available alternative under current circumstances is often to clear forest and grow crops. The farmers who do this are not irrational. They may be ignorant of the ecosystem services their activities destroy (though probably considerably less ignorant than city dwellers), but even if they were aware, their personal gains from deforestation almost certainly far outweigh their share of the loss from ecosystem services destroyed.

Much of the forest within these countries occurs in states or administrative districts that are almost entirely forested, and have very low population densities. It may thus also make sense at the local and regional level to continue deforesting. There is an important caveat, however. Even if continued deforestation may be appropriate at the local level, in many areas it is carried out in a destructive and inefficient manner. For example, where market access is poor, valuable timber and trees providing important nontimber resources (such as Brazil nuts) may be felled and burnt. Unsustainable production techniques that mine the soil are common, even when small investments could yield far more sustainable and lucrative alternatives. Thus, while deforestation may be appropriate from the local point of view, the way in which it is carried out may at the same time be highly inappropriate.

Even at the national level in countries like Suriname, Guyana, and French Guyana that are up to 90% forested, continued deforestation, if carried out without wanton waste, could help improve well-being for the majority of society (though the indigenous cultures that depend on the forest would almost certainly not be part of that majority). It can also strengthen claims to contested boundaries. In other countries where it may be in the national interest to slow or halt deforestation, it may still not be worth the investment of the resources that would be required to do so. Vast areas of intact forest make monitoring and enforcement of regulations intended to curb deforestation expensive and difficult. In fact, in many of these countries, large areas of forests are essentially open access resources. For example, an estimated 80% of timber harvests in Brazil are illegal.⁸

Brazil provides an interesting case study of a country in which the national government over the past two decades has moved from policies explicitly and vigorously promoting deforestation toward considerable

⁸C. Bright and A. Mattoon, The Restoration of a Hotspot Begins, *World Watch Magazine*, November/December, 2001.

legislation designed to reduce deforestation. Much of the national legislation designed to preserve forests is poorly enforced, and other national policies will almost certainly dramatically increase deforestation in some areas, but there is nonetheless a real trend toward greater efforts for preservation at the national level. This trend presumably reflects a growing knowledge of ecosystem services provided by the forest, growing marginal utility of remaining forests, and international pressure. Some of the more heavily deforested states have implemented innovative policies for reducing deforestation, while heavily forested states still promote it. In the Atlantic Forest, of which only some 8% remains, some individual landowners are working to protect their remaining forests, or are even actively restoring the forest. What are conspicuously absent are adequate resources from the global community to preserve the forest.

International Policies

Problems of global scale must ultimately be solved via global policies. What policies are available to the global community to limit deforestation and other forms of ecosystem destruction to a globally acceptable level? We previously discussed the polluter pays principle, but when it comes to deforestation and the destruction of coral reefs, wetlands, and other ecosystems (or greenhouse gas emissions), sovereign nations will not pay for the impacts their activities have on the rest of the world. Although all nations in the world benefit from healthy ecosystems in other countries, they do little or nothing to help pay for their preservation. Ecosystem services are global public goods, and most countries are free riders on the provision of those public goods.

One solution is the application of a “beneficiary pays principle,” where those who benefit pay for the benefits they receive. The fact is that less developed countries, which may now contain the bulk of the world’s most productive ecosystems, often lack the institutions and resources to limit ecosystem conversion to nationally desirable levels. An effective policy for preserving ecosystems at a globally desirable level must provide both incentives and resources for doing so.

International Subsidies for Ecosystem Preservation. One possibility for implementing a market-based “beneficiary pays principle” would be an international Pigouvian subsidy for providing ecosystem services. Take the Brazilian Amazon as a case study. Ideally, to reach point E on Figure 23.1, the global community would need to pay only quantity $D'-E'$ to get from the nationally optimal to the globally optimal level. The first problem is the transaction cost of getting the wealthy nations to agree to paying Brazil to reduce deforestation and deciding how much each must pay.

A potentially larger problem is deciding to whom the subsidy should be paid.

The Amazon is vast, and while deforestation is rapid, it still affects on average less than 1% of the region per year. It would not be efficient to pay all landowners not to deforest, as not all of them plan to deforest in the first place. One possibility would simply be to pay only farmers who are currently deforesting not to do so, but this presents numerous difficulties. Transaction costs of reaching agreements with individual farmers would be enormous, and the monitoring and enforcement costs necessary to ensure landowner compliance would also be substantial. If the farmer agrees not to deforest one area, he may simply deforest another area instead.

Asymmetric information presents another serious problem, as only the landowners themselves know how much they plan to deforest. If payments were made only to farmers in the process of deforesting, other farmers might begin to deforest simply to become entitled to payments. Subsidies directed to the landowners actually involved in deforestation could thus perversely increase the rate of deforestation.

A plausible alternative for a subsidy would be for the international community to pay Brazil for reducing the rate of deforestation below some predetermined baseline. The baseline might be average deforestation over the past several years, or expected deforestation estimated through a more sophisticated model, including variables such as rainfall and economic growth. For example, the most recent available 3-year average of deforestation rates in the Amazon is in the range of 2 million ha per year (average of 1995–1997⁹). The international community could pay a given amount for every hectare deforestation falls short of 2 million. If too much deforestation still occurs, the subsidy can be raised, while if the subsidy is too costly, it can be decreased. Deforesting the full 2 million ha would forego all subsidies, but there still would be no incentive to deforest beyond this point, because there is no incentive to do so in the absence of the subsidy.

Another alternative would be to adopt a strategy currently in use in Brazil known as the ICMS *ecológico*. The ICMS is a tax on merchandise and services, and in certain states, some of this money is refunded to municipalities according to the extent that they meet ecological goals, such as watershed protection and forest conservation. Essentially a payment for the provision of ecological services, it has proven quite effective.

There is no reason a similar approach could not be used at a global level. Initially perhaps, it could be directed toward biodiversity hot-spots,

⁹Instituto Nacional de Pesquisas Espaciais (INPE), Desflorestamento 1995–1997 Amazonia, 1998. Online: <http://www.inpe.br/amz-00.htm>.

25 areas round the world identified by scientists that contain an unusually large number of species and are seriously threatened, with 70% or more of the area destroyed. Given the importance of biodiversity in maintaining ecosystem resilience and function, it is likely that hot-spots offer an unusually large amount of ecosystem services. Similar to the ICMS *ecológico*, a global pool of money could be distributed to the countries that harbor these hot-spots according to how well they meet well-defined conservation criteria. It would then be up to the individual countries to determine how best to meet these criteria, thereby allowing micro-freedom to achieve macro-control.

There are several features that contribute toward the feasibility of such international subsidy schemes. First, transaction costs are minimized. Inexpensive satellite photos are capable of providing increasingly accurate estimates of annual deforestation, so monitoring costs would be small.¹⁰ While interpretation of photos may not be an exact science, computer analysis can at least make it a consistent one, in which case quantitative precision is unnecessary. This approach is currently being used to monitor compliance with land use laws in Mato Grosso, Brazil. A subsidy can thus reflect the amount of forest preserved, though an exact dollars-per-hectare figure for the subsidy may not be accurate.

Second, it would not be necessary at the international level to pinpoint who is deforesting. Enforcement and accountability are not major problems, since funds would only be dispersed after conservation occurs; if deforestation is not slowed, no money is spent. Third, national sovereignty would remain intact, as no country would be under any obligation to change behavior. Finally, a major problem in many less developed countries is that they lack the institutions and resources to enforce environmental policy, especially in such vast areas as the Amazon. A subsidy could provide both the incentive and resources for local and national governments to slow deforestation using some of the previously discussed policies.

A very similar approach would be to pair a national Pigouvian tax with an international subsidy. In this case, the national government would be required to tax deforestation. Such a tax might be costly to implement and administer, both fiscally and politically. While these costs might be less than the revenue gained, they still reduce the incentive to implement such a policy. If, however, a global institution matched such a tax, or some percentage of such a tax (perhaps greater than 100%), it would effectively combine the “polluter pays” and the “beneficiary pays” principles in one policy package.

¹⁰A. Almeida and C. Uhl, Brazil's Rural Land Tax: A Mechanism to Promote Sustainable Land Use in Amazonia, *Land Use and Policy* 12:105–114 (1995).

The problem with subsidies for firms to reduce pollution is that it might lead to more firms. If countries decide to grow forests simply to earn a subsidy not to cut them down, that would hardly be a problem.

Although these suggestions make economic sense, they have not been tried at an international level. This is why we call for adaptive management with all policies. When a policy works, use it. When it doesn't, either fix it or replace it with another. What we cannot afford to do is stand by and do nothing when it is obvious that the status quo is not working.

■ REDEFINING EFFICIENCY

In standard economic practice, allocative efficiency is achieved when we put scarce resources to the use that generates most monetary value (which is taken to be a measure of utility). With a central focus on monetary value, allocative efficiency generally ignores nonmarket goods and services. Typically economists defer to Pareto efficiency, which is an allocation such that nobody can be made better off without making someone else worse off. Pareto efficiency does not permit comparisons between individuals, and it accepts the status quo distribution of wealth, however, unequal that may be. It ignores the diminishing marginal utility of wealth and the potential for gains from redistribution.

Many economists and policy makers favor potential Pareto efficiency (Hicks-Kaldor welfare criterion, see Chapter 17) as an “objective” decision-making tool that favors any allocations that could potentially create a Pareto improvement after redistribution but that does not require that redistribution. Since wealth clearly generates more wealth in the modern economy, potential Pareto improvements are more likely to benefit the already wealthy than the poor.

Technical efficiency, in contrast, is defined as the maximum amount of physical output one can get from a given amount of resource input. While a desirable goal, it alone does little, if anything, to create a more sustainable society. Greater technical efficiency can reduce the demand for a resource. Alternatively, lowering the quantity of resource required to make something can lower the cost of the final product. Quite possibly, the result of increased technical efficiency, thanks to lower price, is greater use of resources, not less.

We have already seen that the goal of economics is not to maximize production, but rather to provide service. We define service as a psychic flux of satisfaction, which is derived from manmade capital as well as from ecosystem services provided directly by natural capital. Manmade capital can only be created through the transformation of natural capital, hence the production of services from manmade capital demands a sacrifice of services from natural capital. We will call this the **comprehensive effi-**

ciency identity. Therefore, an appropriate measure of efficiency is the ratio of services gained from manmade capital stock (MMK) to the services sacrificed from the natural capital stock (NK) as a result. There are several ways to improve this efficiency ratio, as shown in the following identity:

$$\frac{\text{MMK services gained}}{\text{NK services sacrificed}} = \frac{\text{MMK services gained}}{\text{MMK stock}} \times \frac{\text{MMK stock}}{\text{thruput}} \times \frac{\text{thruput}}{\text{NK stock}} \times \frac{\text{NK stock}}{\text{NK services sacrificed}}$$

(1) (2) (3) (4)

Ratio 1 is service efficiency; it is composed of technical design efficiency, allocation efficiency, and distribution efficiency. For example, a well-designed house provides more of the service of shelter than a poorly designed one using the same amount of material; alternatively, glue-laminated beams and laminated I-beam floor joists use less wood to provide the service of structural strength than traditional one-piece solid wood building materials. Allocation efficiency requires that black walnut be used to build fine furniture instead of floor joists. As to distributive efficiency, wood used to provide essential shelter to 50 homeless families provides more service than the same wood used to build a rarely used summer mansion for a billionaire.

Ratio 2 is maintenance efficiency or durability. All MMK stock requires throughput to maintain or replace it, but the less throughput required, the greater the efficiency. A well-built house lasts longer and requires less maintenance than one sloppily slapped together.

Ratio 3 is growth efficiency of natural capital and harvest efficiency. Well-managed forests and plantations of fast-growing species provide more sustainably harvested timber each year than poorly managed forests or plantations consisting of slowly growing species. For example, studies in the Amazon show that carefully selecting trees to be cut, removing vines on those trees, and carefully planning skid tracks can reduce the time between harvests from 90 to 30 years.

Ratio 4 is increased by creating more natural capital stock, or by sacrificing fewer ecosystem services per unit of stock we exploit. Reforestation increases the stock of forest. While a timber plantation may be efficient in terms of growth and timber production, it may provide few other ecosystem services. In contrast, improved management of selective logging of existing forests, as described above, is likely to increase both efficiency ratios 3 and 4.

This definition addresses scale by capturing the tradeoff between services gained (numerator) and services lost (denominator), as shown on the left-hand side of the identity. Uneconomic growth invariably reduces efficiency. On the right-hand side of the identity we see the components of overall efficiency—namely, design, distribution, durability, growth, and harvesting.

THINK ABOUT IT!

Analyze the efficiency of burning coal from a strip mine using the comprehensive efficiency identity. Consider each of the four ratios in your answer. (Hint: MMK stock is inventory of mined coal, NK is coal in the ground.)

BIG IDEAS to remember

-
- | | |
|---|--|
| ■ Pricing nonmarket goods and services | ■ Subsidies for nonmarket goods |
| ■ Contingent valuation and pricing nature | ■ Seigniorage |
| ■ Market vs. nonmarket values | ■ International subsidies for ecosystem preservation |
| ■ Macro-allocation asymmetric information | ■ Comprehensive efficiency identity |
| ■ Asymmetric information | |
-

Looking Ahead

This book has focused on three issues—the allocation of resources, the distribution of income, and the scale of the economy relative to the ecosystem—with special emphasis on the third. A good allocation of resources is efficient; a good distribution of income or wealth is just; a good scale is at least ecologically sustainable.

Allocation and distribution are familiar concepts from standard economics—for any given distribution of income, there is a different optimal allocation of resources with its corresponding optimal set of prices. Standard economics focuses primarily on the allocation issue, paying secondary attention to distribution, first because a given distribution is logically necessary for defining efficient allocation, and second because distributive fairness is important in its own right.

The third issue of scale, the physical size of the economy relative to the containing ecosystem, is not recognized in standard economics and has therefore become the differentiating focus of ecological economics. The preanalytic vision of the economy as an open subsystem of a larger ecosystem that is finite, nongrowing, and materially closed (though open with respect to solar energy) immediately suggests three analytical questions: How large is the economic subsystem relative to the containing ecosystem? How large *can* it be? How large *should* it be? Is there an optimal scale beyond which physical growth in the economic subsystem begins to cost more at the margin than it is worth in terms of human welfare? This text has tried to explain the reasons for an affirmative answer to this last question.

If the economy grew into the void, it would encroach on nothing, and its growth would have no opportunity cost. But since the economy in fact grows into and encroaches upon the finite and nongrowing ecosystem, there is an opportunity cost to growth in scale, as well as a benefit. The costs arise from the fact that the physical economy, like an animal, is a “dissipative structure” sustained by a metabolic flow from and back to the environment. This flow, which we have called throughput (adopting the term from engineers) begins with the depletion of low-entropy useful resources from the environment, is followed by the processes of production and consumption, which, despite the connotations of the words, are only physical transformations, and ends with the return of an equal quantity of high-entropy polluting wastes.

Depletion and pollution are costs. Not only does the growing economy encroach spatially and quantitatively on the ecosystem, it also

qualitatively degrades the environmental sources and sinks of the metabolic throughput by which it is maintained. This forces a continual co-evolutionary adaptation between the economy and the ecosystem. If that adaptation is made in such a way that the throughput remains within the natural capacity of the ecosystem to absorb wastes and regenerate resources, then the scale of the economy is considered “sustainable.” Non-renewable resources, of course, cannot be exploited in a sustained yield manner, but we discussed some rules of thumb for “quasi-sustainable” exploitation, along with the analysis of sustained yield exploitation of renewable resources.

From a policy perspective, we have insisted that optimal allocative prices do not guarantee a sustainable scale any more than they guarantee a just distribution of income. Attaining a sustainable scale, a just distribution, and an efficient allocation are three distinct problems. They are certainly not isolated, but solving one does not solve the others. Achieving three different goals generally requires three different policy instruments. This is illustrated by the cap-and-trade systems, a favored policy of ecological economists. Three policy actions are required in proper sequence. First, a quantitative limit is set reflecting judgments of sustainable scale—that is, a previously unlimited or free good is recognized as scarce and the scale of its use is limited. Second, the newly scarce good or right is now a valuable asset—who owns it? Deciding who owns it is a question of distributive justice. Third, once scale and distribution decisions have been politically made, we can have individualistic trading and efficient market allocation.

As growth pushes us from an empty world to a full world, the limiting factor in production, as we have argued, increasingly becomes natural capital, not manmade capital—for example, the fish catch today is no longer limited by manmade capital of fishing boats but by the complementary natural capital of fish populations in the sea. As we move into a full world, economic logic remains the same—to economize on and invest in the limiting factor. But the identity of the limiting factor changes from manmade to remaining natural capital, and our economizing efforts and policies must change accordingly. Therefore, it becomes more important to study the nature of environmental goods and services in both their stock-flow and fund-service dimensions—are they rival or nonrival, excludable or nonexcludable—in order to know if they are market goods or public goods.

Ecological economics accepts the standard analysis of allocative efficiency, given prior social determination of the distribution and scale questions, and given that the good in question is rival and excludable. Although the main difference has been the focus on scale, that difference has entailed more attention to dimensions of distribution often

neglected—namely, intergenerational distribution of the resource base and distribution of places in the sun between humans and all other species (biodiversity). Also, as more vital resources cease being free goods and are allocated by the market, the fairness of the distribution underlying market allocation becomes more critical. Once growth in scale has become uneconomic, it can no longer be appealed to as the solution to poverty. Poverty reduction requires increased sharing. Other issues of debate include whether natural and manmade capital are primarily substitutes or complements, the degree of coupling between physical throughput and GNP, and the degree of coupling between GNP and welfare.

One question not explicitly addressed in the text, but which students are sure to ask, is: What is the relationship between ecological economics and courses in resource economics or environmental economics that are sometimes taught in economics departments? The difference is that the latter are both subfields of neoclassical economics; they do not consider scale an issue, have no concept of throughput, and are focused on efficiency of allocation. Resource economics deals with the efficiency of allocation of labor and capital devoted to extractive industries. It develops many useful concepts that we have covered in this text, such as scarcity rent and user cost. Likewise, environmental economics also focuses on efficiency of allocation and how it is disrupted by pollution externalities. Concepts of internalizing externalities by Pigouvian taxes or well-defined and enforceable property rights (see Coase Theorem, Chapter 10) are certainly useful, and we have discussed them. Nevertheless, the aim of both resource and environmental economics is allocative efficiency via right prices, not sustainable scale.

Ecological economics connects resource and environmental economics by recognizing the real-world connection between depletion with pollution via the concept of throughput. We have also paid much more attention to impacts on, and feedbacks from, the rest of the ecosystem induced by economic activities that cause depletion, pollution, and entropic degradation. In addition, we have investigated the basic principles (energy flows, material cycles, ecosystem structure and function) governing the containing ecosystem itself, thereby at least partially integrating economics with ecology.

Finally, we have insisted on policy as our guiding philosophical viewpoint. This has led us to recognize and defend the logical presuppositions of policy—namely, nondeterminism and non-nihilism. It really is possible for things to be other than they are, and we really can distinguish better from worse states of the world. If that were not the case, then our effort in writing this book, and your effort in reading it, would both have been in vain.

Glossary

Abiotic resource A nonliving resource that cannot reproduce: fossil fuels, minerals, water, land, and solar energy.

Absolute advantage A country has an absolute advantage if it can produce the good in question at a lower absolute cost than its trading partners. It has a **comparative advantage** if it can produce the good in question more cheaply relative to other goods it produces than can its trading partners, regardless of absolute costs.

Absorptive capacity See “waste absorption capacity.”

Adaptive management A basic policy principle whereby we change policies as conditions change or as we gain new information.

Aggregate macroeconomics The study of the economy from the perspective of key aggregate variables such as the money supply, aggregate price level, the interest rate, aggregate consumption and investment, exports and imports. The main focus of conventional economics is the rate of growth of GNP. In contrast, ecological economics strives to end physical growth while maintaining or improving social welfare.

Allocation The process of apportioning resources to the production of different goods and services. Neoclassical economics focuses on the market as the mechanism of allocation. Ecological economics recognizes that the market is only one possible mechanism for allocation.

Asymmetric information Occurs when either buyer or seller has information that the other does not have, and that information affects the value of the good or service exchanged.

Balance of payments The sum of the current account (exports minus imports), and the capital account (inflow of capital to the nation minus outflow of capital from the nation).

Barter The direct exchange of goods or services without using money as a medium of exchange. It is very inconvenient in that it requires an unlikely coincidence of reciprocal wants.

Basic market equation $MU_x/MU_y = P_x/P_y = MPP_y/MPP_x$, where MU is the marginal utility of good x or good y to person n, and MPP is the marginal physical product of factor a used to produce good x or good y.

Biotic resource A living resource, such as trees, fish, and cattle (elements of ecosystem structure), as well as any of the fund-services they provide, such as climate regulation, water regulation, and waste-absorption capacity (ecosystem functions, or ecosystem services).

Bretton Woods Institutions (IMF and World Bank) Global financial institutions created in 1945 to finance short-term international trade (International Monetary Fund), and to lend for long-term investment projects in developing countries (World Bank).

Capital account A measure of inflows of investments to the nation by foreigners, and outflows of investments by nationals to foreign countries.

Carrying capacity Originally the maximum population of cattle that can be sustained on a given area of rangeland. By extension the population of humans that can be sustained by a given ecosystem at a given level of consumption, with a given technology.

Catch-per-unit-effort hypothesis The assumption of a linear relationship between effort, stock, and harvest.

Circular flow The idea that since every expenditure by anyone is at the same time an equal receipt by someone else and receipts in turn become expenditures, money or exchange value flows in a circle. But physical factors and products do not flow in a circle.

Closed system A system that imports and exports energy only, matter circulates within the system but does not flow through it. The Earth closely approximates a closed system.

Coase theorem States that in a perfectly competitive market allocative efficiency will be achieved whether property rights are given to the polluter or the “pollutee.” All that is required is that someone have the property rights and that transaction costs are zero.

Coevolutionary economics The study of the mutual adaptations of economy and environment. Economic activity induces changes in the environment, and changes in the environment in turn induce further changes in the economy in a continuing process of coevolution.

Command-and-control regulation Flat prohibitions, quotas, or standards as opposed to monetary incentives that operate through prices or taxes.

Comparative advantage See “absolute advantage.”

Competitive market A market in which there are many small buyers and sellers of an identical product. “Many” means “enough that no single buyer or seller is sufficiently large to affect the market price.” Since everyone treats price as a parameter (a given condition) rather than a variable (something one can change), this condition is sometimes called the parametric function of prices.

Complementarity The “opposite” of substitutability—when goods or factors have to be used together in fairly strict combination with each other rather than instead of each other. Even substitutes have some degree of complementarity, unless they are “perfect” substitutes, in which case they are for all practical purposes identical goods or factors.

Comprehensive efficiency The ratio of services gained from manmade capital stock to the services sacrificed from the natural capital stock.

Congestibility Occurs when a basically nonrival asset is used so heavily that one person’s use begins to interfere with or lower the quality of service to other users (e.g., a crowded road or interference from another radio station in a crowded frequency band).

Consumer surplus The maximum that a consumer would be willing to pay for all units of a good he consumes rather than do without the good, minus the amount he actually has to pay for it.

Contingent valuation Hypothetical estimates of prices of nonmarket goods and services based on survey questions asking how much one would be willing to pay for an extra unit of the good, or how much one would accept for the loss of a unit of the good.

Critical depensation The population size below which a population will likely go extinct rather than recuperate, even if exploitation ceases. Also known as “minimum viable population.”

Current account A measure of the international exchange of real goods and services as well as transfer payments in the current year.

Defensive expenditure An expenditure made to protect one against the unwanted consequences of the production and consumption of other goods by other people. Also called regrettably necessary defensive expenditure.

Demand A relationship (inverse) between price of a good and the quantity of the good that consumers would purchase at that price.

Determinism The philosophical doctrine that every event or decision is the inevitable consequence of antecedents, such as physical, psychological, hereditary, or environmental conditions, that are independent of human will or purpose.

Development The improvement in quality of goods and services, as defined by their ability to increase human well-being, provided by a given throughput.

Differential rent See “rent.”

Diminishing marginal physical product See “law of diminishing marginal physical product.”

Diminishing marginal utility See “law of diminishing marginal utility.”

Discount rate The rate at which the present is valued over the future, as a result of uncertainty, or of productivity, or of pure time preference for the present. See “intertemporal discounting.”

Disinflation A reduction in the rate of general price increase. See “inflation.”

Distribution The apportionment of income or wealth among different people.

Doubling time The fixed time period it takes for a population to double when growing at a constant rate. A handy rule of thumb is that doubling time equals 70 divided by the percentage rate of growth (e.g., a population growing at 2% annually will double every 35 years).

Ecological economics The union of economics and ecology, with the economy conceived as a subsystem of the earth ecosystem that is sustained by a metabolic flow or “throughput” from and back to the larger system. See “throughput.”

Ecological reductionism The idea that the human economy is entirely governed by the same laws and forces as the rest of the ecosystem, so there is no necessity to distinguish the human economy as a subsystem.

Economic imperialism The idea that the entire ecosystem can be priced and all values internalized into prices, with the result that price calculations are reliable guides for all decisions.

Ecosystem function An emergent phenomenon in ecosystems, such as energy transfer, nutrient cycling, gas regulation, climate regulation, and the water cycle. As is typical of emergent properties, ecosystem functions cannot be readily

explained by even the most extensive knowledge of system components or ecosystem structure.

Ecosystem services Ecosystem functions of value to humans, though given the tightly interconnected nature of ecosystems, it would be difficult to say with certainty that any particular ecosystem function is not of value to humans. See also “fund-service resources.”

Ecosystem structure The individuals and communities of plants and animals of which an ecosystem is composed, their age and spatial distribution, and the abiotic resources present. The elements of ecosystem structure interact to create ecosystem functions as emergent properties generated of such a large complex system.

Efficient allocation See “Pareto Efficient Allocation.”

Efficient cause An agent of transformation, such as labor or a machine. See also “fund-service resources” and “material cause.”

Elasticity The responsiveness of a change in quantity demanded (supplied) to a change in price—measured by percentage change in quantity divided by percentage change in price.

Entropic dissipation The gradual erosion and dispersion into the environment of the matter of which all human artifacts are composed in a one-way flow of low entropy usefulness to high entropy waste.

Entropy See “Second Law of Thermodynamics.”

Environmental economics The branch of neoclassical economics that addresses environmental problems such as pollution, negative externalities, and valuation of nonmarket environmental services. In general, environmental economics focuses almost exclusively on efficient allocation, and accepts the pre-analytic vision of neoclassical economics that the economic system is the whole, and not a subsystem of the containing and sustaining global ecosystem.

Equimarginal principle of maximization This is known as the “when to stop” rule. The point at which a consumer reaches an allocation that maximizes her total satisfaction or total utility. That point occurs when the marginal utility per dollar spent on each good is equal. Only when utilities were equal at the margin would it no longer be possible to increase total utility by reallocation of expenditure.

Exchange rate The rate at which one nation’s currency is traded for that of another nation. Exchange rates can be either fixed by central banks, floating according to daily supply and demand, or some combination of the two.

Exchange value The value of a good in terms of its ability to be traded for other goods, as opposed to its use value.

Excludability A legal concept that when enforced allows an owner to exclude others from using his asset. An institution is always required to make an asset excludable, but some by their very nature are nonexcludable (e.g., the ozone layer). If the asset is rival (a physical property), then excludability is more or less natural. If the asset is nonrival, then excludability is typically more difficult, but sometimes possible, as in the case of patents.

Exponential growth Growth at a constant percentage rate (fixed doubling time). Exponential growth leads to very large numbers surprisingly quickly.

Externality An unintended and uncompensated loss or gain in the welfare of one party resulting from an activity by another party.

Fallacy of composition The argument that what is true for the part must necessarily be true for the whole, or vice versa.

Fallacy of misplaced concreteness To mistake the map for the territory, to be unmindful of the degree of abstraction in an argument, especially to draw conclusions at a level of abstraction (or concreteness) different from the level of abstraction of the concepts in which the argument is conducted.

Federal Reserve System A coordinated system of district central banks in the U.S. that influences interest rates and money supply by means of open-market operations, discount rate changes, and reserve ratio requirements.

First Law of Thermodynamics Neither matter nor energy can be either created or destroyed.

Fiscal policy The attempt to influence GNP, employment, interest rates, and inflation by manipulating government expenditure and taxes.

Fixed exchange rate regime A regime in which the value of one country's currency is pegged to another country's (typically the US dollar).

Flexible exchange rate regime A regime in which exchange rates are determined by the global supply and demand for currencies, and central banks play no direct role.

Floating exchange rate regime See "flexible exchange rate regime."

Fossil fuels Petroleum, coal, natural gas. Fuels formed over geologic ages from biotic materials, but now treated as nonrenewable abiotic resources.

Fractional reserve banking The practice of keeping on hand reserves against deposits that are only a small fraction of deposits, so that banks can lend the difference. This practice allows the private banking sector to create money, since demand deposits are counted as money.

Free rider One who enjoys the benefit of a public good without paying a share of the cost of its provision and maintenance.

Fund-service resources Resources not materially transformed into what they produce (efficient cause); which can only be used at a given rate, and their productivity is measured as output per unit of time; cannot be stockpiled; and are worn out, rather than used up. See "stock-flow resources."

General equilibrium model The vision of the economy as a giant system of thousands of simultaneous equations balancing the supply and demand, and determining the price and quantity, for each commodity in the economy.

Gini coefficient A measure of the inequality of the distribution of wealth or income across a population. A Gini coefficient of one implies perfect inequality (one person owns everything) and a coefficient of zero indicates a perfectly equal distribution.

Globalization The economic integration of the globe by free trade, free capital mobility, and to a lesser extent by easy migration. It is the effective erasure of national boundaries for economic purposes. See "internationalization."

Gross national product (GNP) The market value of final goods and services purchased by households, by government, and by foreigners (net of what we

purchase from them), in the current year. Alternatively, it is the sum of all value added to raw materials by labor and capital at each stage of production, during the given year.

Growth A quantitative increase in size, or an increase in throughput.

Hotelling rule States that at the margin the rate of return from holding a resource in the ground (its expected price increase) must be equal to the rate of return from exploiting it now and investing the profits.

Hubbert curve A curve showing the cumulative extraction of a nonrenewable resource over time. The vertical distance for each year is annual extraction for that year. The total area under the curve is total reserves. The typical curve is bell-shaped, rising from zero, reaching a maximum, and falling to zero again with exhaustion.

Human needs assessment A multidimensional concept of welfare that goes beyond income and wealth to include capabilities, capacities, and other existential categories.

Hyperbolic discounting The act of giving more weight to what happens now over what happens in the near future, while being nearly indifferent between the same outcome occurring at different times in the more distant future. Empirical studies suggest that this is a far more accurate representation of the human psyche than exponential discounting.

Hyperinflation Inflation greater than 50% per month.

Inalienability rule An entitlement that holds if a person is entitled to either the presence or absence of something, and no one is allowed to take away that right for any reason.

Income The maximum that a community could consume in a given time period and still be able to produce the same amount in the next time period. In other words, the maximum that can be consumed without reducing productive capacity, that is, without reducing capital.

Individual transferrable quotas (ITQs) See “transferable quotas.”

Inflation An increasing general level of prices (not a state of high prices, but a process of increasing prices).

Internationalization The increasing importance of relations among nations, and among citizens of different nations. The nation remains the basic unit of community and policy, controlling to some extent trade, capital flows, and migration. National economies are interdependent, but not integrated. See “globalization.”

International Bank for Reconstruction and Development (World Bank) An international financial institution composed of member nations and created at Bretton Woods, New Hampshire, in 1945. Originally designed to focus on long-term lending to promote the development of underdeveloped countries, in recent decades it has strayed from its charter.

International Monetary Fund (IMF) An international financial institution composed of member nations and created at Bretton Woods, New Hampshire, in 1945. Originally designed to focus on short-term balance of payments financing to promote international economic stability, in recent decades it has strayed from its charter. See also “World Bank.”

Intertemporal allocation The apportionment of resources across different stages in the lifetimes of basically the same set of people (same generation).

Intertemporal discounting The process of systematically weighting future costs and benefits as less valuable than present ones.

Intertemporal distribution The apportionment of resources across different generations (different people).

ISEW Index of Sustainable Economic Welfare—calculated by adjusting Personal Consumption for various factors that affect either sustainability or welfare either positively or negatively, such as depletion of natural capital, increasing inequality in income distribution, defensive expenditures, and so on.

IS-LM model A two-sector general equilibrium model showing how the real and the financial sectors interact to simultaneously determine the national income and the interest rate.

Isolated system An isolated system is one in which neither matter nor energy enters or exits.

Law of diminishing marginal physical product As a producer adds successive units of a variable factor to a production process, other factors constant, then the extra output per unit of the variable factor diminishes with each addition (i.e., total output increases at a decreasing rate). This is sometimes called the law of diminishing returns.

Law of diminishing marginal utility As one consumes successive units of a good the additional satisfaction decreases (i.e., total satisfaction increases, but at a decreasing rate). With a law of constant or increasing marginal utility consumers would spend all their income on only one good.

Law of entropy See “Second Law of Thermodynamics.”

Law of increasing marginal cost As one produces more and more of a product, one must use resources (factors of production) that are of lower quality or are more expensive, so that the cost of producing each additional unit is greater than that of the previous unit.

Liability rule An entitlement rule that holds if one person is free to interfere with another or prevent interference, but must pay compensation.

Linear throughput See “throughput.”

Liquidity preference A general preference for holding assets in a form easily convertible into money, so as to meet unexpected transaction needs and avoid the inconvenience of barter.

Liquidity trap The failure of lowering interest rates to stimulate economies with low demand.

Lorenz curve A curve plotting the cumulative percentage of the population against the cumulative percentage of total income held by that percentage of the population—illustrates the degree of equality/inequality in the distribution of income.

Macro-allocation The allocation of resources between market and nonmarket goods and services.

Marginal cost The increment in total cost resulting from producing one more unit of the commodity in question.

Marginal external cost The cost to society of the negative externality that results from one more “unit” of activity by an economic agent.

Marginal extraction costs The extra total cost required to extract one more unit of a resource from the ground.

Marginal revenue The increment in total revenue from selling one more unit of the commodity in question.

Marginal user cost The value of one more unit of the resource in its natural state. In a perfectly competitive economy, marginal user cost would in theory equal the price of a resource minus its marginal extraction cost.

Marginal utility The additional pleasure or satisfaction to be gained from consuming one more unit of a good or service.

Material cause A resource that is transformed in the production process. See also “stock flow resource” and “efficient cause.”

Materialism The philosophical doctrine that physical matter (its movements and modifications), is the only reality, and that everything in the universe, including thought, feeling, mind, will, and purpose can be explained in terms of physical laws.

Matrix of human needs (Max-Neef) A cross-classification of basic dimensions of existence with basic human values, yielding a much richer and more detailed description of welfare than the abstract notion of “utility.”

Maximum sustainable yield Each level of an exploited population has a growth rate that can be harvested leaving the population undiminished in the following year. There is one level of population for which the sustainable yield is a maximum. In general, however, the biologically maximum sustainable yield is not the economically optimal yield.

Micro-allocation The allocation of resources within the private sector, as opposed to between the private and public sectors. See “macro-allocation.”

Mineral resource A useful element or compound, such as copper, iron, petroleum—a class of abiotic resource.

Minimum viable population The population level below which a population is not likely to recuperate, but rather will dwindle to extinction. See “critical depensation.”

Monetary policy The attempt to influence interest rates, GNP, employment, and inflation by manipulation of the money supply.

Money A unit of account, medium of exchange, and a store of value. Money can be a commodity (gold) whose supply is limited by its real cost of production, or a token (fiat money) whose supply is determined by government authority and social conventions.

Monopoly A single seller of a commodity.

Moral hazard A situation in which the existence of insurance against a hazard actually increases the risk of the hazard by making the insured less vigilant in its prevention.

Multi-tier pricing The act of charging different prices at different times or for different users.

Natural capital Stocks or funds provided by nature (biotic or abiotic) that yield a valuable flow into the future of either natural resources or natural services.

Natural dividend The unearned income from the harvest of renewable resources. As nature and not human industry produces renewable resources, all profits above “normal” profit (included in the total cost) are unearned, and the natural dividend is equivalent to the total return minus the total cost.

Neoclassical economics The currently dominant school of economics, characterized by its marginal utility theory of value, its devotion to the general equilibrium model stated mathematically, its individualism and reliance on free markets and the invisible hand as the best means of allocating resources, with a consequent downplaying of the role of government.

Net present value The amount of money if available today that would generate the future stream of net income in question.

Nihilism In ethics the rejection of all distinctions in moral value, the rejection of all theories of morality. The view that nothing is better than anything else, or basically “anything goes.”

Nonexcludable resource A resource for which no institution or technology exists to make it excludable.

Nonmarket value A value recognized by people but not usually expressed in prices because the valuable thing either is not currently, or cannot conceivably, be traded in markets.

Non-price adjustments The adjustment of the relative desirability of goods by advertising, or the relative possibility of producing goods by research and development. These are ways to satisfy the “basic market equation” without adjusting prices.

Nonrenewable resource Low entropy matter-energy useful to humans and present in fixed stocks whose quantity declines over time. This includes mineral resources, fossil fuels, and fossil aquifers. As fresh water is naturally recycled through the hydrological process, we do not classify it as a nonrenewable resource.

Nonrival resource A resource whose use by one person does not affect its use by another.

Normal profit The opportunity cost of the time and money the entrepreneur has put into an enterprise (i.e., what she could have earned from her time and money in her next best alternative).

Open access A rival resource that all are free to exploit, no one can be excluded, such as noncoastal fisheries.

Open system An open system takes in and gives out both matter and energy. The economy is such a system.

Opportunity cost The best alternative given up when a choice is made, i.e., if a farmer cuts down a forest to expand his cropland, and if the consequent loss of timber, firewood, and water purification is the next best use of the land, then the value of timber, firewood, and water purification is the opportunity cost of the expanded cropland.

Optimal scale of the macroeconomy Occurs when the increasing marginal social and environmental cost of further expansion are equal to the declining marginal benefits of the extra production. Beyond the optimal scale growth becomes

uneconomic, even if we conventionally refer to the expansion of the economy as “economic growth.”

Paradox of thrift Occurs when everyone trying to increase their savings rate in the aggregate leads to declining consumption, growing unemployment and lower aggregate income. People are saving a higher percentage of their income, after increasing their savings rate, but total income has fallen, so that actual savings are less than before.

Pareto efficient allocation Occurs when no other allocation could make at least one person better off without making anyone else worse off. This is also known as a Pareto optimum.

Pareto optimum See “Pareto efficient allocation.”

Pigouvian subsidy See “Pigouvian tax.”

Pigouvian tax A tax designed to equal the marginal external cost of production of a commodity. It is added to the price, which measures only marginal private costs. The price plus tax now measures marginal social cost, thus internalizing the original external cost. If there originally was an external benefit, then a Pigouvian subsidy would be paid to the producer.

Principle of subsidiarity A basic policy principle that the domain of the policy-making unit must be congruent with the domain of the causes and effects of the problem with which the policy deals.

Principle of substitution The assumption that one good or service (or factor of production) can replace another in providing consumer utility (or in the production process). Theoretical utility functions and production functions often exhibit this property, but in reality, many goods, services, and factors of production are complements rather than substitutes. See also “substitutability” and “complementarity.”

Producer surplus The difference between the price (equal to marginal cost) at which the producer sells his total output and the lower marginal costs at which all inframarginal units were produced.

Production function A production “recipe” describing how certain quantities of inputs are combined to yield a certain quantity of output or product.

Property rule An entitlement rule that holds if one person is free to interfere with another, or free to prevent interference.

Public bad Something that is nonrival, nonexcludable, and undesirable.

Public good A resource that is nonrival, nonexcludable, and desired by the public. Because they are nonexcludable, they will not be produced by profit-seeking firms. Because it is nonrival, the marginal cost of another person using one is zero, so its efficient price should also be zero. A public good should be supplied collectively by the government or other social institution.

Pure time rate of preference The rate at which we prefer goods in the present over the future, independently of considerations of productivity.

Renewable resource A living resource that is capable of regeneration and growth in perpetuity if exploited in a sustainable manner and that provides raw materials for the economic process.

Rent A payment over and above minimum necessary supply price (cost of production). Since land has zero cost of production all payment for land is rent. Part

of payment for labor may also be rent if the laborer would still do the job for less. If Tiger Woods would still play golf even if he earned only a million dollars a year, then all his earnings over a million dollars is rent. Producer surplus is also an example of rent.

Ricardian land Land as an extension, surface area, substrate for holding things (i.e., the “indestructible” characteristics of land, excluding its fertility, or underground minerals).

Risk The known probability (relative frequency) of occurrence of an event. Risk is insurable. See “uncertainty.”

Rivalness An inherent characteristic of certain resources whereby consumption or use by one person reduces the amount available for everyone else.

Royalty The payment to the owner of a resource for the right to exploit that resource. Theoretically, in a competitive market, the per-unit royalty should be equal to the marginal user cost.

Say’s “Law” Supply creates its own demand. In production the payments to factors, plus residual profit, generates exactly enough income to purchase, if spent, the total amount produced.

Scale The physical size of the economic subsystem relative to the ecosystem that contains and sustains it. It could be measured in its stock dimension of population and inventory of artifacts, or in its flow dimension of throughput required to maintain the stocks.

Scarcity rent See “rent.”

Second Law of Thermodynamics Entropy never decreases in an isolated system. Although matter and energy are constant in quantity (first law), they change in quality. The measure of quality is entropy, and basically it is a physical measure of the degree of “used-up-ness” or randomization of the structure or capacity of matter or energy to be useful to us. Entropy increases in an isolated system. We assume the universe is an isolated system, so the second law says that the natural, default tendency of the universe is “shuffling” rather than “sorting.” In everyday terms, left to themselves, things tend to get mixed up and scattered. Sorting does not occur by itself.

Seigniorage The benefit that accrues to the issuer of token money, resulting from the fact that the issuer receives real goods and services in exchange for a mere token, whereas everyone else has to give up a real asset to get money to exchange for another asset. There is a transfer of real wealth from the public to the issuer of money, equal to the exchange value of money stock minus the cost of production of the tokens (negligible). That amount is seigniorage. Sometimes the term is used to refer only to the interest that could be earned on such an amount.

Sink That part of the environment that receives the waste flow of the throughput and may, if not overwhelmed, be able to regenerate the waste through biogeochemical cycles back to usable sources.

Social discount rate A rate of conversion of future value to present value that reflects society’s collective ethical judgment, as opposed to an individualistic judgment such as the market rate of interest.

Solar energy Radiant energy flowing from the sun, our basic long-run source of low entropy that sustains life and wealth.

Source That part of the environment that supplies usable raw materials that constitute the throughput by which the economy produces, and which ultimately returns as waste to environmental sinks.

Steady-state economy The economy viewed as a subsystem in dynamic equilibrium with the parent ecosystem/biosphere that sustains it. Quantitative growth is replaced by qualitative development or improvement as the basic goal.

Steady-state subsystem See “steady-state economy.”

Stock-flow resources Resources materially transformed into what they produce (material cause); can be used at virtually any rate desired (subject to the availability of fund-service resources required for their transformation); their productivity is measured by the number of physical units of the product into which they are transformed; can be stock-piled; are used up, rather than worn out. See “fund-flow resources.”

Subsidy A bonus or payment for doing something, the opposite of a tax.

Substitutability The capacity of a one factor (or good) to be used in the place of another, the opposite of “complementarity.” Substitutability is never perfect and the further a substitution is carried the less satisfactory it becomes (the more the force of complementarity is felt). Goods and factors may be thought of as varying in a continuum from perfect substitutes to perfect complements.

Supply The relationship between the price of a commodity and the quantity that would be supplied at each price.

Sustainable yield The amount of an exploited population that can be harvested leaving the population undiminished in the following year; the growth rate of the existing stock. See “Maximum sustainable yield.”

Thermodynamics The branch of physics that tells us that matter and energy can be neither created nor destroyed, and that entropy in the total system always increases. This branch of physics is the most relevant to economics because it helps to explain the physical roots of scarcity.

Throughput The flow of raw materials and energy from the global ecosystem's sources of low entropy (mines, wells, fisheries, croplands), through the economy, and back to the global ecosystem's sinks for high entropy wastes (atmosphere, oceans, dumps).

Total allowable commercial catch (TACC) An aggregate quota limiting the total fish catch per year or season. The quota can be distributed among individuals in many different ways.

Tradeable permits (quotas) Shares of an aggregate quota that are in some way divided up among individuals, who can then buy and sell their quota rights among themselves.

Transaction cost The costs of making a transaction, including legal fees, the cost of gathering information, locating the interested parties, the time costs of bargaining, and so on.

Transaction demand for money The demand for money balances to carry out everyday plus unforeseen transactions and avoid the inconvenience of barter or the delay of converting a nonliquid asset into money.

Ultimate end The vaguely perceived, yet logically necessary ordering principle with reference to which we rank our intermediate ends.

Ultimate means The low-entropy matter/energy, consisting of the solar flow and the terrestrial stock—that which we require to serve our ends and which we cannot ourselves create but can only receive from nature.

Uncertainty A situation in which we may know the range of possible outcomes but do not know the probability distribution of outcomes. Uncertainty is uninsurable. See “risk.”

Uneconomic growth Growth of the macroeconomy that costs us more than it is worth. A situation in which further expansion entails lost ecosystem services that are worth more than the extra production benefits of the expanded economy.

Unemployment Refers usually to involuntary unemployment, the number or percentage of the workforce without a job who are actively looking for a job.

Use value The actual service or utility from using a commodity for its intended purpose, as opposed to its exchange value, its capacity to purchase us another good through exchange.

User cost The opportunity cost of nonavailability of a natural resource at a future date that results from using up the resource today rather than keeping it in its natural state.

Utility function A psychic relationship showing the amount of utility or satisfaction yielded to a consumer by the consumption of differing amounts and combinations of commodities.

Virtual wealth A concept introduced by Frederick Soddy, similar to “seigniorage.” The total value of real assets that the community voluntarily abstains from holding in order to hold money instead. Since individuals can easily convert their money into real assets they count their money holdings as wealth. Yet the community as a whole cannot convert money into wealth because someone has to end up holding the money (“fallacy of composition”). Money wealth is therefore “virtual.”

Waste absorption capacity The capacity of an ecosystem to absorb and reconstitute wastes into usable forms through biogeochemical cycles powered by the sun. This capacity is a renewable resource that can be overwhelmed and destroyed, or used within sustainable limits.

Welfare A psychic state of want satisfaction or enjoyment of life—an experience, not a thing—the basic reason to be of economic activity.

World Bank See “International Bank for Reconstruction and Development.”

World Trade Organization (WTO) The successor organization to the General Agreement on Trade and Tariffs (GATT) that seeks liberalization of both international trade and investment and generally promotes globalization.

Suggested Readings

Part I. An Introduction to Ecological Economics

- Costanza, Robert (ed.). 1991. *Ecological Economics: The Science and Management of Sustainability*. New York: Columbia University Press.
- Costanza, Robert, Charles Perrings, Cutler J. Cleveland (eds.). 1997. *The Development of Ecological Economics*. Cheltenham, U.K.: Edward Elgar.
- Daly, Herman, and K. Townsend (eds.). 1993. *Valuing the Earth: Economics, Ecology, Ethics*. Cambridge, MA: MIT Press.
- Faber, Malte, Reiner Manstetten, and John Proops. 1998. *Ecological Economics: Concepts and Methods*. Cheltenham, U.K.: Edward Elgar.
- Gowdy, John. 1994. *Coevolutionary Economics: Economy, Society, and Environment*. Boston: Kluwer Academic Free Press.
- Martinez-Alier, Juan. 1987. *Ecological Economics*. Oxford, U.K.: Basil Blackwell.
- Meadows, Donella H., et al. 1992. *Beyond the Limits: Confronting Global Collapse, Envisioning a Sustainable Future*. Post Mills, VT: Chelsea Green Publishing Co.
- Nadeau, Robert L. 2003. *The Wealth of Nature: How Mainstream Economics Has Failed the Environment*. New York: Columbia University Press.
- Norgaard, Richard. 1984. "Co-Evolutionary Development Potential." *Land Economics* 60 (May):160–73.
- Krishnan, Rajaraman, Jonathan M. Harris, and Neva R. Goodwin (eds.). 1995. *A Survey of Ecological Economics*. Washington, DC: Island Press.
- Schumacher, E. F. 1974. *Small Is Beautiful: Economics As If People Mattered*. New York: Harper & Row.
- Wackernagel, M. 2002. "Tracking the Ecological Overshoot of the Human Economy." *Proceedings of the National Academy of Sciences* 99(14) (July 2002): 9266–9271.

Part II. The Containing and Sustaining Ecosystem, the Whole

- Cohen, Joel H., 1995. *How Many People Can the Earth Support?* New York: W. W. Norton.
- Georgescu-Roegen, Nicholas. 1971. *The Entropy Law and the Economic Process*. Cambridge, MA: Harvard University Press.
- Georgescu-Roegen, Nicholas. 1976. *Energy and Economic Myths*. New York: Pergamon Press.
- Hall, Charles, Cutler Cleveland, and Robert Kaufmann. 1986. *Energy and Resource Quality: The Ecology of the Economic Process*. New York: John Wiley and Sons.

Hay, Peter. 2002. *Main Currents in Western Environmental Thought*. Bloomington: Indiana University Press. (See especially Chapter 8 on economics.) Boston: Beacon Press.

Hokikian, Jack. 2002. *The Science of Disorder: Understanding the Complexity, Uncertainty, and Pollution in Our World*. Los Angeles: Los Feliz Publishing.

Kauffman, Stuart. 1995. *At Home in the Universe*. New York: Oxford University Press. (A good introduction to complexity theory.)

Meffe, Gary, et al. 2002. *Ecosystem Management: Adaptive Community-Based Conservation*. Washington, DC: Island Press.

Miller, G. Tyler. 1994. *Living in the Environment: Principles, Connections, Solutions*, 8th ed. Belmont, CA: Wadsworth.

Odum, Eugene. 1997. *Ecology: A Bridge Between Science and Society*, 3rd ed. Sunderland, MA: Sinauer Associates.

Part III. Microeconomics

Barnett, Harold, and Chandler Morse. 1963. *Scarcity and Growth: The Economics of Natural Resource Availability*. Baltimore: Johns Hopkins University Press.

Bollier, David. 2002. *Silent Theft: The Private Plunder of Our Common Wealth*. New York: Routledge.

Brekke, Kjell Arne, and Richard B. Howarth. 2002. *Status, Growth, and the Environment: Goods as Symbols in Applied Welfare Economics*. Cheltenham, U.K.: Edward Elgar.

Devarajan, S., and A. Fisher. 1981. "Hotelling's Economics of Exhaustible Resources 50 Years Later." *Journal of Economic Literature* 1 (March).

Heal, Geoffrey. 2000. *Nature and the Marketplace: Capturing the Value of Ecosystem Services*. Washington, DC: Island Press.

Page, Talbot. 1977. *Conservation and Economic Efficiency*. Baltimore: Johns Hopkins University Press.

Perrings, Charles. 1987. *Economy and Environment: A Theoretical Essay on the Interdependence of Economic and Environmental Systems*. Cambridge, U.K.: Cambridge University Press.

Price, Colin. 1993. *Time, Discounting, and Value*. Oxford, U.K.: Blackwell Publishers.

Smith, V. Kerry (ed.). 1979. *Scarcity and Growth Reconsidered*. Baltimore, MD: Johns Hopkins Press.

Part IV. Macroeconomics

Binswanger, Hans Christoph. 1994. *Money and Magic: A Critique of the Modern Economy in the Light of Goethe's Faust*. Chicago: University of Chicago Press.

Cobb, Clifford, John Cobb, et al. 1994. *The Green National Product: A Proposed Index of Sustainable Economic Welfare*. New York: University Press of America.

Collins, Robert M. 2000. *More: The Politics of Growth in Postwar America*. New York: Oxford University Press.

- Ekins, Paul, and Manfred MaxNeef (eds.). 1992. *Real-Life Economics: Understanding Wealth Creation*. New York: Routledge.
- Greider, William. 1987. *Secrets of the Temple: How the Federal Reserve Runs the Country*. New York: Simon & Schuster.
- Heyck, Denis Lynn Daly. 2002. *Surviving Globalization in Three Latin American Communities*. Toronto: Broadview Press.
- Hueting, Roefie. 1980. *New Scarcity and Economic Growth*. North Holland: Netherlands Central Bureau of Statistics.
- Socolow, R., et al. (eds.). 1994. *Industrial Ecology and Global Change*. New York: Cambridge University Press.

Part V. International Trade

- Culbertson, J. M. 1971. *Economic Development: An Ecological Approach*. New York: Alfred A. Knopf.
- Greider, William. 1997. *One World, Ready or Not: The Manic Logic of Global Capitalism*. New York: Simon & Schuster.
- Kaul, I., I. Grunberg, and M. A. Stern (eds.). 1999. *Global Public Goods: International Cooperation in the 21st Century*. New York: Oxford University Press.
- Korten, David C. 1995. *When Corporations Rule the World*. West Hartford, CT: Kumarian Press.
- Mander, Jerry, and Edward Goldsmith (eds.). 1996. *The Case Against the Global Economy*. San Francisco: Sierra Club Books.
- Pincus, Jonathan R., and Jeffrey A. Winters (eds.). 2002. *Reinventing the World Bank*. Ithaca, NY: Cornell University Press.
- Rich, Bruce. 1994. *Mortgaging the Earth: The World Bank, Environmental Impoverishment, and the Crisis of Development*. Boston: Beacon Press.
- Rodrik, Dani. 1997. *Has Globalization Gone Too Far?* Washington, DC: Institute for International Economics.
- Sachs, Wolfgang. 1999. *Planet Dialectics: Explorations in Environment and Development*. New York: Zed Books.
- Stiglitz, Joseph. 2002. *Globalization and Its Discontents*. New York: W. W. Norton.
- Vandana, Shiva. 2000. *Stolen Harvest: The Hijacking of the Global Food Supply*. Cambridge, MA: South End Press.

Part VI. Policy

- Barnes, Peter. 2001. *Who Owns the Sky? Our Common Assets and the Future of Capitalism*. Washington, DC: Island Press.
- Booth, Douglas E. 1998. *The Environmental Consequences of Growth: Steady-State Economics as an Alternative to Ecological Decline*. New York: Routledge.
- Bromley, Daniel. 1991. *Environment and Economy: Property Rights and Public Policy*. Oxford, U.K.: Blackwell.
- Brown, Lester R. 2001. *Eco-Economy: Building an Economy for the Earth*. New York: W. W. Norton.

- Costanza, Robert, et al. (eds.), 1996. *Getting Down to Earth: Practical Applications of Ecological Economics*. Washington, DC: Island Press.
- Crocker, David, and Toby Linden (eds.). 1998. *The Ethics of Consumption*, Lanham, MD: Rowman and Littlefield.
- Daly, Herman. 1996. *Beyond Growth: The Economics of Sustainable Development*. Boston: Beacon Press.
- Daly, Herman, and J. Cobb. 1994. *For the Common Good: Redirecting the Economy toward Community, the Environment, and a Sustainable Future*, 2nd ed. Boston: Beacon Press.
- Hamilton, Clive. 2003. *Growth Fetish*. Crows Neck, NSW Australia: Allen & Unwin.
- Jansson, AnnMari, et al. (eds.). 1994. *Investing in Natural Capital: The Ecological Economics Approach to Sustainability*. Washington, DC: Island Press.
- Kemmis, Daniel. 1990. *Community and the Politics of Place*. Norman: University of Oklahoma Press.
- Munasinghe, Mohan, Osvaldo Sunkel, and Carlos de Miguel (eds.). 2001. *The Sustainability of Long-Term Growth: Socioeconomic and Ecological Perspectives*. Cheltenham, U.K.: Edward Elgar.

About the Authors

Herman E. Daly is professor at the University of Maryland, School of Public Affairs. From 1988 to 1994 he was senior economist in the Environment Department of the World Bank. Prior to 1988 he was Alumni Professor of Economics at Louisiana State University, where he taught economics for twenty years. He holds a B.A. from Rice University and a Ph.D. from Vanderbilt University. He has served as Ford Foundation Visiting Professor at the University of Ceará (Brazil), as a Research Associate at Yale University, as a Visiting Fellow at the Australian National University, and as a Senior Fulbright Lecturer in Brazil. He has served on the boards of directors of numerous environmental organizations, and is co-founder and associate editor of the journal *Ecological Economics*. His interest in economic development, population, resources, and environment has resulted in over a hundred articles in professional journals and anthologies, as well as numerous books, including *Toward a Steady-State Economy* (1973), *Steady-State Economics* (1977, 1991), *Valuing the Earth* (1993), *Beyond Growth* (1996), and *Ecological Economics and the Ecology of Economics* (1999). He is co-author with theologian John B. Cobb Jr. of *For the Common Good* (1989, 1994), which received the 1991 Grawemeyer Award for Ideas for Improving World Order. In 1996 he received Sweden's Honorary Right Livelihood Award and the Heineken Prize for Environmental Science, which was awarded by the Royal Netherlands Academy of Arts and Sciences. In 1999 he was awarded the Sophie Prize (Norway) for contributions in the area of Environment and Development and in 2002 the medal of the Presidency of the Italian Republic for his work in steady-state economics.

Joshua Farley holds an undergraduate degree in biology from Grinnell College, a Master's in International Affairs with a Certificate in Latin American Studies from Columbia University, and a Ph.D. in Agricultural, Resource, and Managerial Economics from Cornell University. A fellowship at the University of Brasília first exposed him to the transdisciplinary field of ecological economics, and in particular the work of Herman Daly. Ecological economics complemented his background in biology and international development and his extensive experience working, studying, and traveling in less developed countries. Following his studies at Cornell, he spent several years teaching ecological economics at the Centre for Rainforest Studies in Far North Queensland Australia, first as resident faculty

and later as program director. His teaching emphasized transdisciplinary, applied problem solving, while his research provided abundant concrete examples of the inability of the unregulated free-market economic system to efficiently and rationally allocate ecological resources. He has spent the last four years with the Gund Institute for Ecological Economics, first as executive director while the Institute was located at the University of Maryland, then as assistant research professor at the University of Vermont. He is now assistant professor at the Department of Community Development and Applied Economics. Combining research and teaching, he enjoys working with nongovernmental organizations, community groups, and local governments to create applied transdisciplinary workshops and field courses in ecological economics.

Index

A

Abatement costs, 375–76, 375n, *table 21.1*
 Abiotic resources, 75, 77–92, *table 5.1*, 429; market failures and, 185–200
 Aboveground mineral stocks, 85–86, *fig. 5.2*, 194
 Absolute advantage, 309, 316–17, 429
 Absolute disadvantage, 334–36
 Absorptive capacity. *See* Waste absorption capacity
 Adaptive management, 362, 429
 Advertising, 412–14
 Aggregate macroeconomics, 223–27, 429
 Agreement on Trade Related Aspects of Intellectual Property (TRIPs), 325
 AIDS medicine, 167, 167n11
 Alaska Permanent Plan, 400
 Allocation, 3, 296–99, 429. *See also* Pareto efficient allocation
 Annual profit maximization, 203–7, *fig. 12.2*, *box 12.1*
 Aral Sea, 117
 Aristotle, 42, 71
 Arkwright, Richard, 326
 Asian Tigers, 337–38
 Asymmetric information, 412–14, 429
 Atlantic cod fisheries, 162–63, *box 10.1*
 Ayres, R. U., 84

B

Balance of payments (BOP), 344–46, 429
 Barnes, Peter, 400–401
 Barter, 245–46, 429
 Basic market equation, 125–40, 429
 Bass, Edward, 106
 Berra, Yogi, 235
 Berry, Wendell, 44–45
 Biophysical equilibrium, 302–3, *fig. 16.6*
 Biosphere Two, 106
 Biotic resources, 75, 93–110, *table 6.2*, 429; abiotic resources vs., 77–78; market failures and, 201–19
 Bond prices, 284–85, 300–301

Boulding, Kenneth, 39, 164n5, 181, 236, *box 13.1*
 Bound energy, 65–66
 Bretton Woods Institutions, 318–19, 429
 Buffet, Warren, 393, 393n
 Bush, George W., 330, *box 18.2*
 Butz, Earl, 340, *box 18.3*
 By-catch, 384, 384n6

C

Camdessus, Michel, 354
 Cap and trade systems, 375n, 379–85
 Capital account, 314, 344, 430
 Capitalist circulation, 246
 Capital mobility, 314–16, 346–48
 Carnot, Sadi, 64, 66
 Carrying capacity, 6, 98–100, 430
 Cassowaries, 102, *box 6.3*
 Catch-per-unit-effort hypothesis, 202–4, 202n2, *fig. 12.1*, 430
 Circular flow, 24–26, *fig. 2.4*, 226–27, *fig. 13.1*, 263, 279, 430
 Clausius, Rudolf, 65
 Cleveland, Cutler J., 177n26
 Clinton, Bill, 328, 330, *box 18.2*, 398n10
 Closed system, 15, 68, 430
 CO₂ emissions, 119, 361
 Coase, Ronald, 176, 179, 325
 Coase theorem, 176–77, 179, 430
 Cobb, Clifford, 238
 Cobb, John, 234
 Cobb-Douglass production function, 151, 151n5
 Coevolutionary economics, 7–10, 430
 Cohen, Ben, 266
 Command-and-control regulations, 374, 430
 Commons, 161, 164–67, 334n17
 Communicable diseases, 172–74
 Community, 262, 267–68
 Community Shareholder Ownership Programs (CSOPs), 397
 Comparative advantage, 309–16, *table 17.1*
 Competitive market, 126, 430

Complementarity, 149–50, 149n, 152–54, 430
 Comprehensive efficiency, 422–23, 430
 Congestibility, 159–60, *table 10.1*, 168, 430
 Consilience (Berry), 45
 Constant (local currency), 251
 Consumer surplus, 143–45, *fig. 9.3*, 430
 Contingent valuation, 407, *box 23.1*, 430–31
 Copyright Act, 167
 Critical depensation, 99, 101–3, *boxes 6.2, 3*, 431
 Crowding out, 290–91
 CSOPs (Community Shareholder Ownership Programs), 397
 Current account, 314, 344, 431

D

Daly, Herman, 234
 Debt crises, 348–55, *box 19.1*
 Defensive expenditures, 231–32, 431
 Deflation, 293
 Deforestation, 416–19, *fig. 23.1*
 Demand, 24–26, *fig. 2.4*, 136–55, *figs. 8.2, 4*, *figs. 9.1–3*, 431
 Democritus, 47, *box 3.1*
 Depletable, 89, 89n15
 Depreciation, 233n8
 Determinism, 44–47, *box 3.1*, 53, 431
 Development, 6–7, 431
 “Diamonds-water paradox,” 197n, 247, *box 14.1*, 406
 Differential rent. *See* Rent
 Diminishing marginal physical product. *See* Law of diminishing marginal physical product
 Diminishing marginal utility. *See* Law of diminishing marginal utility
 Direct regulation, 373–76
 Disinflation, 291–94, 431. *See also* Inflation
 Dissipative structures, 55, 68n11
 Distribution, 12–13, 259–76, *figs. 15.1, 2*, 296–99, 431; justice in, 333–40, 363–65, 389–404. *See also* Intergenerational distribution

Double dividend, 385, 385n
 Doubling time, 112, *box 7.1, fig. 7.1*, 431
 Dualism, 37–43, 50, 54

E

Ecological constraints, 10–13
 Ecological economics, 4, 425–28;
 distribution and, 262, 266–67;
 growth and, 13, 64; intertemporal
 distribution and, 269–70; IS-LM
 model and, 302–4, 302nn; optimal
 scale and, 226; production function
 and, 150–52
 Ecological footprint, 34–35
 Ecological reductionism, 50, *fig. 3.2*,
 53–54, 431
 Ecological tax reform, 145, 385–86
 Economic equilibrium, 302–3, *fig. 16.6*
 Economic imperialism, 50–53, *fig. 3.2*,
 431
 Economic stability, 348–55, *boxes*
 19,2,3
Economist, The, 39
 Economy as subsystem, 17–18, *fig. 2.1*,
 34–35
 Economy of scale, 127n
 Ecosystem functions, 75, 93–97, 94n2,
 118–19, 431
 Ecosystem services, 21–22, *fig. 2.3*, 75,
 97, 103–7, *tables 6.1,2*, 431; full
 world and, 118–19; globalization
 and, 328n11; GNP and, 237,
 237n13; public goods and, 171,
 171nn14,16; renewable resources
 and, 211, *fig. 12.4*. *See also* Fund-
 service resources; Natural capital;
 Nonmarket goods and services
 Ecosystem structure, 93–97, 94nn1,2,
 118–19, 432
 “Efficiency of the commons,” 165–67,
 box 10.2
 Efficient allocation. *See* Pareto efficient
 allocation
 Efficient cause, 71–72, *box 4.1*,
 149–50, 432. *See also* Fund-service
 resources; Material cause
 Ehrlich, Paul R., 44
 Einstein, Albert, 24, 38n1, 46, 66
 Easley, Loren, 39
 Elasticity, 146–48, *figs. 9.4,5*, 432;
 freshwater and, 197–98, *fig. 11.5*,
 197n

Employee Shareholder Ownership
 Programs (ESOPs), 396–98
 Empty-world vision, 17–18, *fig. 2.1*,
 21, 111–21, 154, 302, 370
 Ends and means, 3–5, 37–43
 Ends-means spectrum, 48–50, *fig. 3.1*,
 54, 129–30
 Energy Information Administration,
 113
 Energy return on investment, 78–79
 Entropic dissipation, 38, 432; mineral
 resources and, 84–86, *fig. 5.2*
 Entropy. *See* Second Law of
 Thermodynamics
 Entropy hourglass, 30–31, *box 2.2*
 Environmental economics, 5, 432
 Epicurus, 47, *box 3.1*
 Equilibrium, 143–45, *fig. 9.2*
 Equimarginal principle of
 maximization, 127–28, 128nn3,4,
 130, 133, 432
 ESOPs (Employee Shareholder
 Ownership Programs), 396–98,
 398n10
 Ether, 40, 40n10
 Exchange rate, 344–46, 432
 Exchange value, 28, 28n, 246–47,
 246n, *box 14.1*, 256, 406, 432
 Excludability, 72–74, 157–60, *table*
 10.1, 164–68, 432; abiotic resources
 and, 88, 89, 92; biotic resources
 and, 109, *table 6.2*
 Exclusion zones, 164
 Exponential growth, 112, *box 7.1, fig.*
 7.1, 247, 255–56, 432
 Exports, 27, *fig. 2.5*
 Externalities, 175–80, 176nn21–23,
 432; of fossil fuels, 186–87, *table*
 11.1, *figs. 11.2,3*; globalization and,
 328–29; of mineral resources,
 192–93, *table 11.2*

F

Fallacy of composition, 225, 432
 Fallacy of misplaced concreteness,
 30–33, 432–33
 Federal Deposit Insurance Corporation
 (FDIC), 252
 Federal Open Market Committee
 (FOMC), 300, 300n14
 Federal Reserve System, 252, 254–55,
 284, *box 16.3*, 299–300, *box 16.5*,
 433
 “Fiduciary issue,” 249, *box 14.2*

Financial crises, 350–55, 350n8
 Findlay, R., 310
 Finite planet, 62–64
 First Law of Thermodynamics, 30–32,
 box 2.2, 38, 65, 69, 433; mineral
 resources and, 83, 86; production
 function and, 151
 Fiscal policy: globalization and,
 346–48, 354–55; IS-LM model and,
 289–99, *table 16.1*
 Fisher, Irving, 243, 250, 252, 256
 Fitness of environment, 153–54, 153n
 Fixed exchange rate regime, 345–47,
 433
 Flexible exchange rate regime, 345–48,
 433
 Floating exchange rate regime, 345,
 433
 Food and Agriculture Organization
 (FAO), 118
 Food security, 339–40, *box 18.3*
 Ford, Henry, 337
Fortune magazine, 117
 Fossil fuels, 74–75, 78–83, *box 5.1, fig.*
 5.1, 91, *table 5.1*, 433; full world
 and, 113–16, 113n5; market failures
 and, 186–92, *figs. 11.1–4, table 11.1*;
 solar energy and, 81–82, *fig. 5.1*, 84
 Fractional reserve banking, 252–53,
 433
 Free-rider effect, 170–71, *box 10.3*,
 179, 433
 Freshwater, 75, 87–88, 91, *table 5.1*;
 full world and, 116–17; market
 failures and, 196–98, 196nn14,15
 Full world, 17–18, *fig. 2.1*, 111–21, *box*
 7.1, 154, 303
 Fund-service resources, 70–72, *box 4.1*,
 149–50, 433; abiotic resources as,
 79–82, *box 5.1*, 87–88, 90, 91–92,
 table 5.1; basic market equation and,
 126; biotic resources as, 104, 106–7,
 109–10, *table 6.2*, 211–14;
 excludability and, 158–59;
 production function and, 150–52;
 rivalness and, 159; utility function
 and, 154. *See also* Stock-flow
 resources

G

Gaia hypothesis, 104n9
 Galbraith, James K., 295
 Garbo-junk, 85n, 270
 Gates, Bill, 95, *box 6.1*

- Gates, Bill, Sr., 393, 393n
 GATT (General Agreement on Trade and Tariffs), 319, 328
 General equilibrium model, 224, 277–78, 433
 George, Henry, 145, 199
 Georgescu-Roegen, Nicholas, 30, *box* 2.2, 65–66, 71, 84, 85n, 115n9
 Gilder, George, 39, 41
 Gini coefficient, 263, *table* 15.1, 265–66, 433
 Global climate change, 116–17, 119–20, 272, *box* 15.1
 Global Development Network Growth Database (World Bank), 334n17
 Globalization, 317, 319–21, 323–41, 386, 433. *See also* Internationalization
 Gnosticism, 39, 39n5, 41–42
 Gore, Al, 328
 Gosplan, 51, 51n
 Government expenditure, 26–27, *fig.* 2.5
 Government intervention, 177, 179
 Government regulation, 186–87, 186n2
 Great Depression, 13, 26, 225
Great Transformation, The (Polanyi), 7
 Greenfield, Jerry, 266
 Green Golden Rule, 213, 213n
 Greenspan, Alan, 300
 Gross Domestic Product (GDP), 228, 228n2
 Gross National Product (GNP), 7, 147n, 228–36, *fig.* 13.3, *box* 13.1, 433
 Growth, 6–7, 62–64, 433
 Growth fetish, 13
- H**
 Hardin, Garret, 161
 Hayek, F. A., 133
 Heal, Geoffrey, 213, 213n
 Health, 267–68
 Healthy Forests Plan, 400
 Heilbroner, R., 32
 Helmholtz, Herman, 65
 Hicks, John, Sir, 278, 279, 299n, 301–2, 327, *box* 18.1
 Hicks-Kaldor welfare criterion, 316, 422
 Hotelling, Harold, 189, 194
 Hotelling rule, 189, 433
 Hubbert, M. King, 114, 255
 Hubbert curve, 114, *figs.* 7.2,3, 434
 Human needs, 237–44, 268
 Human needs assessment (HNA), 242–44, 243n23
 Human well-being. *See* Welfare
 Humphrey, Hubert, 340, *box* 18.3
 Hunter-gatherers, 8–9
 Hydrologic cycle, 88, 116–17
 Hyperbolic discounting, 275, *box* 15.3, 434
 Hyperinflation, 292, 434
- I**
 ICMS *ecológico*, 420
 Ignorance, 94–95, *box* 6.1, 179, 328n11, 408
 IMF (International Monetary Fund), 196n15, 318–20, 343, 350, 354, 362, 386, 398n11
 Imports, 27, *fig.* 2.5
 Inalienability rule, 371, 434
 Income, 232–33, 262–63, 262n, *table* 15.1, 334–36, 334n20, 434; caps on, 390–93
 Increasing marginal cost. *See* Law of increasing marginal cost
 Index of Sustainable Economic Welfare (ISEW), 233–36, 234n11, *fig.* 13.3, 243, 434
 Individual transferable quotas (ITQs), 383–85, 434
 Industrial Revolution, 9–10, 64, 268
 Inequality, 265–68
 Inflation, 291–95, 294n8, 301, *box* 16.5, 434
 Information as resource, 39–41, 164–68, 165n7, 172–74
 Information technology (IT) economy, 40
 Injections, 26–29, *fig.* 2.5
 Insatiability, 4, 4n2
 Intellectual property rights, 158, 165, 168, 325–26, *box* 18.1
 Interest rates, 283–86, 300–301
 Intergenerational distribution, 12, 86–87; abiotic resources and, 89, 91–92, *table* 5.1, 198; biotic resources and, 109, *table* 6.2; market failures and, 179–81; welfare and, 235. *See also* Intertemporal distribution
 International Bank for Reconstruction and Development. *See* World Bank
 International Clearing Union, 319, 351
 International flows, 343–56
 Internationalization, 317, 319–21, 434. *See also* Globalization
 International Monetary Fund. *See* IMF
 International Pacific Halibut Commission (IPHC), 382
 International Patent Institute, 326
 International policies, 419–22
 International trade, 309–21; globalization and, 323–41; macroeconomic policy and, 343–56
 Intertemporal allocation, 269–70, 276, 434
 Intertemporal discounting, 52, 181–82, 271–76, *box* 15.1, 434
 Intertemporal distribution, 268–76, 408–10, 434. *See also* Intergenerational distribution
 Investment, 26–27, *fig.* 2.5, 279–82, *figs.* 16.1,2,5, 280n, *box* 16.2, 287–88
 “Invisible hand,” 224
 IS curve, 279–82, *figs.* 16.1,2,5, *box* 16.2, 302, 302n17
 ISEW (Index of Sustainable Economic Welfare), 233–36, 234n11, *fig.* 13.3, 243, 434
 IS-LM model, 224, 277–304, *figs.* 16.1–6, 299n, 435
 Isolated systems, 15, 27–28, *fig.* 2.5, 68, 435
 Ithaca HOURS, 251, *box* 14.3
 Ivanhoe, L. F., 113n5
- J**
 Jevons, William Stanley, 20, 68
 Joule, James, 65
 Junk bonds, 287–88, *box* 16.4
 Just distribution, 333–40, 363–65, 389–404
- K**
 Kapor, Mitch, 327, *box* 18.1
 Keynes, John Maynard: Bretton Woods Institutions and, 318–19, 326, 328; carrying capacity and, 98; International Clearing Union and, 319, 351; monetary sector and, 278; Say’s Law and, 26, 225; user costs and, 190, *box* 11.1; virtual wealth and, 256–57
 Knight, Frank, 95, *box* 6.1, 250, 252, 256

Knowledge as resource, 39–41,
164–68, 165n7, 172–74, 326
Krugman, Paul, 310
Kuhn, Thomas, 23–24
Kyoto protocol, 119

L

Land. *See* Ricardian land
Land tax, 401–4, *fig.* 22.1
Land value, 198–200
Law of diminishing marginal physical
product, 126–29, 127n, 148–49, *fig.*
9.6, 435
Law of diminishing marginal utility,
19–23, *box* 2.1, 127, 152–53, *fig.*
9.7, 352, 435
Law of entropy. *See* Second Law of
Thermodynamics
Law of increasing marginal cost,
19–23, *box* 2.1, 435
Lay, Kenneth, 330, *box* 18.2
Leakages, 26–29, *fig.* 2.5
Leisure, 234, 234n10
Less-developed countries (LDCs),
337–40, 349, *box* 19.1, 352
Lessig, Lawrence, 327, *box* 18.1
Liability rule, 371, 435
Linear throughput. *See* Throughput
Linux operating system, 166–68, *box*
10.2
Liquidation, 233, 233n8
Liquidity preference, 282–83, 287–88,
435
Liquidity trap, 291, 435
Living plant index, 34–35
LM curve, 283–85, *fig.* 16.3, 5, 302,
302n17
Lobbying, 330, *box* 18.2
Local currencies, 250–51, *box* 14.3
Locke, John, 9n13
Lorenz curve, 264–65, *fig.* 15.1, 435
Lovelock, James, 104n9
Lucretius, 47, *box* 3.1

M

MacKenzie, J. J., 113n5
Macro-allocation, 174, 296–99,
412–15, 435
Macroeconomic policy, 343–56
Macroeconomics, 6, 6n, 21, 21n4,
223–27, 223n
Malthus, Thomas, 32–33, 119, 174
Marginal abatement costs (MACs),
375–76, 375n, *table* 21.1

Marginal cost, 16–17, 16n2, 19–23,
box 2.1, 133, 133n8, 435
Marginal disutility (MDU), 20–21, *fig.*
2.2, 230–31, *fig.* 13.2
Marginal external costs (MEC), 175,
177–78, *fig.* 10.2, 195, 435; of
deforestation, 416, *fig.* 23.1; of fossil
fuels, 186–89, *figs.* 11.2, 3
Marginal extraction costs (MEX),
186–89, *figs.* 11.1–3, 186n1, 187n
Marginal net private benefit (MNPB),
177–78, *fig.* 10.2, 416, *fig.* 23.1
Marginal physical product (MPP). *See*
Law of diminishing marginal
physical product
Marginal propensity to consume, 286
Marginal propensity to save, 286–88
Marginal revenue, 133, 133n8, 435
Marginal user costs (MUC), 187–88,
187n, 195, 435
Marginal utility (MU), 19–23, *box* 2.1,
fig. 2.2, 435; basic market equation
and, 126, 130–32; non-price
adjustments and, 134–35. *See also*
Law of diminishing marginal utility
Market failures, 157–83; abiotic
resources and, 185–200; biotic
resources and, 201–19
Marshall, Alfred, 69, 69n15, 327, *box*
18.1
Marshall Plan, 319
Marx, Karl, 32–33, 246
Maslow's hierarchy, 238
Material cause, 71–72, *box* 4.1,
149–50, 436
Materialism, 44–47, *box* 3.1, 436
Matrix of human needs (Max-Neef),
238–44, *table* 13.1, 268, 436
Maximizing annual profits, 203–7, *fig.*
12.2, *box* 12.1
Maximum sustainable yield (MSY),
98–103, *fig.* 6.1, *boxes* 6.2, 3, 180,
202–5, 202n1, *figs.* 12.1, 2, 436
Max-Neef, Manfred, 238–44, 268
Mayer, Robert, 65
Means. *See* Ends and means; Ultimate
means
Measured Economic Welfare (MEW),
233–36, 234n11
Mechanical physics, 66–67, 69
Micro-allocation, 297
Microeconomics, 6n, 21, 21n4,
125–26, 224–27
Mill, John Stuart, 54–55, *box* 3.2

Mineral resources, 74–75, 83–87, *fig.*
5.2, 85n, 91, *table* 5.1, 436; full
world and, 116; market failures and,
192–96, *table* 11.2
Minimum income, 393–95
Minimum viable population, 99, 102,
436. *See also* Critical depensation
Mining, 84–86, *fig.* 5.2
Mining Law (1872), 399
Mishan, E. J., 40
Missing markets, 180–82
Monetary policy: globalization and,
346–48, 354–55; IS-LM model and,
289–99, *table* 16.1
Monetary sector (LM), 278, 282–85
Money, 245–58, 436
Money creation, 252–53
Monisms, 50, 53–54
Monopoly, 133–34, 164–65, 167, 198,
324n4, 325–28, *box* 18.1, 436
Monopsony, 324, 324n4
Moral hazard, 350, 353, 436
Morganstern, Oskar, 278
Morgenthau, Henry, 318
Multi-tier pricing, 168, 436
Mystery, 50, 50n, 54

N

NAIRU, 295, 304
National Income, 25–26
National Institute for Environmental
Studies (Japan), 34
National Product, 25–26
Natural capital, 17, 76, 106–7, 436;
consumption of, 232–33;
globalization and, 335; just
distribution and, 399–404; user
costs and, 190, *box* 11.1; utility
function and, 153–55. *See also*
Ecosystem services; Nonmarket
goods and services
Natural dividend, 213–14, *fig.* 12.6,
436
Natural monopoly, 198
Neoclassical economics (NCE), 4, 436;
efficient allocation in, 13;
intertemporal distribution and,
270–76; Pareto efficient allocation
in, 262; production function in,
151–52; public goods and, 174
Netherlands Ministry for Housing
Spatial Planning and Development,
33–34
Net national product (NNP), 232

Net present value (NPV), 182, 207–11, 272, 409, 436
 Net primary productivity (NPP), 90
 Net recoverable energy, 81–82, *fig. 5.1*
 Newton, Isaac, 24
 Nihilism, 44–47, *box 3.1*, 53, 436
 Nixon, Richard, 330, *box 18.2*
 Nondeterminism, 43–44
 Nonexcludable resources, 73–74, 88, 158–60, *table 10.1*, 437
 Nonmarket goods and services: pricing and valuing of, 405–11; spatial aspects of, 415–22, *fig. 23.1*; subsidies for, 414–15. *See also* Ecosystem services; Natural capital
 Nonmarket value, 410–11, 437
 Nonnihilism, 43–44
 Non-price adjustments, 134–36, 437
 Nonrenewable resources, 75, 437
 Nonrival resources, 73–74, 159–60, *table 10.1*, 164–68, 437
 Nordhaus, William, 233–34
 Norgaard, Richard, 7n6, 195
 Normal profit, 145, 437
 Nuclear wastes, 120
 Nussbaum, Martha, 241–42

O

Ogallala aquifer, 117
 Oklahoma City bombing, 400n17
 Open access, 161–64, *box 10.1*, 437
 Open-source software, 166–67, *box 10.2*
 Open system, 15, 68, 437
 Opportunity cost, 15–17, 181, 271, 437
 Optimal scale, 16–18, 20–21, 53, 226, 437; externalities and, 177, *fig. 10.2*
 Orange roughy, 118, 118n26
 Overdeveloped countries (ODCs), 332, 332n15, 339

P

P10–P90 reserve estimates, 79–80, *box 5.1*
 Paradigm shifts, 23–26, 24n
 Paradox of thrift, 287, 437
 Parametric function of prices, 126, 126n, 131–32, *fig. 8.1*, 135
 Pareto, Vilfredo, 4
 Pareto efficient allocation, 4, 133–36, 259–62, *box 15.1*, 323–31, 363–65, 405–24, 437

Patents, 165–68, 165n7, 167n11, 173, 173n19, 325–28, *box 18.1*
 Per-capita incomes, 334–36, 334n20
 Perfect competition, 324–25
 Perfect information, 134, 134n, 179, 364–65
 Perpetual motion machine, 64–65
 Persistent organic pollutants (POPs), 120–21
 Perverse economic incentives, 79n4
 Petro-dollars, 349, *box 19.1*
 Phillips curve, 294n8
 Phlogiston, 40, 40n9
 Phosphorus, 39, 39n6
 Pigouvian subsidies, 378–79, 438
 Pigouvian taxes, 376–78, 437–38
 Plato, 266
 Plimsoll, Samuel, 5n
 Plimsoll line, 5
 Polanyi, Karl, 7
 Policy: design principles for, 359–72; international, 419–22. *See also* Fiscal policy; Monetary policy
 Policy dialogue, 43–44
 Political power, 330, *box 18.2*, 391, *box 22.1*
 Pollution, 121, 175–79, *fig. 10.2*, 177n26
 Polymers, 84
 Population growth, 63, 63nn2,4, 161, 164, 164n5, 361
 Practical dualism. *See* Dualism
 Preanalytic visions, 23–24
 Presuppositions of policy, 43–44
 Price, 50; of fossil fuels, 191, *fig. 11.4*; freshwater and, 197, *fig. 11.5*; of mineral resources, 193–96; of nonmarket goods and services, 405–11, *boxes 23.1,2*. *See also* Parametric function of prices
 Price vs. quantity, 366–68, *fig. 20.1*
 Prigogine, Ilya, 68n11
 Principle of subsidiarity, 363, 438
 Private property, 9, 9n13, 368–69, 393
 Privatization, 334n17, 398, 398n11
 Producer surplus, 143–45, *fig. 9.3*, 438
 Production function, 148–52, 149n, 151nn, 438
 Profits, 206–11, 206nn, 246
 Property rights, 8–9, 8n10, 157–59, 161–65, 163n; market failures and, 176, 180; policy design principles and, 369–72
 Property rule, 370, 438
 Public bads, 186, 215, 413, 438

Public goods, 168–75, 253–55, 296–99, 296n, 351, 438
 Public subsidies, 399–400
 Puckett, B. Earl, 413
 Pure time rate of preference (PTRP), 271, 274–75, 438

Q

Quality of life (QOL). *See* Welfare
 Quantity of money theory of income, 278, *box 16.1*
 Quotas, 366–68, *fig. 20.1*, 379–85

R

Race to the bottom, 329, 336–39, 353
 Ramsar convention, 119
 Reagan, Ronald, 294, 330, *box 18.2*, 398n10
 Real sector (IS), 278–82
 Recoverable energy, 78–82, *fig. 5.1*
 Recycling, 192; of mineral resources, 84–86
 Reforestation, 175–76, 176n21
 Reich, Robert, 398n10
 Relativism, 44–47
 Renewable resources, 75, 97–103, *fig. 6.1*, *table 6.2*, 438; ecosystem services and, 211–14, *figs. 12.5,6*
 Rent, 143–45, *fig. 9.3*, 189, 228n3, 330, *box 18.2*, 399, 438
 Resource exhaustion, 81, 118, 174
 Ricardian land, 75, 88–89, 91, *table 5.1*, 198–200, 438
 Ricardo, David, 310–13, 316
 Riders, 400, 400n17
 Risk, 95, *box 6.1*, 354, 438
 Risk aversion, 353, *box 19.3*
 Rivalness, 72–74, 159–60, *table 10.1*, 438; abiotic resources and, 86–89, *table 5.1*, 92; biotic resources and, 109, *table 6.2*
 Royalties, 189, 399, 438
 Ruggiero, Renato, 320
 Ruskin, John, 21

S

Savings, 26–27, *fig. 2.5*, 279–82, *figs. 16.1,2,5*, 280n, *box 16.2*, 286–87
 Say's Law, 26, 225, 439
 Scale, 12–13, 259–62, *box 15.1*, 296–99, 439; sustainable, 331–33, 363–65, 373–87

- Scarcity, 74–75; freshwater and, 196; mineral resources and, 193–96, 194nn; public goods and, 171–74
- Schrodinger, Erwin, 46, 68
- Schumpeter, Joseph, 23, 250, 326
- Second Law of Thermodynamics, 29, 31–32, *box* 2.2, 38, 65–66, 69, 439; mineral resources and, 83–87, 85n
- Seigniorage, 249–51, *box* 14.3, 290, 415, 439
- Selling short, 352, *box* 19.2
- Sen, Amartya, 241–43, 339
- Shapiro, Robert, 330, *box* 18.2
- Shortage, 143–45, *fig.* 9.2
- Simon, Julian, 63n2
- Simple commodity production, 246
- Sinks, 121, 368–69, 439
- Sismond, 173
- Sky trust, 400–401
- Slater, Samuel, 326
- Sliding fulcrum function of prices, 126, 126n, 131–32, *fig.* 8.1, 135
- Smith, Adam, 164, 224
- Social contract, 158, 336
- Social discount rate, 273, 439
- Social limits to range of inequality, 267
- Soddy, Frederick, 39, 248–50, 252, 254–56, 290
- Solar energy, 31, *box* 2.2, 62, 62n, 75, 439; as abiotic resource, 89–92, 90n20, *table* 5.1; fossil fuels and, 81–82, *fig.* 5.1, 84; market failures and, 200; water and, 88
- Soros, George, 257, 354–55
- Source limits, 121, 368–69, 439
- Specialization, 330–31
- Speculation, 351–52, *box* 19.2
- Standards-lowering competition, 329, 336–39
- Statistical mechanics, 66–67
- Steady-state economy, 50, *fig.* 3.2, 54–55, *box* 3.2, 439; mineral resources and, 84, 86
- Stiglitz, Joseph, 412
- Stock-flow resources, 70–72, *box* 4.1, 439–40; abiotic resources as, 79–82, *box* 5.1, 88, 91–92, *table* 5.1, 196–98; basic market equation and, 126; biotic resources as, 98, 104, 106–7, 109–10, *table* 6.2, 201–11; excludability and, 158–59; production function and, 150–52; utility function and, 154
- Storage technologies, 8–9, 8n10
- Subsidies, 399–400, 414–15, 419–22, 440
- Substitutability, 149–50, 149n, 440; abiotic resources and, 91–92, *table* 5.1; biotic resources and, 109, *table* 6.2; public goods and, 174, 174n
- Subterranean mineral stocks, 85–86, *fig.* 5.2, 194, 194nn
- Sun Yat-Sen, 337
- Supply, 24–26, *fig.* 2.4, 136–55, *figs.* 8.3, 4, *figs.* 9.1–3, 440
- Surplus, 143–45, *fig.* 9.2
- Sustainable income. *See* Income
- Sustainable scale, 331–33, 363–65, 373–87
- Sustainable yield. *See* Maximum sustainable yield (MSY)
- T**
- Taxation, 26–27, *fig.* 2.5
- Thermodynamics, 29–35, 64–70, 255–58, 440. *See also* First Law of Thermodynamics; Second Law of Thermodynamics
- Third World, 167, 167n11, 238
- Throughput, 6, 24–26, 29–35, 365–66, 440
- Thurrow, L., 32
- Timber companies, 287–88, *box* 16.4
- Tinbergen, Jan, 360
- Tobin, James, 233–34, 249, *box* 14.2
- Torvald, Linus, 166
- Total allowable commercial catch (TACC), 383–85, 440
- Tradeable permits, 375n, 379–85, 440
- Trade secrets, 164, 165n7
- “Tragedy of open access regimes,” 161–63, *box* 10.1
- “Tragedy of the commons,” 161, 164
- Transaction costs, 178–79, 440
- Transactions demand for money, 282–83, 440
- Transnational corporations, 324–25, 330, *box* 18.2
- TRIPs (Agreement on Trade Related Aspects of Intellectual Property), 325
- Tuberculosis treatment, 172–74
- U**
- Ultimate end, 42–43, 48–49, *fig.* 3.1, 440
- Ultimate means, 38–42, 48–50, *fig.* 3.1, 440
- Uncertainty, 94–96, *box* 6.1, 98, 101, *box* 6.2, 328n11, 408, 440; externalities and, 179
- Uneconomic growth, 16, 16n1, 19–23, 64, 440
- Unemployment, 294–96, 294n8, 304, 441
- UN Food and Agriculture Organization (FAO), 118
- U.S. Clean Air Act, 328
- U.S. Congress, 165, 167
- U.S. Constitution, 165, 167
- U.S. Endangered Species Act, 328
- User costs, 187–91, *box* 11.1, 441
- Use value, 246–47, *box* 14.1, 406, 441
- U.S. Marine Mammal Protection Act, 328
- U.S. Public Law 480, 340, *box* 18.3
- U.S. Supreme Court, 167
- Utility function, 152–55, 441
- V**
- Villa, Ferdinando, 166, *box* 10.2
- Virtual wealth, 248–49, *box* 14.2, 256, 441
- W**
- Walras, Leon, 68–69
- Waste absorption capacity, 107–10, *table* 6.2, 119–21, 215–18, *fig.* 12.7, 441; as ecosystem service, 75, 97; intertemporal distribution and, 270
- Waste as resource, 41–42
- Water, 75, 87–88, 91, *table* 5.1, 116–17, 196–98, 334n17
- Wealth, 262–63, 262n, *table* 15.1, 330, *box* 18.2; caps on, 390–93, *box* 22.1; relative, 237
- Wealth effects, 178–79
- Welfare, 17–18, *fig.* 2.1, 230–44, *figs.* 13.2, 3, 234nn10, 11, 330–31, 441
- “When to stop” rule, 17, 128, 128nn3, 4
- White, Harry Dexter, 318
- Whitehead, Alfred North, 30, 43–44
- Whole and part, 15–16, 21–24, 21n4, *fig.* 2.3, 28–29, 218–19
- Wilson, Edward O., 45
- World Bank, 117, 154, 196n15, 318–20, 333, 362, 386, 434
- World Resources Institute (WRI, U.S.), 33
- World Trade Organization (WTO), 118–19, 319–21, 325–26, 328–30, *box* 18.2, 386, 441
- World Wildlife Fund (WWF), 34
- Wuppertal Institute (Germany), 33

Island Press Board of Directors

Victor M. Sher, Esq., *Chair*
Environmental Lawyer, Sher & Leff

Dane A. Nichols, *Vice-Chair*
Environmentalist

Carolyn Peache, *Secretary*
President, Campbell, Peachey & Associates

Drummond Pike, *Treasurer*
President, The Tides Foundation

Robert E. Baensch
Director, Center for Publishing, New York University

David C. Cole
President, Aquaterra, Inc.

Catherine M. Conover
Chair, Board of Directors, Quercus LLC

Henry Reath
President, Collectors Reprints Inc.

Will Rogers
President, Trust for Public Land

Charles C. Savitt
President, Center for Resource, Economics/Island Press

Susan E. Sechler
Senior Advisor on Biotechnology Policy,
The Rockefeller Foundation

Peter R. Stein
General Partner, The Lyme Timber Company

Diana Wall, Ph.D.
Director and Professor, Natural Resource,
Ecology Laboratory, Colorado State University

Wren Wirth
President, The Winslow Foundation