

THE ANNUAL REPORT
OF THE
COUNCIL OF ECONOMIC ADVISERS

LETTER OF TRANSMITTAL

COUNCIL OF ECONOMIC ADVISERS,
Washington, D.C., February 5, 2002.

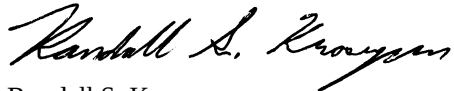
MR. PRESIDENT:

The Council of Economic Advisers herewith submits its 2002 Annual Report in accordance with the provisions of the Employment Act of 1946 as amended by the Full Employment and Balanced Growth Act of 1978.

Sincerely,



Robert Glenn Hubbard
Chairman



Randall S. Kroszner
Member



Mark B. McClellan
Member

C O N T E N T S

	<i>Page</i>
overview	15
chapter 1. restoring prosperity.....	23
Macroeconomic Performance in 2001: Softer Economy, Harder Choices	23
Aggregate Demand During the First Three Quarters	23
Preliminary Evidence on Aggregate Demand in the Fourth Quarter	30
Labor Markets	33
Inflation.....	34
Productivity and Employment Costs	34
Saving and Investment.....	35
The Cyclical Slowdown	36
Moderation After Very Rapid Growth	36
Decline in Equity Values.....	37
Surge in Energy Prices	37
Higher Interest Rates	38
Collapse of the High-Technology Sector.....	38
Lingering Effects of Y2K	38
Effects on Inventories and the Capital Stock.....	39
From Slowdown to Recession	41
Policy Developments in 2001.....	43
Fiscal Policy Before the Terrorist Attacks.....	43
Tax Relief in 2001	44
Monetary Policy Before the Terrorist Attacks.....	45
The Macroeconomic Policy Response After September 11	46
Economic Developments Outside the United States	51
The Economic Outlook	52
Near-Term Outlook: Poised for Recovery	52
Inflation Forecast	54
Long-Term Outlook: Strengthening the Foundation for the Future	54
The Policy Outlook: An Agenda for Economic Security	61
chapter 2. strengthening retirement security.....	65
Rationale for a National Retirement System	66
Insurance Against Uncertainty.....	66
Foresight and Planning.....	67
Redistributive Goals	68

	<i>Page</i>
Sources of Retirement Security	69
Social Security	69
Employer-Sponsored Pensions	71
Individual Savings.....	72
Labor Earnings	72
Public Assistance.....	72
Challenges Ahead	73
Social Security: Past and Present.....	74
Origins of the Current System.....	74
Social Security and National Saving.....	76
The Future of Social Security	79
Advantages of Personal Accounts	79
The Financial Sustainability of Social Security.....	86
Other Sources of Retirement Security	92
Employer-Sponsored Pension Plans	93
Individual Saving.....	94
Fostering Self-Reliance.....	96
Meeting the Challenge of Retirement Security	96
chapter 3. realizing gains from competition	99
Motivations for Organizational Change	101
The Role of Agency Costs in Organizational Change.....	102
Mergers.....	103
Other Organizational Forms: Joint Ventures and Partial Equity Stakes	107
Incorporating Economic Insights into Competition Policy	112
Competition Policy, Corporate Governance, and the Mergers of the 1980s and 1990s	114
The Role of Corporate Governance Changes.....	116
Policy Lessons for Promoting Organizational Efficiencies.....	117
Policy Lessons from Joint Ventures	118
Shaping Policies to Address Partial Equity Stakes.....	120
Policy Toward Vertical Relations	123
Cross-Border Organizational Changes.....	125
Multijurisdictional Review.....	125
Elements of International Policy Convergence.....	127
Core Principles of Competition Policy.....	127
Dynamic Competition and Antitrust Policy.....	130
Sources of Incentives for Innovation.....	132
Fostering Innovation Through Organizational Structure	136
Dynamic Competition as Repeated Innovations.....	137
Implications of Dynamic Competition for Competition Policy.....	138
Conclusion.....	142

	<i>Page</i>
chapter 4. promoting health care quality and access.....	145
Encouraging Flexible, Innovative, and Broadly Available Health	
Care Coverage	149
Recent Trends in Health Care Costs and Coverage.....	149
Addressing Barriers to Effective Competition in Health Insurance .	154
Increasing Health Insurance Coverage	159
Making Medicare Coverage More Flexible and Efficient	166
Better Support for High-Quality, Efficient Care.....	171
Shortfalls in the Quality of Care	171
Disparities in the Health Care System	173
Empowering Providers to Improve Quality of Care	175
Empowering Patients to Make Informed Health Care Choices.....	176
Fulfilling the Promise of Medical Research.....	179
The Benefits of Biomedical Research	180
Many Unanswered Questions About Existing Medical Treatments.	182
The Role of the Federal Government in Supporting Research	184
Conclusion: Fulfilling the Potential of 21 st -Century Health Care.....	185
chapter 5. redesigning federalism for the 21st century.....	187
Institutional Design in a Federal System.....	188
Fostering Partnerships, Competition, and Accountability	191
Elementary and Secondary Education	192
Setting Standards and Measuring Progress	193
Expanding Options	194
Providing for Vulnerable Populations: Government Partnerships....	196
Summing Up: Getting Incentives Right	199
Welfare	199
Focusing on Results	200
The Importance of Measurement	202
The Value of Incentives	203
The Benefits of Flexible Approaches	204
Encouraging Broad Participation	206
Medicaid and SCHIP	207
Limitations and Shortcomings of the Current System	208
Fostering Market-Based Health Insurance	210
Conclusion	213
chapter 6. building institutions for a better environment.....	215
The Government's Role in Environmental Protection	219
Measuring the Benefits and Costs of Environmental Protection.....	221
Types of Environmental Regulation	223
Command-and-Control Approaches	223
Standard Market-Based Approaches: Permit Trading and Fees.....	223

	<i>Page</i>
Other Flexible Approaches: Informal Markets and Tradable Performance Standards	226
Myths About Flexible Approaches	228
Case Studies in Flexible Environmental Protection.....	232
The Sulfur Dioxide Permit Trading Program	233
Tradable Quotas in the Alaskan Halibut and Sablefish Fisheries	237
Informal Permit Trading in the Tar-Pamlico River Basin	240
When Markets Don't Work	243
Lessons for Future Policy: Climate Change	244
Base Policy Action on Sound Science.....	245
Choose a Flexible, Gradual Approach.....	245
Set Reasonable, Gradual Goals	246
Provide Information and Encourage Reductions	248
Give Technology—and Institutions—Time.....	248
chapter 7. supporting global economic integration	251
The United States in the International Economy	252
Trends and Patterns in U.S. and World Trade.....	252
Trends and Composition of Capital Flows.....	260
The Benefits of Globalization.....	265
The Benefits of Trade.....	265
The Benefits of Capital Flows.....	268
The Role of Migration.....	270
Some Myths About Trade and Globalization.....	271
Trade and the Environment.....	271
Trade and Employment	272
Trade and Relative Wages	273
The Effects of Trade on Developing Nations	274
International Policy Issues and the Role of International Institutions.	275
International Trade Institutions and the Benefits of Trade	275
Role and Reform of International Financial Institutions.....	281
Conclusion.....	300
appendixes	
A. Report to the President on the Activities of the Council of Economic Advisers During 2001	301
B. Statistical Tables Relating to Income, Employment, and Production.....	313

list of tables

1-1. Administration Forecast.....	53
1-2. Accounting for Growth in Real GDP, 1960-2012	55
1-3. Accounting for the Productivity Acceleration Since 1995.....	61
7-1. U.S. Manufacturing Trade as Share of Shipments and Consumption, 2000.....	255
7-2. Estimated Gross Private Sector Capital Flows.....	261
7-3. Estimated Net Private Sector Capital Flows.....	263
7-4. Estimated World Cross-Border Claims and U.S. International Investment Position, Year-End 2000.....	264

list of charts

1-1. Real GDP Growth.....	24
1-2. Real Consumption Growth	27
1-3. Growth in Real Gross Private Domestic Investment	29
1-4. Standard & Poor's 500 Composite Stock Index	33
1-5. Growth in the Real Capital Stock	41
1-6. Consumer Sentiment.....	42
1-7. Discount Window Borrowing	49
1-8. Effective and Target Federal Funds Rates.....	50
1-9. Productivity Growth Around Business Cycle Peaks.....	60
2-1. Income Sources of Aged Households, 1998.....	70
2-2. Pension Plan Participants by Type of Plan	71
2-3. Ratio of Working-Age to Retirement-Age Persons	88
3-1. Announced Mergers and Acquisitions Involving U.S.-Headquartered Firms.....	104
3-2. Fraction of U.S. Mergers and Acquisitions Involving a Foreign Buyer or Seller	104
3-3. Industry-Funded Research and Development and Patents Granted for Inventions	131
3-4. Sales Revenue of Selected Prescription Anti-Ulcer Drugs.....	140
4-1. Health Care Expenditures.....	150
4-2. Expenditures on Components of Health Care	150
4-3. Mortality Rates for Coronary Heart Disease	181
4-4. Survival Rate After AIDS-Defining Infection	181
5-1. Children in Federally Supported Programs for the Disabled.....	199
5-2. The Changing Allocation of Welfare Funds in Six States.....	205
6-1. Emissions of Major Air Pollutants	217

	<i>Page</i>
6-2. Reductions in Average Ambient Concentrations of Major Air Pollutants, 1981-2000	217
6-3. Sulfur Dioxide Emissions and GDP in Canada, United Kingdom, and United States	218
6-4. Sulfur Dioxide Allowances Traded Between Economically Distinct Organizations	236
6-5. Emissions from Phase I Facilities in the Sulfur Dioxide Trading Program	236
6-6. Nutrient Loading by the Tar-Pamlico Basin Association	243
7-1. World Merchandise Trade Volume.....	253
7-2. U.S. Trade Relative to National Output	254
7-3. U.S. Trade by Sector in 2000.....	267
7-4. Import-Weighted Average Tariffs, 1999	277

list of boxes

1-1. Better Tools: Improving the Accuracy and Timeliness of Economic Statistics.....	24
1-2. Capital Overhang and Investment in 2001.....	39
1-3. Increased Security Spending and Productivity Growth	56
1-4. Is There Still a New Economy?	58
2-1. National Saving, Personal Saving, and Growth	74
2-2. Does Social Security Alter Retirement Behavior?	78
2-3. The Effect of Social Security on Income Distribution	82
2-4. The Effectiveness of Saving Incentives	95
3-1. A Co-Production Agreement and a Partial Equity Stake: Pixar and Disney	111
3-2. The Primestar Acquisition	122
3-3. Dynamic Competition in the Market for Prescription Anti-Ulcer Drugs.....	139
4-1. Managed Care: Good, Bad, or Somewhere in Between?	151
4-2. The Need for Good Risk Adjustment.....	156
4-3. Federal Employee Health Insurance Plans	169
4-4. The Puzzle of Geographic Variations in Medicare Expenditure	174
4-5. Survival Rates and Mortality Rates	183
5-1. Why Have Welfare Caseloads Declined?.....	201
5-2. The State Children's Health Insurance Program	209
5-3. Community Health Centers	213
6-1. Trends in National and International Environmental Quality	216
6-2. Environmental Fees in Other Countries	225
7-1. Vertical Trade and Production Sharing.....	256
7-2. Crisis and Restructuring in Ecuador	288
7-3. Elliott Associates versus Peru.....	296

Overview

The events of 2001 brought new challenges for the U.S. economy and for economic policy. The war against terrorism has increased the demands on our economy, and we must do everything in our power to build our economic strength to meet these demands. At the same time, we must take pains to ensure that the benefits of economic growth are shared as widely as possible, both within and beyond our borders.

Economic growth is not an end in itself. As it raises standards of living—consumption, in the language of economists—growth also provides resources that may be devoted to a variety of activities beyond the traditional marketplace. Growth can fund environmental protection, the work of charitable organizations, and many other activities of interest and value to the United States, other industrialized economies, and developing economies alike. These uses of our economic growth contribute to achieving the President's vision of “prosperity with a purpose.”

Restoring Prosperity

The economy entered 2001 growing slowly, and growth continued to decelerate through most of the year. After expanding at an annual rate of 5.7 percent in the second quarter of 2000, gross domestic product (GDP)—a standard measure of economy-wide production—began to falter later that year, and the weakness persisted into 2001. Some sectors stumbled into outright decline; for example, industrial production peaked in June 2000 and then entered a prolonged slump. After several quarters of increasingly weak growth, the terrorist attacks of September 11 tipped the economy into recession, the first in 10 years.

The economic difficulties that began in 2000 and continued through 2001 should not blind us to the fact that the outlook for the economy remains strongly positive. What matters most for long-term growth is improvements in productivity. Productivity growth in the United States accelerated during the second half of the 1990s, and economists generally believe that much of that faster productivity growth is permanent. New technology deserves much of the credit—but by no means all of it. Better, more efficient ways of doing business also contributed, and only a fraction of the many possible improvements have yet been made. Our economic challenge is, in large measure, to discover how to reap the benefits of the remainder.

The United States is unique among industrial economies in having experienced this recent boom in productivity growth. In principle, nothing prevents businesses in all of the world's industrial and industrializing economies from adopting the same technologies available here. Yet only the United States has enjoyed an increase in sustained productivity growth since 1995. This stronger productivity performance therefore likely derives from uniquely American advantages: notably, the strength of our institutions and the flexibility of our business culture. Accordingly, this *Report* focuses on those institutions and on that culture, and proposes strategies for improving them and putting them to use, to sustain our growth and broaden our prosperity.

The *Report* begins, in Chapter 1, by reviewing the important economic events of 2001. The chapter goes on to present the economic outlook for the United States and to sketch an agenda for the institutions needed to speed the Nation's growth and enhance its economic security.

Strengthening Retirement Security

No area of American life could benefit more from enhancements to its institutional underpinnings than retirement security, and the President has made the reform of the Social Security system a central part of his economic agenda. As he has stressed, "Ownership in our society should not be an exclusive club. Independence should not be a gated community. Everyone should be part owner in the American Dream."

Chapter 2 of this *Report* examines the changing nature of retirement security and the institutional changes needed to meet this challenge. There is little dispute about the need for reform, and there is growing agreement that personal accounts within the Social Security system are an indispensable part of any reform plan. Personal accounts would enhance individual choice—the very foundation of the success of our market economy. The current Social Security system collects 12.4 percent of all covered wages and essentially constrains all working Americans to place that sizable share of our wealth in a single entity—one that demographic change is rendering increasingly inadequate to support the system's obligations.

Personal accounts would permit individuals to diversify their retirement portfolios, thus increasing their retirement security. They would for the first time acquire rights of ownership, wealth accumulation, and inheritance within Social Security. These advantages are widely recognized. Less well appreciated, however, is that ownership and inheritability will enhance Social Security's role in making our economic system more equitable. Some groups in our society with lower average incomes also have lower life expectancies, and as a consequence, they benefit less today from Social Security than do other, wealthier groups. Under a system of personal accounts, the early death of a worker would no longer mean the loss to that worker's heirs of much of

what he or she has paid into Social Security. Instead, those assets could be passed on to the next generation. For all these reasons, personal accounts are an important part of reforming Social Security, and thereby of strengthening retirement security for all Americans.

Realizing Gains from Competition

One source of the United States' superior economic performance over the past decade has been the success of its institutions for promoting open, competitive markets. Strong incentives to compete are what drive firms to exploit new opportunities, and so achieve faster growth throughout the economy. Deregulation of several key industries during the 1970s and 1980s brought substantial benefits to consumers and to the economy as a whole—however, it took time for all of those benefits to be realized, and this counsels patience in evaluating more recent deregulation initiatives in, for example, electricity markets.

The task of competition policy—as detailed in Chapter 3 of this *Report*—is to promote competition in a way that ensures the efficient allocation of resources and serves the interests of consumers. In doing so, however, competition policy must walk a fine line: efforts to prevent anticompetitive changes in the behavior and organization of firms may inadvertently keep firms from taking steps that could lower their costs or improve their products. Such ill-advised interventions would ultimately harm consumers rather than benefit them.

The recent past has witnessed a remarkable shift in the competitive landscape. Mergers and acquisitions have reshaped and continue to reshape the organization of firms and the nature of competition itself. Our competition policy must be flexible enough to acknowledge and support the quest for efficiency that drives these changes, while remaining vigilant against efforts to restrain competition. To fail in this task would be to hinder the growth of innovative firms, the adoption of new technology, and the enhancement of productivity.

The markets in which American firms compete today are increasingly global markets, and globalization motivates further changes in firms' organization. Our competition policy should acknowledge and reflect these motivations. But other countries have their own competition policies, and inefficient policies in any one of them may impose costs on firms and consumers in the United States and around the world. The United States should therefore pursue the harmonization of national competition policies—but should do so in a way that spreads best-practice, efficient competition policy worldwide.

Finally, competition policy must also deal with the increased importance of “dynamic competition,” in which firms compete not just for increments of market share but for absolute (if temporary) market dominance, through rapid innovation. Policies should recognize that, at any given moment, high profits and substantial market share—indicators that might warrant concern about competition in some industries—need not preclude vigorous dynamic competition among firms in industries undergoing rapid technical change.

Promoting Health Care Quality and Access

Health care is one of the largest and most vibrant sectors of the economy. Biomedical research, both public and private, has generated stunning advances in our understanding of biology and disease and achieved major therapeutic discoveries. As a result, Americans today are living longer lives with less disability. However, the health care delivery system today is troubled, as medical expenditures are again rising rapidly. The costs of private health insurance to working Americans and the costs to taxpayers of government health programs, including Medicare and Medicaid, are increasing at rates far surpassing the growth of the economy. Managed care is under fire from patients and physicians alike. With the economic slowdown and rising costs, concerns about the growing number of uninsured are again coming to the fore.

Much of the discussion about Federal policies to address these concerns has been framed through a narrow lens that focuses on “guarantees” for access and treatment, to be achieved largely through expanding government programs that rely on regulation and price setting. Yet this approach does not ensure access to innovative care that meets the diverse needs of patients in an efficient way.

Chapter 4 of this *Report* explores an alternative framework, one that focuses on achieving better health care through solutions that emphasize both shared American values and sensible economics. These solutions build on existing support; they encourage flexible, innovative, and broadly available health care coverage; they emphasize the central role of the patient in making health care decisions; and they improve those decisions by creating an environment for medical practice that encourages steps to improve quality and reduce costs. This approach emphasizes patient-centered health care, with individual control and individual responsibility.

If we move toward a system of informed choice and well-crafted economic incentives, and away from rigid regulation, the health care system will benefit from the resulting flexibility and competition. In this vision, government support would be used to broaden access and to encourage competition in both the private and the public sectors. Support should be targeted to improving the health care of those most in need: the uninsured and those

with significant health expenses. New incentives should strengthen the market by improving information about quality and cost, broadening choice, rewarding quality, and addressing costs by encouraging value purchasing by both employers and patients.

The Administration's emphasis on patient-centered health care reform centers on three objectives. First, we must develop flexible, market-based approaches to providing health care coverage for all Americans. Second, we must support health care providers in their efforts to meet the demand for higher quality and value, in part by making better information available about providers, options, outcomes, and costs. And finally, we must provide the foundation for further innovation through strong support for biomedical research. Providing competitive choices for all Americans, and meaningful individual participation in those choices, will encourage innovation in health care delivery and coverage. Improving incentives and information, and taking steps to help patients and providers use information effectively, will help ensure continued improvements in the health of Americans in the future.

Redesigning Federalism for the 21st Century

Throughout its history the United States has relied heavily on State and local governments to provide certain goods and services. Our federal system has been a source of greater efficiency and of innovation in government practice. History reveals several tensions as well, most vividly evidenced by Washington's all-too-frequent practice of providing funds to State and local governments without allowing flexibility in their use. As discussed in Chapter 5 of this *Report*, this tension between flexibility and control can be resolved efficiently by specifying standards for outcomes but leaving it to State and local providers to determine how best to achieve those outcomes.

Focusing on outcome standards and flexibility to improve efficiency can also imply a role for the private sector in providing public services. The choice of where to draw the line between the public and the private sector depends on the characteristics of the services to be provided. The nature of some services makes it difficult for markets to meet the needs of the population effectively. Even then, it may be efficient to rely on the private sector to *produce* the service, but to let State and local governments decide what and how much shall be provided.

Chapter 5 of this *Report* discusses the principles underlying the roles of differing levels of government, and of for-profit firms and not-for-profit organizations, in identifying and meeting needs for public goods and services. Specifically, the chapter shows how allowing public and private organizations to compete in meeting preset standards can improve the efficiency of programs in education, welfare, and health insurance for needy populations.

In education, evidence supports the benefits of competition in improving quality, with public, private, and charter schools vying with each other to provide the best education most efficiently. When the right institutions are in place, school systems can be held accountable for results. Similarly, the providers of safety net benefits—such as welfare and Medicaid—must be accountable to taxpayers for the quality of services they provide and the resources they use to provide them. By tying payments to these providers to results, and by allowing private nonprofit providers to compete with them on an equal footing, the market discipline that yields innovation and efficiency in the private sector can be brought to bear in the public sector as well.

Building Institutions for a Better Environment

Not so long ago, environmental protection and market-based economic growth were widely regarded as fundamentally in conflict. The past 30 years, however, have seen dramatic improvements in environmental quality go hand in hand with robust growth in GDP. Releases of many toxic substances have been reduced, and many of our natural resources are better protected. Rivers are cleaner and the air is clearer.

In many of these early environmental interventions, the anticipated benefits were clear, large, and achievable at relatively low cost. The next generation of environmental issues, however, is certain to be more challenging. Ongoing efforts to protect endangered species, maintain biodiversity, and preserve ecosystems will require tradeoffs between the welfare interests of current and future generations. But those early initiatives also taught us that the costs of environmental protection can be minimized through careful policy design. Part of the challenge for environmental protection today is to identify the best institutions to address each of an array of stubborn environmental problems. Another part is to design those institutions so that they can evolve to address new problems in the future.

Chapter 6 of this *Report* describes how flexible, market-based approaches to environmental protection—using tradable permits, tradable performance standards, and similar mechanisms for a fixed overall standard—allow businesses to pursue established performance goals or emission limits in the manner they find most efficient. The chapter documents, through several case studies, that such an approach can often achieve equal or greater environmental benefits at lower cost than one based on inflexible government mandates. The chapter concludes by illustrating how—and how not—to apply this experience with flexible mechanisms to the long-term challenge of global climate change.

Supporting Global Economic Integration

The final chapter of this *Report* examines our institutions for international trade and finance. International flows of goods, services, capital, and people have played an increasingly important role in the world economy, raising the standard of living in the United States and around the world. These gains from international interaction stem from an improved allocation of resources. A more efficient global allocation of productive inputs such as capital and labor translates into higher global output and consumption. Today, however, signs of a slowing global economy, and threats to the freedom that is part and parcel of a well-functioning economic system, make it more important than ever to rededicate ourselves to the free exchange of goods, services, and capital across borders.

It is therefore critical that the United States continue to lead the world in the liberalization of trade. The restoration of the President's Trade Promotion Authority (TPA) will provide the Administration the flexibility and the bargaining power to promote this liberalization most effectively. By streamlining the system for approving trade agreements, TPA will allow the United States to keep pace with our trading partners in the timely adoption of trade liberalization.

The United States must also continue to encourage efforts to strengthen the international financial architecture. A stronger global financial system is needed to support the cross-border flows of capital that are vital to increasing world output. The Administration is taking the lead in the debate over principles for reform of international lending by the International Monetary Fund and the World Bank. In addition, the Administration is seeking to shift the multilateral development banks' emphasis toward grants for low-income countries: this is consistent with continued efforts to make these institutions more efficient and more focused on growth in living standards in developing countries. U.S. leadership in this area is essential to safeguarding and enhancing both our own economic prospects and those of the rest of the world.

Conclusion

The past year has shown that we cannot be complacent about America's rate of economic growth, gains in productivity, and successes in global markets. Nor can we afford to be parochial. We seek growth and prosperity for the whole world, and we will achieve it by wise economic policy and farsighted institutional reform.

Restoring Prosperity

Over the past two decades, the Nation has witnessed an impressive increase in prosperity. Over 35 million jobs were created, and real income nearly doubled, producing an unprecedented standard of living. This economic success also serves as an example of what an open, free market economy—one that relies on the private sector as the engine of growth—can achieve.

A hallmark of the economy has been its ability to weather adverse economic developments in a flexible and resilient manner. This is not an accident but rather a characteristic of an economic system that relies on market forces to determine adjustments in economic activity. But such an economy, even in the presence of sound fiscal and monetary policies, is not immune to business cycles. Economic activity in 2001 is an example of how a series of adverse developments can cause setbacks on the road to greater prosperity. The last year also highlighted the value of continued efforts to strengthen the policy environment in a way that allows the private sector both to recover more quickly and flourish more strongly in the future.

Macroeconomic Performance in 2001: Softer Economy, Harder Choices

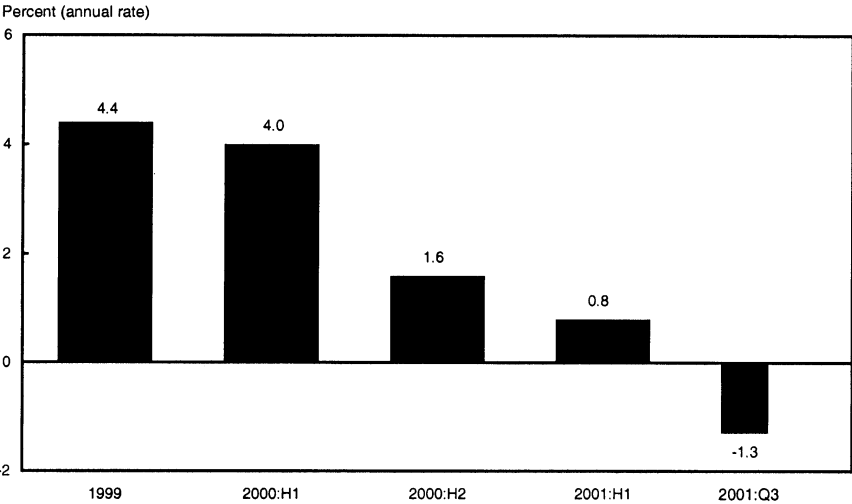
U.S. economic growth continued to decelerate during 2001. It was apparent early in the year that policymakers would face considerable challenges as the rate of growth slowed from the rapid rates of past years. The momentum placing downward pressure on economic activity appeared to subside by midsummer, however, by which time growth of real gross domestic product (GDP) had come to a virtual standstill. Economic conditions showed some tentative signs of firming, and growth prospects were brightening. All that changed on September 11. The President, Congress, and other policymakers responded decisively to the damage and disruptions caused by the terrorist attacks, while continuing to work to strengthen the long-run economic fundamentals.

Aggregate Demand During the First Three Quarters

The deceleration of real GDP in 2001 continued a slowdown in economic activity that had begun the previous year (Chart 1-1). Real GDP growth over the first three quarters remained barely positive, at 0.1 percent on an annualized basis; however, the economy steadily weakened through this period,

ending with a 1.3 percent annualized contraction in real GDP in the third quarter. Although several key components of aggregate demand rose moderately, overall growth was dragged down by unusually weak investment spending. Preliminary evidence indicates a further decline in the fourth quarter due to weaker economic conditions—especially during the early months of the quarter—in the aftermath of the September terrorist attacks. This assessment, however, may be subject to large revision because of the limitations of existing statistical sources (Box 1-1).

Chart 1-1 Real GDP Growth
The economy has been decelerating since mid-2000.



Note: Growth is measured to the final quarter of the indicated period from the final quarter of the preceding period. H1 and H2 denote the first half and second half of the year.

Source: Department of Commerce (Bureau of Economic Analysis).

Box 1-1. Better Tools: Improving the Accuracy and Timeliness of Economic Statistics

Economic statistics are valuable tools that economists, policy-makers, business leaders, and individual investors use to increase our understanding of the economy. The Bureau of Economic Analysis, the Bureau of Labor Statistics, the Bureau of the Census, the Federal Reserve, and other departments and agencies combine thousands of bits of information from market transactions, consumer and business surveys, and numerous other sources to produce scores of economic estimates every month.

continued on next page...

Box 1-1.—*continued*

The ability of government, consumers, workers, and businesses to make appropriate decisions about work, investments, taxes, and a host of other important issues depends critically on the relevance, accuracy, and timeliness of economic statistics. At turning points in the economy, such as those marking the beginning or the end of an economic slowdown, the accuracy and timeliness of data are especially critical, because at these times fiscal and monetary policy can be most useful in steering the economy.

Recent economic events have emphasized the importance of timely economic information. Thus one area deserving considerable attention is the need for readily accessible real-time data. Investment in sources of these data could yield handsome dividends, especially at key junctures in the business cycle.

Moreover, the quality of existing statistics is far from perfect and could be enhanced with further investment. Even real GDP, generally thought of as a reliable measure of overall activity in the U.S. economy, is susceptible to considerable revisions. For example, in the third quarter of 2000, real GDP was first estimated to have grown 2.7 percent at an annual rate—a subpar but respectable growth rate. That rate was then revised downward to 2.4 percent and then again to 2.2 percent. Seven months later it was further revised downward to 1.3 percent, providing evidence that the economy had begun to slow dramatically at that time. A key component of the revision came from revised data on gross private domestic investment, initially estimated to have risen 3.2 percent but later revised to show a contraction of 2.8 percent. Such revisions lead to uncertainty for both government and private decisionmakers, which can cause costly delays. Although most revisions are not that large, the average quarterly revision of real GDP growth over 1978-98 was about 1.4 percentage points in either direction, while real GDP growth averaged 2.9 percent.

In addition to these problems with large revisions, the national accounts statistics are beset by some growing inconsistencies. Gross domestic product, the sum of final expenditures for goods and services produced by the U.S. economy, and gross domestic income, the sum of the costs incurred and income received in the production of those goods and services, are theoretically equal. Because of statistical discrepancies, there has always been some divergence between these two reported numbers. However, this discrepancy has been growing lately, raising concerns among policy experts and business leaders as well as among the producers of the data themselves. These differing estimates can lead to different readings of such critical indicators as output and productivity growth.

continued on next page...

Box 1-1—*continued*

A number of steps can be taken to improve the accuracy and timeliness of economic statistics. In particular, targeted improvements to the source data for the national accounts would go a long way toward illuminating the causes of the growing statistical discrepancy. Another cost-effective measure would be to ease the current restrictions on the sharing of confidential statistical data among Federal statistical agencies. Such data sharing, which would be done solely for statistical purposes, is currently hindered by lack of a uniform confidentiality policy. Confidentiality is of key importance to all agencies and to the individuals and businesses who participate in Federal surveys, but a uniform confidentiality policy would allow agencies such as the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the Bureau of the Census to cost-effectively compare and improve the quality of their published statistics while preserving confidentiality. In the past, attempts have been made to pass legislation, together with a conforming bill to modify the Internal Revenue Code, allowing such data sharing under carefully crafted agreements between or among statistical agencies. In 1999 such legislation passed the House but stalled in the Senate. The Administration will continue to seek passage of data sharing legislation to improve the quality and effectiveness of Federal statistical programs.

In addition to data sharing legislation, the Administration is proposing new and continued funding for the development of better and more timely measures to reflect recent changes in the economy. For example, these resources would allow for tracking the effects of the growth in e-commerce, software, and other key services, and for developing better estimates of employee compensation. The latter are increasingly important given the expansion in the use of stock options as a form of executive compensation, as well as for tracking the creation and dissolution of businesses, given the importance of business turnover in a constantly evolving economy. Improved quality-adjusted price indexes for high-technology products are also an important area for future research. The direct contribution of these products accounted for nearly a third of the 3.8 percent average annual growth rate in real GDP during 1995-2000, but current estimating techniques fail to capture productivity growth in high technology-using service industries. This shortcoming may lead to underestimates of annual productivity growth of 0.2 to 0.4 percentage point or more. As the economy continues to change and grow, the need persists to create and develop such new measures, to provide decisionmakers with better tools with which to track the economy as accurately as possible.

Consumption

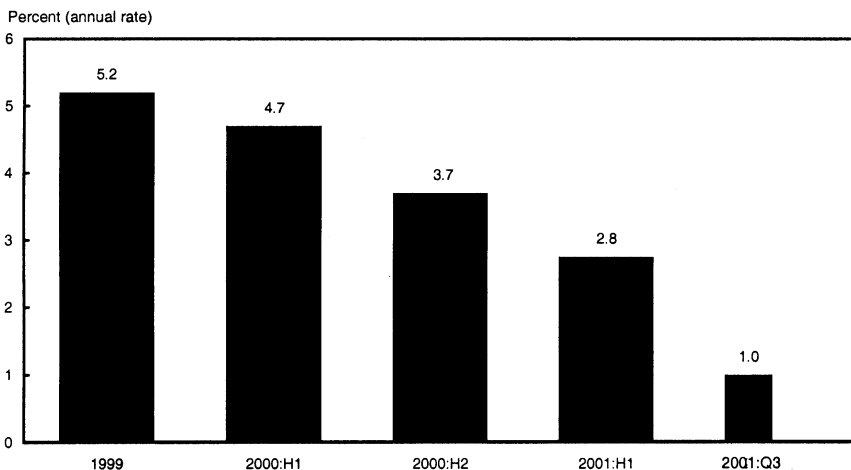
Personal consumption expenditures grew 2.8 percent at an annual rate in the first half of 2001, followed by a 1.0 percent increase in the third quarter (Chart 1-2). Consumption growth in the first three quarters was 2.2 percent—notably slower than the 4.8 percent rate of the previous 3 years.

Spending for all types of consumption slowed in 2001. Growth in spending on nondurable goods declined to a 1.1 percent annual rate through the third quarter, from a 4.5 percent rate in 1998-2000. The sharp decline in nondurable consumption is somewhat surprising, because swings in this category of consumption tend to be more muted than those in overall consumption. Consumption of food and of clothing and shoes decelerated sharply, in a significant deviation from recent trends. The Bureau of Economic Analysis estimates that food consumption edged down 0.4 percent in the first three quarters of 2001, after averaging 3.8 percent growth in the previous 3 years; clothing and shoes consumption rose 1.9 percent after averaging nearly 7 percent growth in 1998-2000. Energy consumption continued to be weak, reflecting higher energy prices early in the year.

Growth in durable goods spending also subsided, but remained relatively strong, in the first three quarters of 2001: purchases rose 6.1 percent at an annual rate compared with 9.7 percent on average in 1998-2000. This recent strength has been atypical because, during most economic downturns,

Chart 1-2 Real Consumption Growth

Consumer spending has been slowing since mid-2000, but remained positive in 2001 despite the contraction in overall economic activity.



Note: Growth is measured to the final quarter of the indicated period from the final quarter of the preceding period. H1 and H2 denote the first half and second half of the year.

Source: Department of Commerce (Bureau of Economic Analysis).

durable goods spending tends to slow more sharply than nondurable goods spending. Part of the explanation is that two key durable goods industries have proved more resilient to the slowdown than in the past. Furniture and household equipment grew robustly, as the housing sector stayed healthy in 2001. And although growth in sales of motor vehicles and parts was anemic early in the year, these sales remained remarkably high for a period of such marked slowing in overall activity.

Finally, consumption of services—the least cyclical component of consumption—grew at a 1.9 percent annual rate in the first three quarters of 2001, down from a 4.0 percent rate over 1998-2000. Medical care spending, however, continued its strong upward trend.

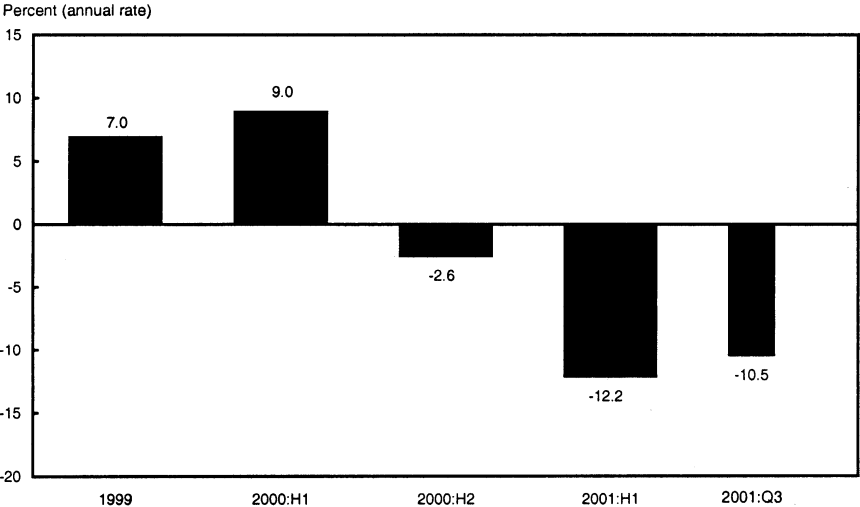
These patterns in consumption spending—which constitutes two-thirds of GDP—reflected several key economic crosscurrents. On the downside, the decline in equity markets and the deterioration in labor markets (discussed below) reduced wealth and consumer confidence. On the upside, housing prices continued to climb, rising at roughly an 8 percent annual rate. In addition, lower mortgage interest rates sparked the strongest wave of home refinancing ever, transforming housing equity into more liquid forms of wealth. Refinancing is estimated to have increased household liquidity (from increased cash flow and cashouts) by about \$80 billion during the year. In addition, real disposable personal income, aided somewhat by provisions of the President's tax cut—reduced withholding and the payment of rebates for the new 10 percent personal income tax bracket—rose at a solid 4.5 percent annual rate during the first three quarters.

Investment Spending

Real gross private domestic investment fell at a double-digit annual rate (roughly 12 percent) in each of the first two quarters of 2001—the steepest decline in investment spending in a decade (Chart 1-3). The year began with a sizable inventory liquidation, which accounted for most of the decline in gross private domestic investment in the first quarter and subtracted 2.6 percentage points from the growth rate of real GDP. Inventory reduction remained a drag on GDP growth in subsequent quarters, with manufacturing industries shedding inventories at a faster pace than wholesalers and retailers. By the end of the third quarter, the inventory-to-sales ratio had returned to a level close to the average over the previous 3 years, indicating that the downward phase of the inventory cycle may soon be ending.

Nonresidential business fixed investment contracted sharply in 2001, in stark contrast to the investment boom from 1995 to early 2000. In the first quarter this category of investment fell at only a 0.2 percent annual rate—the first decline in 9 years. In the second quarter, however, it fell at a 14.6 percent annual rate, with declines in investment in structures and in equipment

Chart 1-3 Growth in Real Gross Private Domestic Investment
The pronounced deceleration in investment spending since mid-2000 led the economic slowdown.



Note: Growth is measured to the final quarter of the indicated period from the final quarter of the preceding period. H1 and H2 denote the first half and second half of the year.

Source: Department of Commerce (Bureau of Economic Analysis).

and software of 12.3 percent and 15.4 percent, respectively. Investment in information processing equipment and software alone fell at a 19.5 percent rate in the second quarter. The widespread decline in business fixed investment continued in the third quarter with an 8.5 percent contraction, combining a 7.6 percent drop in structures investment with an 8.8 percent decline in equipment and software spending. Capital spending on computers and peripherals during the second and third quarters was hit particularly hard, plunging at a 28.6 percent rate.

The housing sector was a bright spot in 2001. Lower mortgage rates and rising real income helped to support rising residential investment in each of the first three quarters; growth for the period averaged 5.6 percent at an annual rate. Investment in single-family structures rose 6.0 percent, after declining during most of 2000. Investment spending on multifamily structures rose briskly at a 15.3 percent rate. Investment in residential building improvements increased at a 3.2 percent rate.

Government Spending

Government spending—Federal, State, and local levels combined—added to economic activity over the first three quarters of the year. Federal Government spending increased at a 2.9 percent annual rate during this

period. In contrast, Federal spending in 2000 fell by 1.4 percent, and over 1995-2000 it grew at only a 0.1 percent average rate. Last year's increase was driven by national defense expenditure, which rose 4.4 percent through the first three quarters. Defense spending on research and development as well as personnel support accounted for most of the increase. Nondefense expenditure grew only 0.2 percent in the first three quarters of 2001.

State and local government spending increased 3.8 percent at an annual rate in the first three quarters. State and local spending has increased steadily over the past decade, averaging 2.8 percent annual growth from 1990 to 2000 and 3.2 percent from 1995 to 2000. Investment by State and local governments rose much faster (4.6 percent a year on average) than their consumption (2.8 percent) during 1995-2000. However, consumption expenditure accounts for 80 percent of State and local spending.

Net Exports

Net exports exerted a smaller drag on economic activity in 2001 than in 2000. Both imports and exports fell significantly during the year, but the drop in imports was larger. Real exports of goods and services, measured at an annual rate, declined \$95.3 billion through the third quarter, mostly because of a decline in exports of capital goods—especially high-technology goods—as a result of the global economic slowdown (discussed further below). Over the same period, real imports declined \$105.3 billion. Real imports of services suffered one of the largest declines on record in the third quarter, largely because international travel was disrupted in September.

Overall, net exports contributed 0.1 percentage point to real GDP growth in the first three quarters of the year. By comparison, in 2000 net exports depressed real GDP growth by 0.8 percentage point.

Preliminary Evidence on Aggregate Demand in the Fourth Quarter

The terrorist attacks of September 11 changed the direction of the macroeconomy. Before the attacks, the economy had been showing tentative signs of stabilizing after its long deceleration, and many forecasters expected real GDP growth to accelerate in the third and fourth quarters of 2001. Immediately after the attacks, however, the economy turned down because of the direct effect of the assault on the Nation's economic and financial infrastructure and because of the indirect, but more significant, effect on consumer and business confidence. The drop was sufficient to turn the sluggish period of economic activity into a recession.

The disruptions to lower Manhattan's telecommunications and trading facilities temporarily interfered with the normal operations of key components

of the Nation's financial center and caused dislocations in the Nation's payment system, which processes trillions of dollars in transactions on a typical business day. Equity markets shut down temporarily, and when they reopened a week later, the value of shares fell by \$500 billion. Money markets and foreign exchange markets continued to function during this period but faced considerable difficulties.

In the New York City area, the closure of much of lower Manhattan weakened economic activity, especially employment, and had serious consequences for local businesses that depend on sales from that part of the city. The local tourism and business travel industries also sagged. The attack on the Pentagon had less of a direct effect on the private sector because of the limited destruction of private infrastructure. Nonetheless, economic activity in the Washington, D.C., area slumped, primarily because of the need to temporarily close Reagan National Airport for national security reasons. Local businesses, such as hotels and restaurants, that provide ancillary services for travelers were hit particularly hard. As in the New York City area, small businesses were especially affected, because many operate from only one business location, whereas large businesses with operations throughout the country are often better able to weather local dislocations.

The terrorist attacks also had a significant macroeconomic effect. The Nation's airspace was shut down for several days after the attacks, halting passenger travel and deliveries of airfreight. In addition, cross-border ground shipping was delayed because of increased security measures. Businesses that rely on highly synchronized deliveries of inputs were forced to slow down their assembly lines, and in some cases close plants, creating disruptions up and down the stream of production.

Beyond the initial impacts, the attacks continued to have a significant negative effect on the economy as uncertainty about the future led to a steep decline in consumer and business spending. Consumers retrenched as they mourned the loss of life and reevaluated the risks inherent in even the most mundane activities, such as shopping at malls and traveling by air. Meanwhile businesses adopted a more pessimistic outlook about the prospects for a speedy recovery. The underlying psychology was affected again in October, by the discovery of anthrax spores delivered through the mail distribution system, although the direct macroeconomic effects of this attack have been fairly limited.

Preliminary evidence indicates that economic activity at the beginning of the fourth quarter of 2001 suffered a pronounced decline. The industrial sector contracted at a faster pace in October than earlier in the year, and job losses mounted. By November, however, some tentative signs had emerged that business conditions were deteriorating at a slower pace. For example, the decline in industrial production was milder, and nondefense capital goods

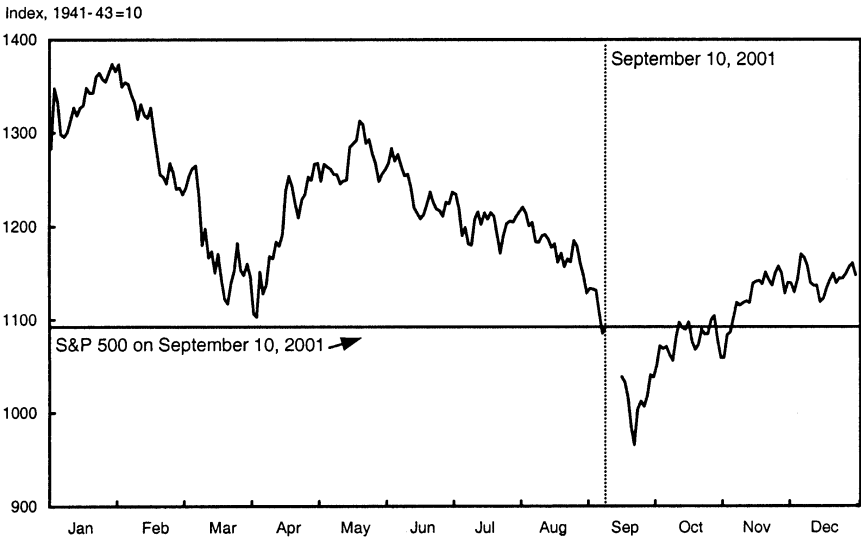
spending appeared to have bottomed out, with new orders recovering from the trough in September. Construction spending also performed well, as weather in the fall was unseasonably warm. By December the manufacturing sector, which had been particularly hard hit in 2001, witnessed increases in the length of the average workweek and in factory overtime. Meanwhile the Purchasing Managers' Index (PMI) of the Institute for Supply Management (formerly the National Association of Purchasing Management) rebounded sharply, with a jump to 48.2 in December from 39.8 in October. The production component of the PMI rose to 50.6 from 40.9 in October; the new orders index surged to end the year at 54.9. Moreover, industrial production in December was nearly unchanged after several months of sizable declines.

Despite the initial dropoff in consumer confidence after the terrorist attacks, consumer spending bounced back within the quarter from its September plunge. Real personal consumption expenditures on durable goods, nondurable goods, and services rose considerably in October and November. Purchases of automobiles and light trucks contributed substantially to the rebound, as consumers responded favorably to the incentive programs offered by manufacturers and dealers, such as zero-percent financing and rebates. Automobile and light truck sales surged to a record 21 million units at an annual rate in October, then moderated to something closer to the average 17-million-unit selling pace of the first three quarters. Even though nominal retail sales of goods excluding motor vehicles edged down in November and December, falling prices for energy and consumer goods suggest that real consumption spending continued to rise.

The performance of financial markets confirmed the view that economic conditions were firming in the fourth quarter. Stock market prices rebounded from a sharp decline after September 11 (Chart 1-4). The Standard & Poor's 500 Composite Stock Index had returned to its pre-September 11 level by mid-October, and it ended the year near 1150, up 19 percent from its post-September 11 low. Other market indexes such as the Dow Jones Industrial Average and the Wilshire 5000 rose in a similar pattern. In addition, credit markets were active in providing funds to businesses. Low interest rates made bond financing attractive, especially for investment grade issuers. Lending by commercial banks for real estate and consumer purchases was rising and generally higher in the fourth quarter than earlier in the year. Commercial and industrial lending, in contrast, was lower in the quarter than earlier. According to the Federal Reserve, banks tightened credit standards and terms on commercial and industrial loans by late summer and early autumn. The tightening of non-price-related loan terms was especially apparent for small firms.

Chart 1-4 **Standard & Poor's 500 Composite Stock Index**

Stock prices generally declined in 2001, with a precipitous drop after the terrorist attacks on September 11. Stock prices returned to pre-attack levels by mid-October.



Note: The New York Stock Exchange was closed on September 11 and reopened on September 17.

Source: Standard & Poor's.

Labor Markets

Private nonfarm payrolls dropped by roughly 1.5 million in 2001, reflecting the weak economy. The bulk of the decline occurred in manufacturing, especially in durable goods-producing industries, where over 1 million jobs were shed after December 2000. In addition, employment in help supply services, which provide labor to other industries, fell by about 550,000 jobs. Job losses in manufacturing and help supply services were offset in part by increases in some other service industries during the year. The health services industry logged strong increases in 2001. In recent months, service employment has been hurt by cutbacks in business travel and tourism, which have adversely affected employment in air transportation and travel-related services such as travel agencies, hotels, and amusements and entertainment.

Labor markets became substantially less tight in 2001. The total unemployment rate rose from 4.0 percent in December 2000 to 5.8 percent a year later, still below the average rate for the past 20 years of 6.2 percent. The average duration of unemployment rose by 2 weeks during 2001, ending the year at 14.5 weeks. More than half of this increase occurred in the last 3 months of 2001.

Every region saw its unemployment rate rise, as the slowdown in economic activity was national in scope. The Mountain States experienced the largest increase, 1.8 percentage points. The smallest increase occurred in the West North Central States; this region had one of the lowest unemployment rates in the country at the end of 2000.

The labor force participation rate (the share of the working-age population either working or seeking work) fell 0.4 percentage point over the year. Labor force participation has hovered near 67 percent since 1997, after rising from near 60 percent in 1970. The average number of discouraged and displaced workers has risen nearly 30 percent since the beginning of 2001 but remains below the average for the past 5 years.

Inflation

Inflation remained low and stable in 2001. The consumer price index (CPI) rose only 1.6 percent during the 12 months ending in December. Consumer energy prices for fuel oil, electricity, natural gas, and gasoline tumbled 13.0 percent, reflecting a collapse in crude oil and in wellhead natural gas prices. In contrast, energy price inflation a year ago was 14.2 percent. Food prices rose 2.8 percent, the same rate as a year ago. The CPI excluding the volatile food and energy components—often referred to as the core CPI—posted another year of stable inflation. Core inflation was 2.7 percent, up somewhat from its 2.3 percent average rate over the past 4 years.

The absence of price pressures in the production pipeline helped hold consumer price increases in check. The producer price index (PPI) for finished goods fell 1.8 percent in the 12 months ending in December. At the start of the year, producer prices had been rising rapidly, largely reflecting rising energy prices; but PPI inflation fell all year long as energy prices slumped and economic activity weakened. Excluding the volatile energy and food components, the PPI for finished goods rose 0.7 percent during 2001. PPI inflation for intermediate and crude materials declined throughout the year, sometimes experiencing periods of steep price declines.

Productivity and Employment Costs

Despite the economic slowdown, nonfarm business labor productivity grew at a 1.2 percent annual rate during the first three quarters of the year. Although below the 2.4 percent average rate recorded during 1995–2000, productivity growth has been remarkably strong for this stage of the business cycle. During previous postwar recessions, productivity growth averaged 0.8 percent.

Manufacturing productivity, in contrast, edged down at a 0.2 percent annual rate for the first three quarters of the year, compared with a 0.6 percent

decline in the 1990-91 recession. The 2001 figure represents the first decrease in manufacturing productivity in the past 8 years, and it reflects the pronounced slump in the industrial sector that began in mid-2000. A sharp deceleration in durable manufacturing productivity from a nearly 7 percent rate of growth in 2000 to a 0.8 percent rate of decline during the first three quarters of 2001 accounted for much of the change. Nondurable manufacturing productivity grew at only a 0.1 percent rate over the first three quarters of 2001.

Employment costs rose at a slower rate in 2001 than in 2000. Total wages and salaries for private workers as measured by the employment cost index (ECI) rose 3.7 percent at an annual rate through the first three quarters of 2001—slightly less than the 3.9 percent increase in 2000. The total cost of benefits for private industry workers increased at a 5.1 percent rate through September 2001, down from a 5.7 percent increase in 2000. The ECI for manufacturing rose 3.3 percent, combining a 3.8 percent rise in wages and salary with a 2.7 percent increase in benefit costs. This slowdown in the rate of employment cost increases should help to moderate future inflationary pressure.

Saving and Investment

National saving, which comprises private saving and government saving, fell in 2001. As a share of gross national product, national saving edged down to 17.2 percent during the first three quarters of 2001 from 17.9 percent in 2000. Shrinking Federal Government saving accounted for most of the decline, as the economic slowdown reduced revenue and caused some types of automatic expenditure to rise. The personal saving rate (personal saving as a share of disposable income) averaged 2 percent in the first three quarters of 2001, up from 1 percent in 2000. Part of the increase was due to the downpayment on the President's tax cut, which was sent out in the form of "rebate" checks in July through September. Although the personal saving rate rose in the third quarter, Federal Government saving declined, the natural consequence of returning surpluses to taxpayers.

As the current account deficit shrank with the slowing economy, net foreign investment flows slowed in 2001. As a result, despite the decline in the national saving rate, domestic sources of saving funded a larger share of domestic investment. Over the previous 3 years, net foreign investment had been growing by roughly \$100 billion a year. After reaching a peak of just over \$450 billion in 2000, net foreign investment fell steadily in 2001, its first decline since 1997. By the third quarter, net foreign investment had dropped to \$355 billion, although this was exaggerated somewhat by the one-time insurance payment of roughly \$40 billion (at an annual rate) from foreign sources on claims (recorded on an accrual basis) related to the terrorist attacks.

National saving and investment are key to our long-run prosperity, and the President's 2001 fiscal initiatives improved incentives for private saving and investment. Because budget resources ultimately depend on the health of the economy as a whole, this approach serves as the best way to enhance budget surpluses over the long run.

In June the President signed the Economic Growth and Tax Relief Reconciliation Act (EGTRRA, described in more detail later in this chapter), which removes impediments to private saving by expanding contribution limits for Individual Retirement Accounts (IRAs), 401(k) plans, and education savings accounts. Education savings accounts raise incentives not only to save for education, but also to improve the quality and productivity of the Nation's work force in the future. Other provisions of the act, such as lower marginal tax rates, a reduced marriage penalty, and elimination of the estate tax, provide strong incentives to work, save, and invest. Another important initiative is the President's Commission to Strengthen Social Security, which in December issued its final report on meaningful reform options to strengthen the Social Security system and improve the ability of individuals to accumulate and pass along wealth.

The Cyclical Slowdown

Several factors contributed to the deceleration in economic activity during 2000 and 2001 from its very high levels in the preceding years: the decline in stock market wealth, the spike in energy prices, an increase in interest rates, the collapse of the high-technology sector, and the lingering effects of preparations against the year-2000 (Y2K) computer bug. With this backdrop setting the stage for sluggish growth, the economic aftermath of the terrorist attacks in September and the subsequent precipitous decline in consumer and business confidence late in 2001 were sufficient to tip the Nation into its seventh recession since 1960.

Moderation After Very Rapid Growth

The strong growth recorded from 1995 through 1999 was a welcome and beneficial development, as the private sector reaped the rewards from its investments in high technology. In particular, the productivity gains offered by the more intensive use of computers, fiber optic technologies, and the Internet drove an investment boom in which the Nation's businesses retooled and upgraded their workplaces for the 21st century. Not surprisingly, the rapid pace of investment then slowed as the need to adopt the new technologies began to be satisfied and a more mature investment phase began. Although the transition to a more moderate growth rate could in principle

have been smooth, in practice additional economic developments created swings in investment spending that contributed to the significant slowing of economic activity.

Decline in Equity Values

The decline in equity values starting in early 2000 also helped slow economic activity by dampening both consumption and business fixed investment spending. Equity in businesses (both in corporations and in noncorporate businesses) fell from its peak of \$17.5 trillion in the first quarter of 2000 to just under \$13 trillion in the third quarter of 2001, according to the latest quarterly estimate from the Federal Reserve's flow of funds accounts. Various studies suggest that every one-dollar decline in stock market wealth ultimately reduces annual consumption spending by 3 to 4 cents. Thus the observed \$4.5 trillion decline in wealth could be expected to reduce consumption by \$135 billion to \$180 billion, or roughly 1 to 2 percentage points of GDP. Downward pressure from the equity decline may continue to affect consumption spending into 2002, because a drop in wealth typically has lagged effects for 1 to 2 years. Offsetting some of the decline in equity wealth, however, has been a continued increase in housing wealth. From the start of 2000 to the middle of 2001, housing prices rose at a steady 9 percent annual pace, increasing housing wealth by \$1.7 trillion.

The effect of the decline in equity prices on investment demand was both direct and indirect. Lower equity prices reduced investment spending directly by raising the cost of capital for corporations, and indirectly by causing growth in aggregate demand for final goods and services to wane.

Surge in Energy Prices

Energy prices surged in 1999 and 2000, reaching extremely high levels at the start of 2001. Oil prices rose dramatically from \$12.00 a barrel to peak in November 2000 at \$34.40 a barrel for West Texas Intermediate crude, its highest monthly average price since October 1990. Even more dramatic was the spike in natural gas prices, to the highest price on record, \$8.95 per million Btu in December 2000. This was more than 3½ times the average price over the preceding 6 years. These developments in energy prices had important ramifications for 2001. Personal disposable income available for goods and services other than energy fell as gasoline, heating, and electricity prices soared. Producers of nonenergy goods and services also suffered as their costs of production rose—especially in the energy-intensive manufacturing sector. The decline in demand and the rise in input costs squeezed profit margins, slowing corporate cash flow and reinforcing the downdraft on stock market values and capital spending plans.

Higher Interest Rates

Higher interest rates in 2000 and early 2001 also contributed to the deceleration in activity. The 10-year Treasury yield peaked at 6.7 percent in January 2000, and the 10-year corporate Baa yield hit 8.9 percent in May. Short-term interest rates rose consistently for a full year before reaching 6.2 percent in November 2000. The higher interest rate environment slowed economic activity as consumers were given the incentive to consume less, and investment in plant and equipment became less attractive.

Collapse of the High-Technology Sector

The collapse of stock prices in the high-technology sector—especially the dot-coms, or Internet-related firms—contributed an additional drag on economic activity. Prices for high-technology stocks as measured by the NASDAQ composite index fell 67 percent from their monthly peak in March 2000 to their monthly trough in October 2001, returning the NASDAQ to levels last seen in early 1998. By contrast, during the same period the Wilshire 5000 index fell by a much smaller 32 percent. The drop in the high-technology stocks represented an important reduction in equity wealth, but it also signaled a sea change in the fortunes of these businesses—especially those in the information and communications technology industries—which had been an important source of economic gains in the 1995-99 period. Investors both ratcheted down the earnings prospects of these firms and perceived a greater risk of investing in both established and more speculative high-technology businesses. This fundamental reevaluation of information and communications technology firms led to a swift downturn in the sector's activity and a reversal of the capital investment boom.

Lingering Effects of Y2K

The runup in capital spending by firms nationwide in anticipation of and in response to the Y2K event created conditions that exacerbated swings in high-technology capital spending. Instead of primarily upgrading existing capital and software, which might have remained vulnerable to the Y2K bug, most businesses replaced them with the latest technologies. The resulting bulge in investment spending around January 2000 generated a tendency toward a subsequent investment lull. Given that the typical replacement cycle for high-technology goods is about 3 to 5 years, it is not surprising that the investment decline that began in 2000 lingered in 2001.

Effects on Inventories and the Capital Stock

The factors just discussed—the transition to more moderate growth rates, the decline in equity values, the surge in energy prices, higher interest rates, the collapse of high-technology industries, and the lingering effects of Y2K—constituted a potent set of adverse economic circumstances for investment in 2000, with consequences for 2001. The declining stock market and higher interest rates increased the cost of external financing of new investment. At the same time, higher energy prices ate into corporate cash flow, which was already slowing as the economy decelerated. As a result, the financing gap (capital expenditure less internally generated funds) hit an all-time high in 2000. Also, by mid-2000 businesses found themselves with unplanned inventories as demand began to soften, and the result was a traditional inventory cycle. The accumulation of unwanted inventories led businesses to slow production further, with consequences for employment growth. This in turn fed the reduction in demand that had left businesses with rising inventories in the first place.

As the economy slowed, firms found themselves with the desire to defer future capital spending plans. By some estimates, a “capital overhang” developed in which the actual capital stock exceeded that desired by firms to meet the lower expected demand in 2000. By late 2001, however, the decline in investment spending had likely eliminated the capital overhang (Box 1-2).

Box 1-2. Capital Overhang and Investment in 2001

A capital overhang develops when the amount of capital in the economy exceeds the amount that businesses desire for the production of goods and services. The emergence of such an overhang complicates both business planning and policymaking. Businesses often have to alter their capital spending plans and curtail their investment spending—sometimes quite abruptly. A large overhang may also reduce the stimulative effects of tax policies designed to boost investment, possibly lengthening the recovery time during a period of sluggish economic activity, especially for the manufacturing sector.

An overhang can arise in various ways. If, for example, rapid growth is expected in the future, businesses will begin increasing their investment in advance. If the faster growth is not realized, these businesses will find themselves with too much capital. A capital overhang can also arise during a short period of unexpectedly sluggish growth. If the decline in demand is thought to be sufficiently deep and persistent, businesses may want to reduce their capital spending plans,

continued on next page...

Box 1-2.—*continued*

and possibly sell off part of their capital stock, especially those capital goods that are readily marketable. However, if the slowdown is sufficiently short, businesses may prefer to reduce their use of the capital stock rather than sell it, especially because the market price of capital goods is likely to fall during such periods. Selling capital and buying it back at a later date can then be more costly than simply holding onto it and not using it to its full capacity. Reducing the utilization rate thus helps to prevent the desired capital stock from falling.

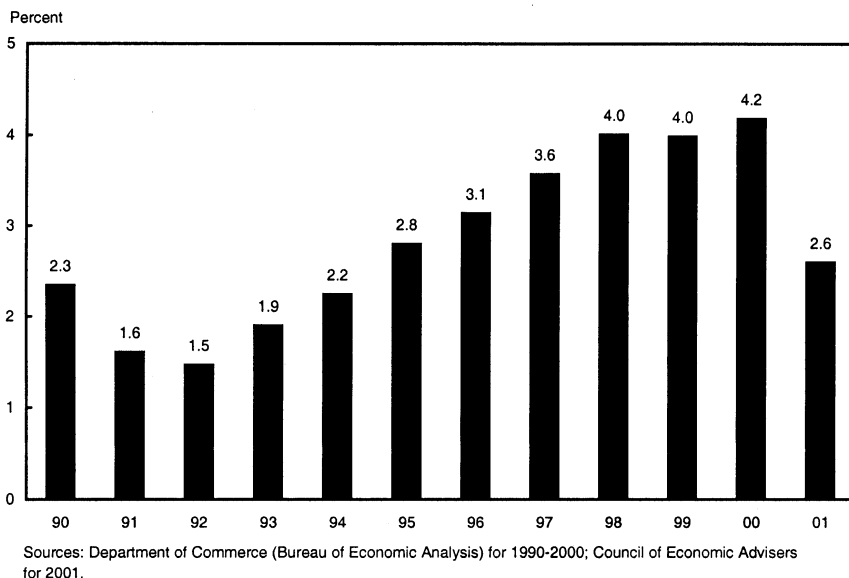
Policymakers have lately been concerned that the changing business climate may have given rise to a capital overhang over the past 2 years. Some businesses, especially in the information and communications technology sector, may have overestimated the potential of the “New Economy” and therefore overinvested in productive capacity. In addition, businesses throughout the economy were surprised by the extent of the slowdown in aggregate demand in 2000 and 2001, and they therefore had to revise downward the path of their desired capital stock.

Empirical evidence suggests that a capital overhang did develop in 2000. The overhang was modest for the economy on average, but various types of capital equipment such as servers, routers, switches, optical cabling, and large trucks were disproportionately affected. Estimates of the total overhang must be interpreted with caution. There is considerable uncertainty about its size, because it is difficult to estimate precisely both the capital stock that businesses desire and the capital stock they actually possess. Better data collection (see Box 1-1) could help solve this problem in the future. In any case, over the past year and a half, the decline in investment spending and depreciation of the existing capital stock combined to slow capital accumulation sufficiently to eliminate the overhang. Chart 1-5 shows that the capital stock, which had been growing at an annual rate above 4 percent over the past several years, is estimated to have grown just over 2 1/2 percent in 2001.

The remarkable slowdown in capital accumulation during 2001 underscores the importance of the President's tax relief recommendations for economic stimulus. The partial expensing provisions and the elimination of the corporate alternative minimum tax will encourage business investment, stimulating economic activity in the short run and laying the foundation for stronger growth in the long run. The reductions in marginal income tax rates will help spur investment by providing incentives for flow-through entities, mainly small businesses, to grow and create jobs. The President's tax relief will also help foster a smooth and more predictable transition to a period of sustainable growth.

Chart 1-5 Growth in the Real Capital Stock

Growth in the Nation's fixed nonresidential capital stock slowed considerably in 2001 as investment spending plummeted from its rapid pace during the preceding 5 years.



From Slowdown to Recession

Even though economic activity had begun to soften in the first half of 2000, the onset of recession did not arrive until March 2001, according to the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER), the arbiter of U.S. business cycle dates. The committee based this date on its reading of the economic data through November 2001, especially the four measures of economic activity it considers most important: industrial production, the real volume of sales in manufacturing and trade, employment, and real personal income less transfer payments. Industrial production peaked in June 2000, real sales in manufacturing and trade peaked in August 2000, employment peaked in March 2001, and real personal income less transfers may not have peaked yet.

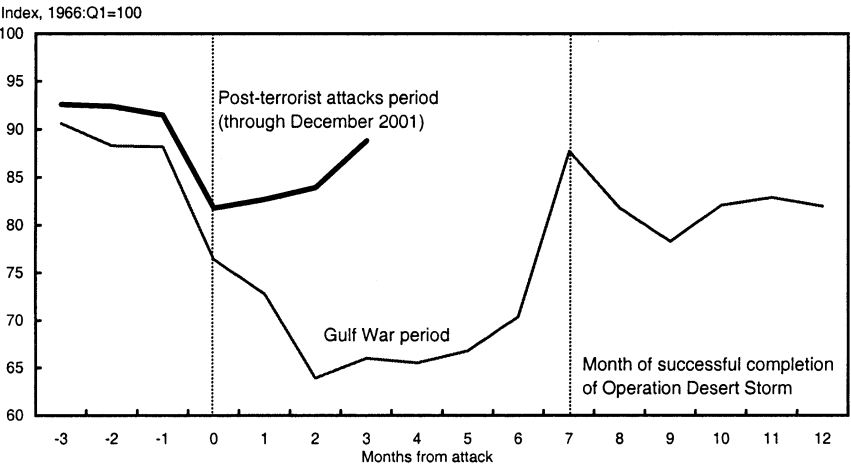
As the variation in these dates suggests, picking “the” month for the start of a recession involves considerable judgment and is not without controversy. The employment series appears to play a dominant role in the NBER committee’s decisions. Without a doubt, employment is a key resource for economic activity, representing about two-thirds of all inputs into production. In recessions since 1960, however, the peak in employment has tended to follow the peak in economy-wide activity. In addition, total industrial capacity utilization, a standard measure of the employment of capital—the other key input in production—peaked in mid-2000, suggesting an earlier

economy-wide turning point. These statistical arguments notwithstanding, the evidence is clear that the industrial sector was already well into a contraction, and real sales volumes were sagging, before March 2001. Finally, the economic consequences of the terrorist attacks were critical to the business cycle dating. As the committee noted in its decision, “before the attacks, it is possible that the decline in the economy would have been too mild to qualify as a recession. The attacks clearly deepened the contraction and may have been an important factor in turning the episode into a recession.”

The decline in consumer and business confidence following the terrorist attacks in September had a larger and more durable macroeconomic effect than the physical destruction and was sufficient to scuttle any possibility of avoiding a recession. Chart 1-6 shows, however, that the decline in the University of Michigan consumer sentiment index following September 11 was less than the sharp drop following Iraq’s invasion of Kuwait in 1990. Since September, consumer confidence has rebounded noticeably, to close to the preattack level. By comparison, during the Gulf War period, consumer confidence remained subdued for a longer period but then surged when the successful completion of Operation Desert Storm largely resolved uncertainty about the future.

Overall, the deceleration of economic activity since mid-2000 has been dramatic. Unemployment has risen, business earnings have suffered, and government budgets have been strained. As in past recessions, no single key

Chart 1-6 Consumer Sentiment
Consumer sentiment fell sharply after September 11. But sentiment rebounded more quickly than after the Iraqi invasion of Kuwait.



Note: The month of the attack initiating the Gulf War was August 1990, when Iraq invaded Kuwait.

Source: Surveys of Consumers (University of Michigan).

factor caused the slowdown and subsequent recession; rather it took the confluence of a series of unforeseen adverse events. Despite some similarities shared with previous episodes of sluggish growth, the 2000-01 slowdown has been unique in many respects and has required policies to address the particular challenges of these developments.

Policy Developments in 2001

Both fiscal and monetary policy became expansionary in 2001. The Federal budget surplus, although still substantial by historical standards, fell because of deteriorating economic conditions and changing fiscal priorities after the terrorist attacks. Falling short-term interest rates and rapid expansion of the money supply indicated that monetary policy was eased significantly during the year.

Fiscal Policy Before the Terrorist Attacks

In February 2001 the President's budget for fiscal 2002 outlined major policy initiatives for the Nation. These included continuing the retirement of the Federal debt, providing tax relief for American families, strengthening and reforming education, modernizing and reforming Social Security, modernizing and reforming Medicare, revitalizing national defense, and championing faith-based initiatives. Although tangible progress has already been made, fiscal vigilance will be essential to continuing toward these goals. The Federal budget process needs to be more disciplined, and spending limits previously agreed upon should be respected. Too often in the past, budget deadlines were missed and legislation was consolidated into omnibus spending bills that exceeded the agreed spending limits. Appropriations in fiscal 2001, even before the emergency funds made available after September 11, were over \$50 billion higher than in 2000—the largest 1-year appropriations increase in history. The events in September and October precluded an expeditious completion of the appropriations process in the fall, but the President and Congress agreed to limit discretionary spending to \$686 billion excluding emergency spending. This new level provides reasonable spending growth, ensures funding for Medicare and Social Security, and sets an example for future budget negotiations.

In fiscal 2001 the Federal Government ran the second-largest budget surplus in history and paid down the second-largest amount of debt in history, despite the weak economic conditions. Looking forward, the Federal budget will be in deficit during fiscal 2002 but, with spending restraint and pro-growth policies, is projected to return to surplus beginning in 2005. About two-thirds of the decline in the projected baseline fiscal position since

last year may be traced to the weaker economy and technical revisions. Spending accounts for nearly 20 percent of the decline, and the EGTRRA provisions account for under 15 percent.

A sound long-run fiscal position holds down unnecessary spending and removes tax-based impediments to economic growth. As noted earlier, the tax cut in 2001 was key to mitigating the severity of the slowdown and simultaneously improving growth incentives. The deterioration in the surplus from a weak economy is the mirror image of the experience of the late 1990s, when budget surpluses were fueled largely by a strong economy. In general, faster economic growth causes budget surpluses, not the other way around. Moreover, policies that promote job creation and entrepreneurial activity ultimately increase the size of the economy and hence provide the resources for future spending obligations.

Tax Relief in 2001

The President laid a strong foundation for growth in 2001 with the Economic Growth and Tax Relief Reconciliation Act. This package provides a powerful stimulus for future growth, with reductions in marginal tax rates that improve incentives and leave in the hands of Americans a greater share of their own money to spend on consumption, education, and retirement investment.

The first reduction in marginal tax rates was effective for 2001 and was reflected in lower withholding during the second half of the year. In addition, the new 10 percent tax rate bracket, carved out of the beginning of the 15 percent rate bracket, was reflected in rebate checks totaling \$36 billion, which were mailed to 85 million taxpayers during the second half. The timing of these reductions in withholding and rebates proved propitious: they added significant economic stimulus by boosting purchasing power in the hands of consumers during a period of sluggish economic activity. The 2001 tax rate reductions were just the first step in a series of income tax rate reductions to be phased in by 2006; by that year the 39.6 percent tax rate will have dropped to 35 percent, the 36 percent rate to 33 percent, the 31 percent rate to 28 percent, and the 28 percent rate to 25 percent.

The tax cut package also provided incentives for saving, investment, and capital accumulation. Higher IRA and 401(k) retirement contribution limits are to be phased in over time, with those for persons over 50 phased in more quickly. Beginning in 2002 and continuing through 2009, the highest estate tax rates are reduced and the effective exemption amount is increased, reducing an important impediment to the growth of entrepreneurial enterprises and the overall accumulation of wealth. In 2010 the estate tax is eliminated. Small businesses will benefit from the lowering of individual income tax rates for owners of flow-through business entities such as sole

proprietorships and partnerships. In 1998 there were close to 24 million flow-through businesses in the United States, including 17.1 million sole proprietorships, 2.1 million farm proprietorships, 1.9 million partnerships, and 2.6 million S corporations. By 2006, when the personal income tax cut is fully phased in, the Treasury Department estimates that over 20 million tax filers with income from flow-through businesses will receive a tax reduction.

Finally, the President's tax cut strengthens families and reduces the burden of financing education. The marriage penalty is reduced, and the annual child tax credit is increased from \$500 to \$600 per child in 2001 and gradually increased to \$1,000 by 2010. Adoption credits are doubled in 2002 from \$5,000 per child; in addition, the credit will apply to more taxpayers, because the income threshold at which the credit begins to phase out will rise to \$150,000 from \$75,000. Contribution limits for education savings accounts (formerly called educational IRAs) are raised to \$2,000 a year, and distributions are made tax-exempt. The law also increased the income phaseout range for student loan interest deductions and made certain higher education costs tax-exempt for households with less than \$130,000 in income.

The initial macroeconomic effects of tax relief have been positive, strengthening aggregate demand in the face of other downward pressures. The rebate checks and the lower marginal tax rates alone reduced taxpayer liabilities by \$44 billion in 2001 and by \$52 billion in 2002. Adding in the effects of the other provisions of EGTRRA (such as the education incentives, child credits, the individual alternative minimum tax, and marriage penalty relief) brings the liability reduction in 2001 and 2002 to \$57 billion and \$69 billion, respectively.

In short, the President delivered important tax relief in 2001, providing a solid foundation for renewed growth in consumer spending once confidence rebounds, and for an improved investment climate for businesses. The boost in aggregate demand should help provide a foundation for economy-wide recovery in 2002.

Monetary Policy Before the Terrorist Attacks

The Federal Reserve aggressively pursued an easier monetary policy during 2001. With clear evidence that economic activity was sharply decelerating at the end of 2000 and that inflation pressures were minimal, the Federal Open Market Committee (FOMC) began cutting the target Federal funds rate by 50 basis points (hundredths of a percentage point) at an unscheduled meeting on January 3, 2001. By mid-August the FOMC had lowered its target Federal funds rate on seven occasions, from 6½ percent at the start of the year to 3½ percent (the lowest rate since early 1994). The target rate reductions were also notable for their rapid succession. The Federal Reserve

lowered the target rate at every scheduled meeting and at two unscheduled meetings—a sequence of events rare in its history, and one that underscored the seriousness of the deterioration in economic conditions. At each meeting the committee also reaffirmed its view that the risks of weaker economic activity outweighed the risks of higher inflation. Over the first 8 months of 2001, easier monetary policy pushed growth in M2 (a broad definition of the money supply) to an annualized 10 percent rate.

Market interest rates responded to the lower targets for the Federal funds rate. Short-term interest rates followed in lockstep, with the 3-month Treasury bill rate declining roughly 240 basis points from December 2000 to early September 2001. Three-month commercial paper rates, credit card rates, personal loan rates, and 1-year adjustable mortgage rates also moved down. Long-term rates decreased as well, but by a smaller amount. Ten-year Treasury yields slid almost 20 basis points, and rates on 30-year fixed rate mortgages fell about 25 basis points. Corporate bond yields also receded: yields on corporate Baa-rated bonds fell roughly 15 basis points. The Merrill Lynch high-yield bond index was off about 20 basis points.

The pattern of short-term and long-term interest rates during 2001 is consistent with similar periods in the past. History shows that when the economy has slowed sharply or is in a recession, and monetary policy has eased significantly, short-term interest rates have tended to fall more than long-term rates, but the large decline in short-term rates often proves temporary. In addition, the widening interest rate spread during 2001 reflected the fact that long-term rates had edged down in 2000 in anticipation of lower short-term rates in 2001. On the whole, the pattern of the yield spread is more a reflection of the circumstances of the recession, not a factor contributing to it.

The Macroeconomic Policy Response After September 11

In the days and weeks following the September terrorist attacks, fiscal and monetary actions were taken to address the new challenges. The President expeditiously requested emergency funds to assist in meeting humanitarian, recovery, and national security needs. The Federal Reserve added substantial liquidity through various channels to help markets function in an orderly fashion in the immediate aftermath of the attacks, and it continued to ease monetary policy.

Fiscal Policy

In the wake of the attacks, the President took action to ensure the security of Americans. The President signed the 2001 Emergency Supplemental Appropriations Act for Recovery from and Response to Terrorist Attacks on the United States. The \$40 billion in funding assisted victims and addressed

other consequences of the attacks. Funding was provided for debris removal, search and rescue efforts, and victim assistance efforts of the Federal Emergency Management Agency; emergency grants to health providers in the disaster-affected metropolitan areas; investigative expenses of the Federal Bureau of Investigation; increased airport security and sky marshals; initial repair of the Pentagon; evacuation of high-threat embassies abroad; additional expenditures of the Small Business Administration disaster loan program; and initial crisis and recovery operations of the Department of Defense and other national security operations. These measures took needed initial steps toward restoring security and confidence in the economy. The President also proposed additional funding to help displaced workers and to extend unemployment insurance in impacted areas.

In September the President signed the Air Transportation Safety and System Stabilization Act, which provided the tools necessary to aid the transition of the air transport system to the new security and economic environment. The law provides \$5 billion to compensate for losses to the industry directly resulting from the attacks; it also allows the President to issue up to \$10 billion in Federal loan guarantees.

The terrorist attacks introduced new risks into the economic environment. One of the challenges has been to provide an umbrella of support for economic security that draws on the strengths of the private sector. The Administration has proposed measures designed to provide economic growth insurance, or economic stimulus. The central focus of this effort is to address the immediate needs of those displaced workers directly affected by the recession and the terrorist attacks, while also mitigating the effects of these events on the broader economy. In response to the President's leadership, the House of Representatives passed such stimulus legislation on two separate occasions, but the Senate failed to pass such legislation.

In choosing among alternative economic stimulus policies, the government should favor those that are pro-growth—enhancing long-term incentives to work, invest, take risks, and expand productive capacity—as well as remain cognizant of short-term needs. The Administration's approach includes tax relief for low-income families and extended unemployment insurance benefits. These types of policies address short-term needs while also providing purchasing power that helps to ensure steady demand for businesses.

However, the real solution to the economic woes of displaced workers is employment. Fully addressing these workers' needs and buttressing confidence on the part of all households and businesses requires a focus on job growth. One key to this effort is small businesses and entrepreneurs, traditionally an important source of new jobs in the economy. The best policy to help businesses and entrepreneurs is to reduce their marginal tax rates. The

Administration proposes moving forward the implementation of the marginal tax rate cuts passed by Congress in the spring of 2001. Lower marginal tax rates both improve incentives and augment the cash flow of small businesses. Research shows that entrepreneurs will respond to these stronger incentives and increased cash flow by expanding their payrolls and increasing their investments.

A second policy to provide incentives for private sector job creation is to help businesses overcome uncertainty and restart investment spending. At the aggregate level, the return to rapid growth requires a resumption in the growth of capital expenditure. Employment losses have been concentrated in the manufacturing sector—a sector heavily dependent on the health of business investment. For this reason the Administration has focused on growth incentives, such as partial expensing and reform of the corporate alternative minimum tax, that target the source of the problem, namely, an investment slump that has diminished private sector job creation.

Property and casualty insurance is one mechanism by which economies respond efficiently to risks in the business environment. Insurance spreads these risks, converting, for each business that takes out insurance, a potential cost of unknowable size and timing into a set of smaller premium payments of known magnitude. The events of September 11 induced a dramatic revision in businesses' perceptions of the risks facing them. In normal circumstances, such increased risks are translated into higher premiums. This serves the useful economic function of pricing risk, leading the private sector toward those activities that present a risk worth taking, and away from foolhardy gambles.

In the aftermath of September 11, however, one concern was that the economy faced disproportionate increases in terrorism risk insurance premiums or, in the extreme, a complete withdrawal of this type of coverage. With this concern in mind, the Administration proposed legislation to provide a short-term backstop for terrorism risk insurance that would encourage rather than discourage private market incentives to expand the economy's capacity to absorb and diversify risk, and which would expire as soon as the private market is capable of insuring these losses on its own.

Taken as a whole, the President's policies have improved the Nation's security, compensated the direct victims of the September attacks, and aided displaced workers. If the President's terrorism risk insurance and economic stimulus proposals are passed, they will further enhance economic security.

Monetary Policy

In the hours, days, and weeks following the terrorist attacks, the Federal Reserve used its financial resources to provide liquidity and ensure the functioning of financial markets. The Nation's central bank injected substantial liquidity into financial markets by promoting the use of the discount

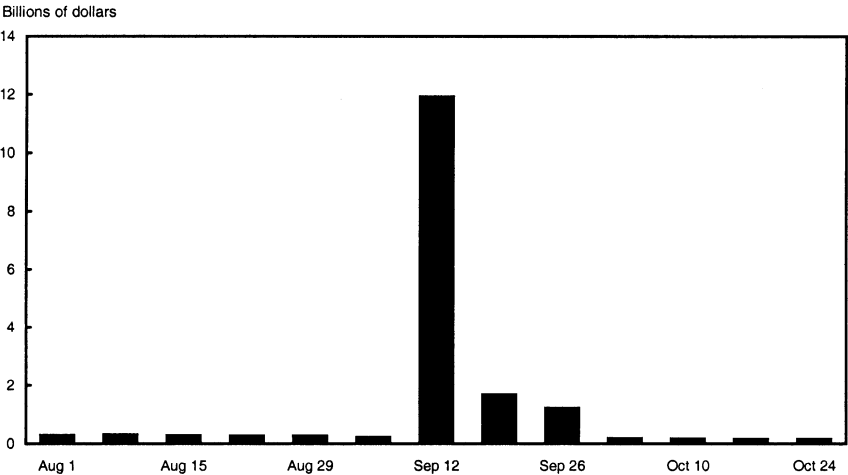
window by depository institutions, increasing the volume of open market operations, and arranging temporary reciprocal currency swaps (swap lines) with several foreign central banks.

On September 11, the Federal Reserve made it clear through a press release that the discount window was available to meet liquidity needs, and depository institutions responded by employing the discount window at an unprecedented level. Before September 11 average weekly discount borrowing during 2001 had been \$143 million. During the week of the attack, however, borrowing ballooned to an all-time high of \$11.8 billion (Chart 1-7). In the next 2 weeks, as liquidity pressures waned, borrowing quickly dropped to the \$1 billion to \$1.5 billion range and then returned to levels seen earlier in the year. On the days that followed the attack, the Federal Reserve also allowed reserves in the Federal funds market to rise as Federal Reserve float surged because of the closure of the Nation's air transportation system. In addition, the Federal Reserve made liquidity available by arranging temporary swap lines with the European Central Bank (ECB) and the Bank of England, and by augmenting existing swap lines with the Bank of Canada.

In the week following the attacks, the Federal Reserve eased monetary policy further at an unscheduled meeting of the FOMC, lowering its target Federal funds rate ½ percentage point, to 3 percent. The FOMC reiterated, in a press release accompanying its decision, that it would continue to supply large amounts of liquidity to counter the extraordinary strains in the

Chart 1-7 Discount Window Borrowing

The banking system's liquidity needs in the immediate aftermath of the September 2001 terrorist attacks were addressed in part through unprecedented levels of discount window borrowing.



Note: Data are averages of daily figures for the week ending on Wednesday.

Source: Board of Governors of the Federal Reserve System.

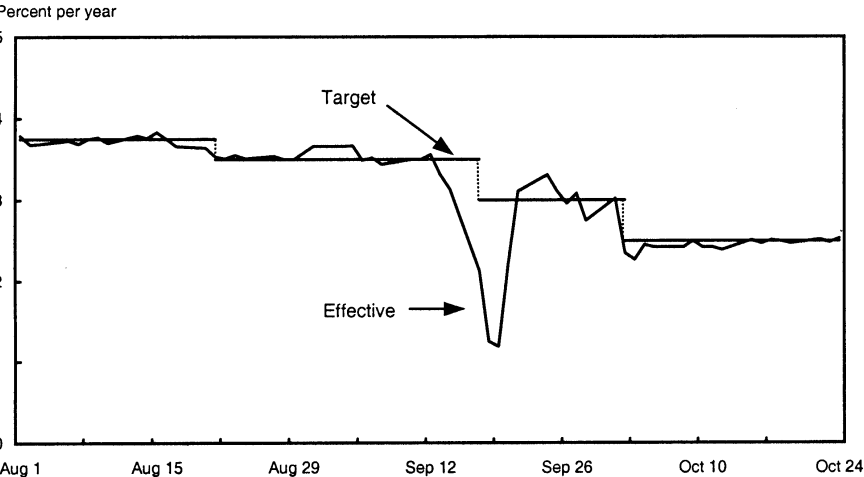
financial markets as well as to help ensure the effective functioning of the banking system. The committee recognized that providing ample liquidity in the short run could lead to the Federal funds rate trading well below its target. In fact, in the week following September 11, the effective Federal funds rate fell to an average of 1.2 percent for the 2 days of the week when liquidity issues were of primary concern (Chart 1-8).

Despite the devastation to New York’s financial center, financial markets and the banking system resumed business quickly and were operating at near-normal conditions within just weeks of the terrorist attacks. The remarkable resiliency of the financial markets and the longstanding policy of the Federal Reserve to provide ample liquidity to stabilize markets in the wake of unusual developments combined to mute the effects of the initial shock.

Since mid-September the FOMC has continued its easing of monetary policy to help counter the deterioration of economic activity. By the end of the year the Federal Reserve had lowered its Federal funds target to 1¾ percent, its lowest level in 40 years, leaving the real Federal funds rate near zero. Meanwhile there was no evidence of increasing inflation pressures. The lowering of the Federal funds rate target led to further declines in short-term and long-term market interest rates. At the end of the year, short-term market interest rates were below 2 percent. The 10-year Treasury yield was 5.2 percent, and 30-year conventional mortgage rates averaged 7.2 percent.

Chart 1-8 Effective and Target Federal Funds Rates

The Federal Reserve added liquidity into markets immediately after the September 2001 terrorist attacks, causing the effective Federal funds rate to plunge. In addition, the target Federal funds rate was lowered.



Note: The daily effective rate is an average of the rates on a given day weighted by the volume of transactions at these rates.

Source: Board of Governors of the Federal Reserve System.

Economic Developments Outside the United States

Growth in the rest of the world slowed markedly in 2001. The global slowdown is attributable to many of the same factors that affected the United States: weakened investment demand (especially for high-technology goods), relatively high oil prices in 2000 and early 2001, and the increased costs and loss of confidence associated with the September terrorist attacks.

Canada and Mexico, our largest trading partners, saw their economies soften in 2001. Canadian economic growth began to fall in 2000 as the deterioration in U.S. economic conditions particularly affected Canadian exports. Late in 2001 Canada's exports and domestic demand were weakened further by disruptions and increased uncertainty following the terrorist attacks. Real GDP growth was 1.4 percent for 2001 as a whole, down from 4.4 percent in 2000, and the unemployment rate stood at 8 percent at year's end. Mexico experienced zero growth in 2001, following a long period of expansion; real GDP growth had been 6.9 percent in 2000. The unemployment rate edged up to 2.5 percent for 2001.

Growth also faltered in Europe. In the euro area (the 12 European countries that have adopted the euro as their common currency), output growth slowed significantly in 2001, after weak growth in the second half of 2000. The unemployment rate remained above 8 percent last year. Because of constraints imposed by member countries' commitments to the monetary union, fiscal policy in the euro area remained only slightly stimulative. With regard to monetary policy, the European Central Bank cut interest rates by a total of 150 basis points in 2001. Growth in the United Kingdom declined in 2001, but by less than in continental Europe, bolstered in part by a 200-basis-point reduction in short-term interest rates. Over the year, growth fell to 2.3 percent from 2.9 percent in 2000. The unemployment rate declined to 5.1 percent in 2001, its lowest in 26 years.

Japan fell into its third recession in 8 years during 2001, with its unemployment rate reaching an all-time high of 5.5 percent as of November. Although Japan, too, suffered from the effects of the slowing global economy, it also continued to struggle with its moribund banking and corporate sectors. Fiscal stimulus and monetary easing have done little thus far to improve the country's economic prospects.

The newly industrialized economies in East Asia were particularly hard hit by economic stagnation in Japan and the slump in global technology investment. High-technology goods account for roughly 40 percent of these economies' exports. After increasing 8.2 percent in 2000, output in these economies registered only a 0.4 percent increase in 2001.

In the developing economies as a group, economic growth moderated from almost 6 percent in 2000 to 4 percent in 2001. Meanwhile growth for the developing economies in Asia declined from almost 7 percent to just over 5½ percent. In China, fiscal measures aimed at infrastructure investment helped maintain rapid growth: Chinese GDP growth for 2001 was roughly 7 percent. The Middle East and developing countries in the Western Hemisphere saw GDP growth fall dramatically, to just 1 to 2 percent in 2001. In contrast, Africa saw growth edge up from just under 3 percent to 3½ percent.

Two of the world's larger developing economies—Turkey and Argentina—faced significant financial turmoil in 2001. In Turkey, a banking crisis and political uncertainty led to high real interest rates and a sharp drop in output. The Turkish lira was floated in February 2001 and depreciated sharply against the dollar before stabilizing. Late in the year Argentina also experienced severe financial distress, with unsustainable fiscal policy leading to loss of confidence and a run on bank deposits, culminating in a default on the country's sovereign debt and dramatic political unrest.

The Economic Outlook

The Administration expects that the economy will recover in 2002. The economy continues to display characteristics favorable to long-term growth: productivity growth remains strong, and inflation remains low and stable.

Near-Term Outlook: Poised for Recovery

Real GDP growth is expected to pick up early in 2002 (Table 1-1). The pace is expected to be slow initially, followed by an acceleration thereafter; over the four quarters of 2002 real GDP is expected to grow 2.7 percent. The unemployment rate is projected to continue rising through the middle of 2002, when it is expected to peak around 6 percent.

As discussed earlier, the decline in aggregate demand during the past year was concentrated in inventory investment, business fixed investment, and exports. Of these downward pressures, that from inventory disinvestment is projected to reverse its course soonest and most rapidly, as the pace of liquidation is forecast to recede dramatically in the first quarter of 2002. By the end of 2001 inventories had become quite lean, making it likely that, once sales resume their growth, stockbuilding will boost real GDP growth.

Growth in business investment and exports may take longer to reassert itself. Nonresidential investment fell sharply in 2001, and some downward momentum probably remained at the start of 2002. Still, the financial foundations for investment remain positive: real short-term interest rates are low,

TABLE 1-1.—*Administration Forecast*¹

Year	Nominal GDP	Real GDP (chain-type)	GDP price index (chain-type)	Consumer price index (CPI-U)	Unemployment rate (percent)	Interest rate, 91-day Treasury bills (percent)	Interest rate, 10-year Treasury notes (percent)	Nonfarm payroll employment (millions)
	Percent change, fourth quarter to fourth quarter				Level, calendar year			
2000 (actual)	5.3	2.8	2.4	3.4	4.0	5.8	6.0	131.8
2001	1.9	-5	2.4	2.0	4.8	3.4	5.0	132.3
2002	4.7	2.7	1.9	2.4	5.9	2.2	5.1	132.2
2003	5.6	3.8	1.7	2.2	5.5	3.5	5.1	135.2
2004	5.5	3.7	1.7	2.3	5.2	4.0	5.1	138.3
2005	5.4	3.5	1.9	2.4	5.0	4.3	5.1	140.9
2006	5.0	3.1	1.9	2.4	4.9	4.3	5.2	143.2
2007	5.0	3.1	1.9	2.4	4.9	4.3	5.2	145.4
2008	5.0	3.1	1.9	2.4	4.9	4.3	5.2	147.5
2009	5.0	3.1	1.9	2.3	4.9	4.3	5.2	149.6
2010	5.0	3.1	1.9	2.3	4.9	4.3	5.3	151.7
2011	5.0	3.1	1.9	2.3	4.9	4.3	5.3	153.9
2012	5.0	3.1	1.9	2.3	4.9	4.3	5.3	156.1

¹ Based on data available as of November 30, 2001.

Sources: Council of Economic Advisers, Department of Commerce (Bureau of Economic Analysis), Department of Labor (Bureau of Labor Statistics), Department of the Treasury, and Office of Management and Budget.

prices of computers are again falling rapidly, and equity prices moved up during the fourth quarter. Indications late in the year suggested that these factors were contributing to an upturn in new orders for nondefense capital goods in October and November. The Administration projects that business fixed investment will return to positive growth around the middle of 2002 and resume rapid growth thereafter.

The past year's decline in exports reflects stagnating growth among the United States' trading partners. Consensus estimates of foreign growth in 2002 are anemic as well. In these circumstances any rebound in exports is likely to lag behind the expected recovery of U.S. GDP as a whole. Imports meanwhile are projected to grow faster than GDP. As a result, net exports and the current account deficit are likely to become increasingly negative during 2002.

Consumption growth slowed during the past year but has remained in positive territory. This slowing may be attributable to the decline in the stock market from its peak in March 2000. But in the absence of further stock market declines, such restraint is expected to wane. Consumption will also be supported by fiscal stimulus and interest rate cuts. The major provisions of EGTRRA will lower tax liabilities by about \$69 billion in 2002 (up from its contribution of \$57 billion in 2001).

Inflation Forecast

As measured by the GDP price index, inflation was stable at about 2.3 percent during the four quarters ending in the third quarter of 2001. The Administration expects this measure of inflation to fall to 1.9 percent over the four quarters of 2002. The unemployment rate is now above the level that the Administration considers to be the center of the range consistent with stable inflation, and capacity utilization in the industrial sector is substantially below its historical average. Despite faster-than-trend growth of output in 2003 and 2004, some downward pressure will be maintained on the inflation rate, because the unemployment rate is projected to remain high over that period. As a result, inflation in terms of the GDP price index is expected to inch down to 1.7 percent in 2003 before edging up to 1.9 percent over the forecast period.

In contrast, consumer price inflation is likely to edge up temporarily over the four quarters of 2002, to 2.4 percent, reflecting energy price fluctuations. (Petroleum-related goods make up a larger share of consumer budgets, on which the CPI is based, than of the production of final goods in the economy, on which the GDP price index is based.) In 2001 CPI inflation was held down by a 13 percent decline in energy prices. In 2002 petroleum prices are expected to stabilize, and energy price inflation is projected to be positive, but still moderate. Following a temporary increase in 2002, overall CPI inflation is projected to edge down and eventually flatten out at about 2.3 percent from 2003 forward.

Long-Term Outlook: Strengthening the Foundation for the Future

The Administration forecasts real GDP growth to average 3.1 percent a year during the 11 years through 2012. The growth rate of the economy over the long run is determined primarily by the growth rates of its supply-side components, which include population, labor force participation, productivity, and the workweek. The forecast is shown in Table 1-2.

The Administration expects nonfarm labor productivity to grow at a 2.1 percent average pace over the forecast period, the same as over the entire period since the previous business cycle peak in the third quarter of 1990. This forecast is noticeably more conservative than the 2.6 percent average annual growth rate of actual productivity from 1995 to 2001. The pace is projected to be slower as a caution against several downside risks:

- Nonresidential fixed investment has fallen about 6 percent from its peak in the fourth quarter of 2000, while the level of the capital stock—and therefore depreciation—remain elevated. This combination implies

that the near-term growth of capital services is likely to be reduced from its average pace from 1995 to 2001, leading to slower growth in labor productivity from the use of these capital services.

- The diversion of capital and labor toward increased security (which is largely an intermediate product) may reduce the growth of productivity modestly over the next few years (Box 1-3). Once the transition phase has been completed, the enduring restraint on productivity growth is likely to be small.
- As discussed in Box 1-4, about one-half of the post-1995 structural productivity acceleration is attributable to growth in total factor productivity (TFP) outside of the computer sector, perhaps due to technological progress and better business organization. (The latter aspect is discussed in Chapter 3.) Although there is no reason to expect this process not to continue, the Administration forecast adopts a cautious view in which the pace of TFP growth is near its longer term average.

TABLE 1-2.—*Accounting for Growth in Real GDP, 1960-2012*
[Average annual percent change]

Item	1960 Q2 to 1973 Q4	1973 Q4 to 1990 Q3	1990 Q3 to 2001 Q3	2001 Q3 to 2012 Q4
1) Civilian noninstitutional population aged 16 or over	1.8	1.5	1.0	1.0
2) Plus: Civilian labor force participation rate	.2	.5	.0	.0
3) Equals: Civilian labor force ¹	2.0	2.0	1.0	1.0
4) Plus: Civilian employment rate ¹	.0	-.1	.1	.0
5) Equals: Civilian employment ¹	2.0	1.9	1.1	1.0
6) Plus: Nonfarm business employment as a share of civilian employment ^{1 2}	.1	.1	.3	.3
7) Equals: Nonfarm business employment	2.1	2.0	1.5	1.3
8) Plus: Average weekly hours (nonfarm business)	-.5	-.4	-.1	.0
9) Equals: Hours of all persons (nonfarm business)	1.7	1.7	1.4	1.3
10) Plus: Output per hour (productivity, nonfarm business)	2.9	1.4	2.1	2.1
11) Equals: Nonfarm business output	4.6	3.1	3.4	3.5
12) Plus: Ratio of real GDP to nonfarm business output ³	-.3	-.2	-.4	-.4
13) Equals: Real GDP	4.2	2.9	3.0	3.1

¹ Adjusted for 1994 revision of the Current Population Survey.

² Line 6 translates the civilian employment growth rate into the nonfarm business employment growth rate.

³ Line 12 translates nonfarm business output back into output for all sectors (GDP), which includes the output of farms and general government.

Note.—The periods 1960 Q2, 1973 Q4, and 1990 Q3 are business cycle peaks.
Detail may not add to totals because of rounding.

Sources: Council of Economic Advisers, Department of Commerce (Bureau of Economic Analysis), and Department of Labor (Bureau of Labor Statistics).

Box 1-3. Increased Security Spending and Productivity Growth

The Nation will spend more on security in the wake of the terrorist attacks. Economic growth will likely slow because more labor and capital will be diverted toward the production of an intermediate product—security—and away from the production of final demand. In addition, lower output from these direct effects will lower national saving and investment, and this reduces output a bit further. The eventual increase in the private security budget is unknown, but for calibration purposes it is assumed that it doubles. Smaller or larger changes would produce proportionally smaller or larger effects. Under these assumptions, increased security costs reduce the level of output and productivity by about 0.6 percent after 5 years below what they would have been otherwise.

The United States spends roughly \$110 billion a year on security. This includes the services of Federal, State, and local police (but not the armed forces). Of this, private business spends about \$55 billion, or 0.53 percent of GDP. It is assumed that one-third of the incremental spending goes to security capital and two-thirds to security labor.

The diversion of two-thirds of \$55 billion for additional security labor diverts about 760,000 workers from productive employment, lowering labor input to the economy by 0.69 percent. This diversion lowers production by about two-thirds of 0.69, or about 0.46 percent. The diversion of one-third of \$55 billion from productive investment in the first year lowers the “productive” capital stock by 0.10 percent and lowers production by one-third of that, or about 0.03 percent.

In addition, by reducing output, the diversion also reduces saving and investment, in turn reducing output further. The diversion in each subsequent year lowers capital services even more. Assuming a 25 percent depreciation rate, capital services will have fallen by 0.39 percent after 5 years, lowering output by 0.13 percent.

The effect of the labor diversion is relatively large and immediate. The effect of the capital diversion, in contrast, takes a few years to accumulate. By the fifth year, output will be about 0.6 percent lower, with 85 percent of that effect arising in the first year or two. Thus productivity growth will be lower by 1/4 percentage point during the first 2 years but will be affected only marginally thereafter.

The other components of potential GDP growth shown in Table 1-2 are more easily projected. In line with the latest projection from the Bureau of the Census, the working-age population is projected to grow at an average 1.0 percent annual rate through 2012. The labor force participation rate and the work week are projected to remain approximately flat. In sum, potential real GDP growth is projected to grow at about a 3.1 percent annual pace, slightly above the average pace since 1973.

The rate on 91-day Treasury bills fell about 4 percentage points during the 12 months of 2001, reflecting the series of cuts in the Federal Reserve's interest rate target in response to the slowing economy. By the end of December, the Treasury bill rate had fallen to about 1.7 percent. At this nominal rate, real short-term rates (that is, nominal rates less expected inflation) are close to zero. Real rates this low are not expected to persist once recovery becomes firmly established, and nominal rates are projected to increase gradually to 4.3 percent by 2005. At that level the real rate on Treasury bills will be close to its historical average.

The Administration projects that the yield on 10-year Treasury notes will remain flat at 5.1 percent. The Administration's expectation for the 10-year rate reflects the assumption that the market yield embodies all pertinent information about the path of future interest rates. In 2003 and thereafter, the real 10-year rate is projected to remain slightly below its historical average. The projected term premium (the premium of the 10-year rate over the 91-day rate) of about 1 percentage point is projected to remain slightly (about 30 basis points) below its historical average.

One important purpose of the Administration forecast is to estimate future government revenue. To this end, the forecast of the components of taxable income is crucial. The Administration's income-side projection is based on the historical stability of the long-run labor and capital shares of gross domestic income (GDI). During the first three quarters of 2001, the labor share of GDI was on the high side of its historical average of 57.7 percent. It is projected to decline to this long-run average and then remain at this level over the forecast period. Nevertheless, the Administration forecasts that wages and salaries as a share of GDI will decline and that other labor income, especially employer-provided medical insurance, will grow faster than wages. The capital share of GDI is expected to rebound in the short run, reflecting an expected cyclical rebound in productivity, and to remain flat at roughly its historical average thereafter. Within the capital share, a near-term decline in the depreciation share (a consequence of the recent decline in equipment investment) implies an increase in the profit share from its current level. (Profits before taxes had fallen to 6.7 percent of GDP by the third quarter of 2001, well below the post-1969 average of 8.1 percent.) The Administration projects an increase in the profit share over the next several years, so that it averages 8.1 percent over the forecast period.

Box 1-4. Is There Still a New Economy?

The late 1990s witnessed what many regard as the birth of a “New Economy”—one characterized by the dominance of high-technology industries, immunity from cyclical downturns, and, most of all, rapid productivity growth. In the past year, however, high-technology stocks, especially Internet and communications stocks, led the stock market’s retreat; the 1990s expansion ended; and July’s annual revision to the national income and product accounts caused productivity to be revised downward. It is useful, therefore, to examine the evidence for a resumption of the post-1995 acceleration in productivity.

Productivity growth is cyclical: it typically slows relative to its trend immediately before and after a business cycle peak. Yet over the four quarters ending in the third quarter of 2001, productivity growth grew faster than in any comparable period during the last four decades (Chart 1-9).

Table 1-3 presents the results of an analysis of the factors that influence productivity growth and compares their influences in two periods: 1973 to 1995, and 1995 to 2001. According to a model designed to capture its cyclical behavior, the productivity acceleration after 1995 would have been stronger by 0.48 percentage point a year but for the hiring that took place during this period to accommodate the increase in demand that occurred before and during 1995. (See the second line in Table 1-3.) This model estimates that business cycle effects raised productivity growth noticeably in 1992-94 as the economy emerged from recession, and reduced it noticeably in 1999, 2000, and 2001 (by 0.8, 0.4, and 1.4 percentage points, respectively). Adjusted for this cyclical effect, structural productivity has accelerated by 1.70 percentage points. In short, the latest evidence shows structural productivity growth continuing to exceed its pace during the period from 1973 to 1995. Because it was reduced by the effects of the business cycle slowdown, actual productivity growth accelerated somewhat less than structural productivity: by 1.21 percentage points, to a 2.60 percent annual rate of growth.

In general, an acceleration in structural productivity can come from increases in any of the following four sources of growth:

- growth in the amount of capital services per worker-hour throughout the economy (capital deepening),
- improvements in the measurable skills of the work force (labor quality),
- total factor productivity (TFP) growth in computer-producing industries, and
- TFP in other industries.

continued on next page...

Box 1-4.—*continued*

TFP growth is the increase in aggregate output over and above that due to increases in capital or labor inputs. For example, TFP growth may result from a firm redesigning its production process in a way that increases output while keeping the same number of machines, materials, and workers as before.

Business investment was relatively strong during the past 6 years, so that even after declining during the past year, nonresidential fixed investment remained (at 12.0 percent of GDP in the third quarter of 2001) well above its postwar average (10.7 percent of GDP). Investment in information equipment and software was especially strong after 1995, and likewise remains above its historical average share of GDP, although it, too, has fallen from levels of a year ago. As Table 1-3 shows, investment in information technologies added 0.60 percentage point to the increase in structural productivity growth after 1995. The buildup of capital outside of information technology maintained about the same pace after 1995 as before, and so did not contribute to the acceleration of productivity.

The Bureau of Labor Statistics measures labor quality in terms of the education, gender, and experience of the work force. The agency uses differences in earnings paid to workers with different characteristics to infer relative differences in productivity. Measured in this way, labor quality has risen as the education and skills of the work force have increased. Because that increase occurred at about the same rate before and after 1995, however, the contribution of labor quality to the recent acceleration in productivity has been negligible.

The rate of growth of TFP in computer-producing industries has been rising, as evidenced by the rapid decline in computer prices. Computer prices did not fall as rapidly in 2000 as they did from 1997 to 1999; however, their rapid descent resumed in 2001. Using computer prices as an indirect measure of productivity growth in the computer-producing industries, calculations indicate that computer manufacturing accounts for 0.16 percentage point of the economy-wide acceleration in productivity.

The final contribution comes from accelerating TFP in the economy outside the computer-producing industries. The contribution of this source is calculated as a residual; it captures the extent to which technological change and other business and workplace improvements outside the computer-producing industries have boosted productivity growth since 1995. This factor accounts for about 0.90 percentage point of the acceleration, or about half of the total. Taken at face value, it implies that improvements in the ways capital and labor are used

continued on next page...

Box 1-4.—*continued*

throughout the economy are central to the recent acceleration in productivity, but it is equally an illustration of the limits on our ability to account for the acceleration.

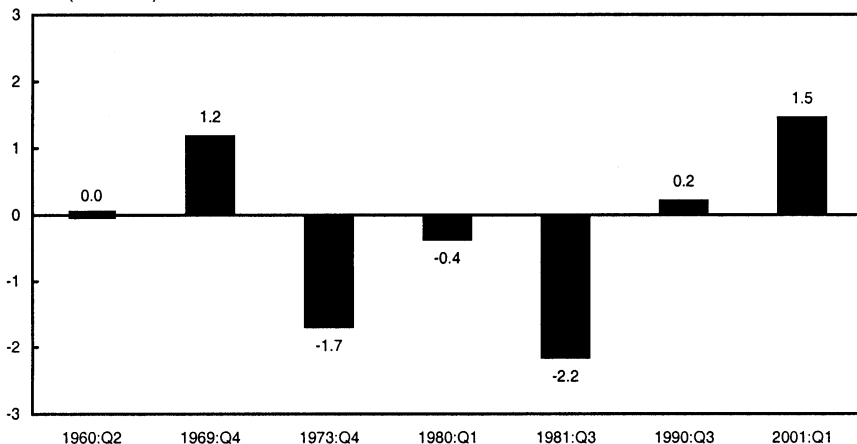
In summary, structural labor productivity growth and TFP growth remained strong through 2001. This growth argues that the New Economy remains alive and well.

The Administration believes that the economy may be able to grow faster than assumed in the budget, once the new tax policy is in place. The reductions in marginal tax rates are expected to lead to increases in labor force participation and increased entrepreneurial activity. The budget, however, uses economic assumptions that are close to the consensus of forecasters. As such, the assumptions provide a prudent, cautious basis for the budget projections.

Chart 1-9 Productivity Growth Around Business Cycle Peaks

Nonfarm productivity growth was higher around the 2001 business cycle peak than around any other peak in the past four decades.

Percent (annual rate)



Note: Dates denote quarter in which a business cycle peak occurred. Growth is measured as the four-quarter percent change for the period ending two quarters after each business cycle peak. Thus, growth shown for 2001:Q1 is for the four quarters ending in third quarter 2001.

Sources: Department of Labor (Bureau of Labor Statistics) and National Bureau of Economic Research.

TABLE 1-3.—*Accounting for the Productivity Acceleration Since 1995*
[Private nonfarm business sector; average annual rates]

Item	1973 to 1995	1995 to 2001	Change (percentage points)
Labor productivity growth rate (percent)	1.39	2.60	1.21
<i>Percentage point contributions:</i>			
Less: Business cycle effect.....	.02	-.46	-.48
Equals: Structural labor productivity	1.37	3.07	1.70
Less: Capital services.....	.72	1.29	.57
Information capital services.....	.41	1.01	.60
Other capital services.....	.31	.28	-.03
Labor quality.....	.27	.31	.04
Equals: Structural TFP	.37	1.44	1.07
Less: Computer sector TFP	.18	.35	.16
Equals: Structural TFP excluding computer sector TFP.....	.19	1.09	.90

Note.—Labor productivity is the average of income- and product-side measures of output per hour worked. Total factor productivity (TFP) is labor productivity less the contributions of capital services per hour (capital deepening) and labor quality.

Productivity for 2001 is inferred from data for the first three quarters.

Detail may not add to totals because of rounding.

Sources: Department of Commerce (Bureau of Economic Analysis) for output and computer prices; Department of Labor (Bureau of Labor Statistics-BLS) for hours and for capital services and labor quality through 1999-but the BLS figures have been adjusted for the effects of the July 2001 annual revision to the national income and product accounts; and Council of Economic Advisers for the business cycle effect, and for capital services and labor quality for 2000-2001.

The Policy Outlook: An Agenda for Economic Security

The events of 2001 have brought home to us a simple lesson: We cannot be complacent about the security of American lives. Nor can we be complacent about our rate of economic growth, our gains in productivity, or our successes in the international marketplace. The war against terrorism steps up the demands on our economy. We must seek every opportunity to remove obstacles to greater efficiency and seek new ways to combine our workers' skills, our new technologies, the drive of our entrepreneurs, the efficiency of our financial markets, and the strength of our small businesses to yield faster growth. As we integrate ever more closely our own resources, so must we also extend this integration abroad, addressing the economic roots of terrorism and securing the gains from worldwide markets in goods and capital. This is our economic challenge.

The United States boasts a more rapid long-term rate of productivity growth than do other major industrialized economies. Nonetheless, the Administration is committed to seeking opportunities to enable the economy to grow even more rapidly in the future. Growth, of course, is not an end in itself. As the President has said, we seek “prosperity with a purpose.” Economic growth raises standards of living and generates resources that may be devoted to a variety of activities in the market and beyond. Growth can fund environmental protection, the good works of charitable organizations, and a wide variety of nonmarket goods and services that benefit the United States, other industrialized economies, and developing economies alike.

To build upon our past success and rise to our new challenges, we must remove impediments to growth and build the institutions necessary to foster improved economic performance. For example, as noted in Chapter 7, one of the President’s top priorities is the U.S.-led effort toward more open global trade. Trade raises the productivity of Americans, and the United States has an opportunity to reap significant gains from future trade agreements.

Another area of interest is science and technology, long an important source of economic growth. For example, although information technology-producing industries account for roughly one-twelfth of total output, they contributed nearly a third to economic growth between 1995 and 1999. They generate some of the best and highest paying new jobs and contribute strongly to productivity growth. Technology also improves our quality of life. New agricultural technologies are increasing crop yields while reducing the need to spray herbicides and insecticides on our foods or into the atmosphere. More generally, however, it is important to establish incentives that will ensure continued growth in innovation and the new technologies that will define the 21st century. We must not only invest in basic research, but also ensure that the intellectual property of innovators is secure at home and abroad.

Getting the most out of the economy’s resources also means avoiding unnecessary costs. Prominent among these are the costs—in terms of slower economic growth and waste—associated with the Federal tax code. The entire tax system would benefit from changes to address its complexity and inefficiency. With the President’s leadership, progress has been made with the individual income tax by reducing marginal tax rates and improving tax fairness. Much more needs to be done, however, to ease the burden of taxation on the economy, to help it generate resources and increase productivity.

The current tax code imposes multiple layers of taxation, whose inefficiency costs may be as high as ½ percent of GDP a year, according to the Treasury Department. In addition, tax complexity is much more than an irritant around April 15: it, too, imposes real costs on taxpayers and the economy. Taxpayers bear the cost in terms of the billions of dollars they

spend—on recordkeeping, tax help, and their own valuable time—trying to comply. Tax compliance costs range from \$70 billion to \$125 billion a year. The economy also suffers because tax complexity raises the uncertainty surrounding business decisions, wastes resources, reduces our international competitiveness, and lowers productivity. These are costs that produce few benefits. They are largely avoidable. To get the most out of our economy, we must investigate options for tax reform.

The deregulation of the economy over the past 25 years has been a tremendous source of economic flexibility and productivity growth. We must build on that success. Deregulation of several key sectors during the 1970s and 1980s has brought substantial benefits to consumers and to the economy at large. In the 20 years following the beginning of airline deregulation, the average fare declined 33 percent in real terms. Rates for long-distance telecommunications dropped 40 to 47 percent in the 10 years following deregulation of that market.

Partly because of increased competition arising from reductions in banking regulations, banks have greatly expanded the financial services they offer customers, including important new tools for diversifying risk. Together these price declines and quality improvements across a range of deregulated industries have yielded substantial economic benefits. One study estimates the combined economic benefit of deregulating just three industries—airlines, motor carriers, and railroads—at about ½ percent of GDP each year.

This important strength of our economy must be protected against unintended interference and extended to new spheres. Competition and incentives to compete are at the core of exploiting opportunities to achieve faster growth. (Chapter 3 discusses competition policy.) The rule of law is central to efficient markets. Today, however, frivolous lawsuits and the lure of windfall recoveries are transforming America from a lawful society to a litigious one. The litigation explosion imposes a variety of costs on all of us—as much as 2 percent of GDP by one estimate—and damages the prospects for growth. The inefficiencies in our tort system are a pure waste, an unnecessary tax on our attempts to grow faster. To reduce this wasteful distortion we must address the incentives that lead to unnecessary torts and unreasonably large settlements.

We must reexamine the provision of economic security for every individual American. For example, Chapter 2 of this *Report* examines the changing nature of retirement security and documents the widely accepted need for reform. Personal accounts within the national retirement system would enhance the ability to diversify retirement portfolios, including diversifying part of retirement security away from the unsustainable current system. In doing so, they could for the first time provide rights of ownership, wealth accumulation, and inheritance within the Social Security framework.

We must design an efficient set of institutions that meet the short-run needs of displaced workers and move them quickly toward productive activities. The past year has displayed an extreme form of the shocks to which our economy may be subjected. The President's vision of economic security recognizes that many events impact the economy all the time. We should think comprehensively about these policies and focus our efforts on incentives for getting workers back to work, and quickly. Resources should be devoted flexibly to basic needs and retraining, without creating an incentive for unnecessarily long spells between jobs, because benefits extended under the wrong conditions create a "tax" when a new job is taken and those benefits are lost.

Finally, getting the most out of the economy will require an emphasis on efficiency in government as well. If government spending grows without discipline, billions of dollars will be siphoned away from private sector innovation, taxes will rise, and growth will suffer. The President's Management Agenda seeks to shift the emphasis of government toward results, not process. It aims to replace the present Federal Government hierarchy with a flatter, more responsive management structure and to establish a performance-based system. Chapter 5 of this *Report* examines fiscal federalism and shows how this approach to the structure of Federal programs may usefully be extended to the conduct of intergovernmental relations, particularly in education, welfare, and health insurance for low-income Americans.

Strengthening Retirement Security

Over the course of the 20th century, longer life expectancies and increased personal prosperity fostered a virtual revolution in the way Americans approach work and retirement. At the turn of the last century, male and female life expectancies at birth were 51.5 years and 58.3 years, respectively. Today, in contrast, life expectancy at birth is 79.6 years for males and 84.3 years for females. Because of these patterns, retirement security was not nearly the important policy issue in 1900 that it is just over a century later. And this issue is likely to grow in importance. Thanks to lifestyle improvements, less dangerous jobs, and advances in medical technology, among other reasons, the average life expectancy of a 65-year-old is projected to increase by more than 2 years over the next half century and to continue increasing even after that.

Changes in life expectancy and in fertility—American women are having fewer children—are among the forces working at the individual level that have demographic implications at the national level. These trends, together with the aging of the baby-boom generation, ensure that the population of the United States will grow older on average and remain older. Whereas in 1950 only 8 percent of the population were aged 65 or over, today those in that age group account for more than 12 percent of the population. Thirty-five years hence, they will represent more than a fifth of all Americans.

Not only are Americans living longer, but work and living arrangements have changed as well. In 1900, when fewer than 4 in 10 people reached the age of 65, approximately two-thirds of these survivors continued to work, the vast majority as farmers or laborers. In contrast, more than half of all workers today retire before their 62nd birthday, and only about 12 percent of the population work past 65. The few elderly Americans at the turn of the last century who were lucky enough to retire by 65 typically counted on extended family to support them in their old age: over 72 percent of retired men in 1900 were living with adult children. Today, fewer than one in five retirees live with extended family.

In addition to longer lives and earlier retirements, increased personal and national prosperity means that most Americans, including those in retirement, can now pursue leisure and recreational activities that were the exclusive privilege of the most affluent a century ago. To take full advantage of these changes, however, we must confront issues that previous generations of Americans, who often labored until life's end, did not have to. Planning

ahead for a comfortable, independent lifestyle during several decades without earnings from labor has become an important issue for most of the population. Amassing the resources necessary to live unsupported by others for an indefinite length of time is a task that demands forethought and preparation from the time a worker first enters the labor force. The growing importance of retirement security demands that, as we enter the 21st century, we reevaluate the strength of the Nation's many institutions for supporting workers' retirement planning efforts.

Rationale for a National Retirement System

As a starting point for thinking about retirement security, it is useful to consider a simplified scenario in which each individual passes through two distinct phases of adult life, with the length of each known with certainty. During the "working" phase, the individual uses earnings from work both to purchase goods and services for current consumption and to accumulate assets for future use. In the "retirement" phase, the individual ceases to work and instead lives on savings accumulated during the first phase. If these individuals are forward looking, then because they know how many years they will spend in retirement, they will save enough while working to ensure that they can maintain through retirement their previous level of consumption, and perhaps make a bequest to their heirs as well. Put differently, they will use their savings to "smooth" their consumption over their entire lifetime, instead of living well only while working.

In this highly simplified world, retirement security is not an issue of national concern. Prudent individuals have the incentives and the means to successfully plan for their retirement so that they will always have enough resources in their nonworking years. There is no need for government involvement in workers' planning and saving decisions.

Why, then, is retirement security a public policy concern? Traditionally, the rationale for a public system for retirement planning derives from three broad sources: insurance against uncertainty, foresight and planning failures on the part of individuals, and redistributive goals.

Insurance Against Uncertainty

So far we have deliberately ignored the many sources of uncertainty an individual faces when planning for the future. But in fact none of us who are working today knows how long we will be able to work, how much we will earn along the way, how long we will live, or what our costs of living in retirement will be. A person may plan to work for 45 years and may save accordingly, only to discover after just 40 years that, for health reasons, he or

she simply cannot work any longer. Exactly how long we will live in retirement is likewise subject to a great deal of uncertainty. Although the average remaining life expectancy of a 65-year-old today is about 18 years, nearly a quarter of those alive at 65 will live into their 90s. To guard against the pleasant “surprise” of a longer-than-expected life, an individual needs a larger nest egg than if he or she were certain of living to the average life expectancy. Uncertain and unexpected health care costs pose another potential obstacle to an individual’s retirement planning. Out-of-pocket medical expenses are fairly low for most retirees, but for some they will be catastrophically high.

Can private insurance markets effectively safeguard individuals against these contingencies? Although insurance is available against disability and against large medical costs, not all the potential shocks to an individual’s retirement security can be insured against. For example, an insured worker may find it difficult to continue to work, and therefore apply for benefits, but for various reasons the insurance company may be unable to verify that the person can indeed no longer work and is therefore entitled to benefits. This creates what economists call moral hazard: once a person is insured against running out of money in retirement, he or she has an incentive to retire earlier than in the absence of insurance, and this raises the insurer’s costs.

It has been argued that the inadequacy of existing insurance contracts against a long life without work constitutes a market failure that only a national social insurance system can address. Some have pointed to the small size of the private U.S. market for life annuities as evidence of market failure due to adverse selection: those who expect to live longer than the average will be more inclined to buy annuities; this self-selection of higher risk (from the insurers’ perspective) individuals raises the cost to insurers of providing annuities, and thus, ultimately, their price. The higher price in turn discourages still more potential annuity purchasers, further shrinking the market. But although there is evidence of some adverse selection in the U.S. annuity market, studies have shown that this is not a sufficient explanation of its small size. Among the leading alternative explanations is the existence of Social Security, which itself provides a substantial annuity to most disabled workers and retirees. Thus the seeming failure of markets for insurance against a long life may not actually be a sufficient motive for government involvement in retirement security.

Foresight and Planning

Some have suggested that even if workers could insure against all uncertainty in planning for retirement, a portion of the population may nonetheless fail to save adequately for retirement. Why might this be the case? Some people may simply be shortsighted, failing to consider fully the long-run implications of their consumption and saving decisions. Also, some “free-riders”

might intentionally neglect to accumulate retirement assets, in the expectation that they can throw themselves at the mercy of a family or government safety net that will guarantee them a minimally acceptable living standard in retirement.

Even a worker who intends to save adequately for retirement may not fully appreciate the necessity of saving enough, early enough, in his or her working life. Or that worker may miscalculate the level of savings necessary to finance a retirement that may span several decades. Saving for retirement is a continuous, lifelong process, but inadequate preparation early in life, perhaps due to lack of experience in saving for large expenditures, may have lifelong implications. Although some empirical research suggests that most people do plan and save adequately for retirement, it is ultimately unclear, given widespread expectations of government support in old age, how much people would save in the absence of existing government programs.

Redistributive Goals

For some, a third rationale for a public pension system is as a way of redistributing resources from higher income to lower income individuals. There are two reasons why government institutions for retirement security may be especially well suited for achieving redistributive goals. The first is that, because retirement benefits are provided after a person's working years are over, it is possible to redistribute based on lifetime rather than annual income. Because income in a given year is not perfectly correlated with income over a lifetime, redistribution on a lifetime basis should allow for more accurate targeting of the lifetime needy. However, as discussed below, evidence suggests that the current Social Security system accomplishes very little lifetime income redistribution. Another task for which a social security system might be uniquely suited is redistribution between generations. This sort of redistribution might be desirable if each generation is substantially wealthier than its predecessors. Indeed, in a continually growing economy this is normally the case, but it was especially the case for the generation following the Great Depression. The institution of Social Security transferred a large amount of resources from those who were younger during the Depression to those who were older, many of whom had lost much of their wealth, or were unable to accumulate it, during those years.

Unlike most events against which individuals insure, retirement and old age are not unforeseen. Accordingly, individual workers can and should take primary responsibility for their own retirement preparation. For a variety of reasons, however, retirement planning in the real world may not reflect the ideal, simplified world in which each worker can and does optimally provide for his or her own retirement. To the extent that obstacles to an individual's

ability to save adequately for retirement do exist and cannot be removed by private markets, or if certain social goals can only be achieved through government involvement in retirement planning, retirement security can be a national concern as well as a personal one. The appropriate public policy in this area depends on the nature of the impediments to successful retirement planning at the individual level, and the potential benefits from government intervention. Given the wide variety of circumstances facing individuals, however, retirement security must ultimately be the fruit of government policy that supports and enhances individuals' efforts to plan for themselves.

Sources of Retirement Security

A traditional metaphor for retirement security is that of the “three-legged stool,” where the legs—the principal sources of income in old age—are Social Security, employer-sponsored pensions, and individual savings. For elderly households as a group, the largest share of income today comes from Social Security, providing 38 percent of the total (Chart 2-1). Personal savings, which include both individual savings and employer pensions, also remain important, but a fourth income source has taken on increased salience in recent years, namely, earnings from labor. In fact, earnings from work are second only to Social Security in their contribution to the total income of the elderly. Other sources of income, including Supplemental Security Income (SSI) and other forms of public assistance, account for only a small fraction of all income for this group. In the future, the relative importance of each of these income sources will likely change; for example, many of today's younger workers will receive a larger share of income from private pensions upon retirement than did previous generations.

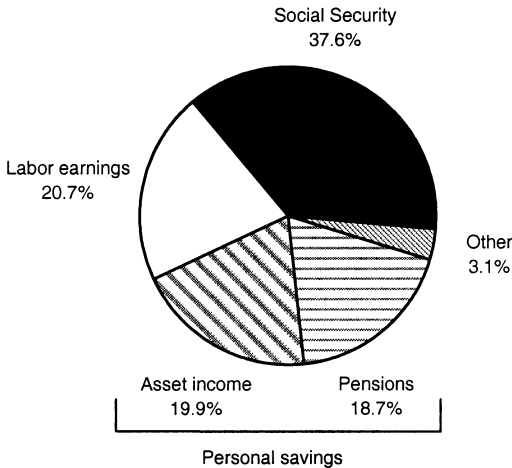
There are other sources of retirement security as well. Many people have the advantage of owning a home that they can occupy. Private, employer-provided health insurance benefits for retirees, as well as Medicare and Medicaid, also help mitigate the need for income flows in retirement.

Social Security

Social Security plays a central role in the household budgets of older Americans as a group. On average, Social Security benefits account for 58 percent of total income for elderly households (defined in this chapter as households with at least one member aged 65 or over). For the poorest elderly, Social Security is even more important. Those in the lowest income quintile obtain an average of 77 percent of their money income from Social Security benefits; for half of that group, Social Security is the sole source of income.

Chart 2-1 **Income Sources of Aged Households, 1998**

Income from work and personal savings are, together with Social Security, fundamental components of retirement security.



Note: Aged households are households with at least one member aged 65 or over.
Source: Social Security Administration.

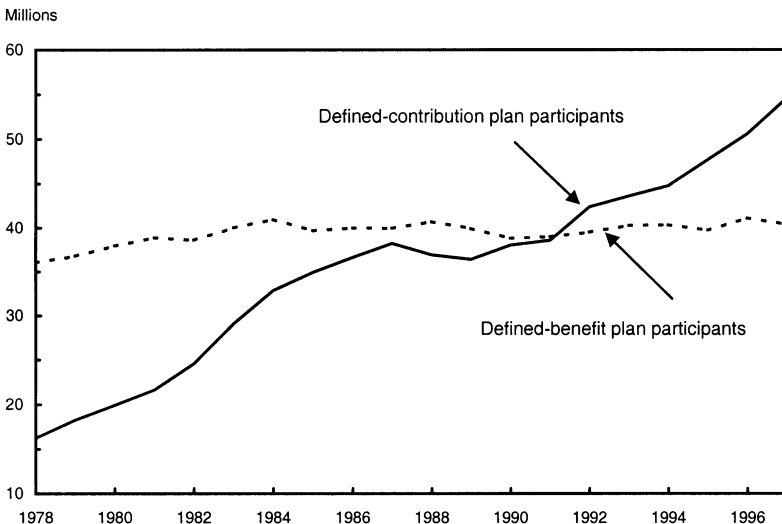
The importance of Social Security benefits in the retirement portfolios of most American households does not necessarily mean, however, that most U.S. households would be poorly prepared for retirement without it. It is sometimes suggested that, were it not for Social Security, elderly poverty rates would be much higher than they are today. But this claim is generally based on the premise that benefit payments to current Social Security beneficiaries would suddenly be ended without warning, and that workers who had contributed to the system their entire lives would be given nothing in return. That is not the same as saying that, if Social Security had never existed, the elderly poverty rate today would necessarily be higher than it is. In the absence of a national retirement security program, people would have higher after-tax income and would not expect future retirement benefits. Therefore it is reasonable to suppose that today's retirees would have saved more on their own for retirement than they actually did. Private pension coverage might also have been dramatically different in the absence of a public pension system. Consequently, it is important not to conclude, based solely on the current distribution of retirement income sources, that people would be poorly prepared for retirement under a different set of savings institutions.

Employer-Sponsored Pensions

Outside of Social Security, saving for retirement occurs in two main ways: individuals may save independently, or they may save through an employer-sponsored pension plan. Savings accumulated in employer plans have increased dramatically over the past few decades, growing from \$852 billion (in 1997 dollars) in 1978 to almost \$3.6 trillion in 1997. At the same time, there has been a pronounced trend away from defined-benefit plans, in which employees are promised specified benefit levels upon retirement, and toward defined-contribution plans, including 401(k) plans, in which employers and, often, employees make specific periodic contributions toward the employees' pension savings. The number of participants in defined-contribution plans has skyrocketed, from 16.3 million in 1978 to 54.6 million in 1997, while the number of participants in defined-benefit plans increased only slightly, from 36.1 million to 40.4 million (Chart 2-2). The growth in defined-contribution plans primarily reflects the popularity of 401(k)-type plans; participation in these had increased to 33.9 million by 1997, compared with only 7.5 million in 1984. Age-specific trends in plan participation, as well as a trend toward more companies offering plans, indicate that the rapid growth of 401(k)-type plans is likely to continue.

Chart 2-2 Pension Plan Participants by Type of Plan

Participation in defined-contribution plans has surpassed participation in defined-benefit plans.



Source: Department of Labor (Pension and Welfare Benefits Administration).

Individual Savings

Income from assets accumulated outside of private pension accounts is another important component of retirement income, accounting for about a fifth of all income for elderly households. With more than half of elderly households reporting income from nonpension assets in 1998, individual retirement savings are a widespread, but not yet ubiquitous, phenomenon. At the same time, the distinction between pension savings and other personal savings has become increasingly blurred. For example, balances from 401(k) and other pension plans may be rolled over into Individual Retirement Accounts (IRAs), which are regarded as nonpension savings. Also, small firms may establish IRAs on behalf of their workers rather than provide traditional pensions or 401(k)-type plans; such accounts would be counted as individual savings even though the employer contributes the funds.

Labor Earnings

Older workers are a vital part of the work force today and will become even more important in the future, as growth in the work force slows in response to population trends. Earnings from labor are an important component of income for a significant minority of older households. In 1998, 21 percent of elderly households reported income from labor earnings. Apparently, working is a feasible and perhaps even a desirable option for those elderly who wish to supplement income from Social Security and savings. And for those who determine that they have undersaved, or whose assets decline in value close to or during retirement, working in the traditional retirement years can be an important adjustment mechanism. Finally, today's elderly tend to be in better health than the elderly of 50 years ago, and it is likely that many more than in the past have valuable skills whose use does not require physical exertion. These considerations make the choice of continued work even easier.

Public Assistance

Compared with the four primary sources—Social Security, savings in pension plans, individual savings, and labor earnings—public assistance programs such as SSI account for an insignificant share of total income for the elderly. Nevertheless, SSI, as the retirement security program of last resort, is an important part of the safety net for a civilized society, guaranteeing a minimum income for those elderly who have little or no income from other sources. Five percent of all aged households receive some form of public assistance, and for a quarter of these it is their sole income source. Medicare and Medicaid, which provide in-kind assistance rather than cash benefits and which may have a substantial insurance value, also are an important form of public support for the elderly.

Challenges Ahead

At the beginning of the 21st century, America is taking stock of its institutions for retirement security. A monumental demographic shift is taking place, in the United States and around the world, with the result that the elderly, and programs for the elderly, will consume a growing proportion of the Nation's output. The aging of the baby-boom generation, whose oldest members will reach the age of 65 in just 9 years, together with continuing low fertility rates and increasing life expectancies, will mean that relatively fewer workers will be available to support a growing elderly population. Over the next 35 years, the number of workers for every retiree will fall from 3.3 to just 2.1—a 36 percent drop.

One clear imperative arises from this trend: Americans must take even greater responsibility for their own retirement security by increasing their personal saving. Higher personal saving has a twofold benefit. Not only will it improve personal retirement security by expanding personal wealth, but it will also have a salutary effect on the economy as a whole. When individuals save more, they add to national saving (Box 2-1). Higher national saving, in turn, means a larger capital stock and, consequently, an expanded national productive capacity for the future. This larger economic pie improves the ability of the Nation to ensure a minimum level of consumption for those members of the growing elderly population who did not earn enough while working to accumulate a large base of assets.

Public policy has an important role to play in encouraging personal saving as the foundation of retirement security. As outlined earlier, personal saving can take several different forms. Individuals may save for retirement on their own initiative. This form of saving can be encouraged through incentives in the tax system, such as the exemption of capital income from taxation. These incentives reduce the tax burden that might otherwise inhibit personal saving; however, they also have a cost in terms of forgone tax revenue, which can mean that national saving does not increase by the full amount of the increase in personal saving (see Box 2-4 below). Personal saving may also take place through employer-sponsored pension plans, which likewise receive favorable treatment under the tax code. Finally, personal saving may even take place through a public pension system, if the program allows individuals to save in accounts that they personally own. The rest of this chapter examines each of these important retirement security institutions, beginning with the institution that dominates the current retirement saving landscape: Social Security.

Box 2-1. National Saving, Personal Saving, and Growth

National saving is the sum of saving by individuals, businesses, and all levels of government, Federal, State, and local. Augmented by saving from abroad, national saving represents the total resources available for investment: the purchase of factories, equipment, houses, and inventories. When a country saves more than is necessary to replace worn-out capital goods with new capital, so that net national saving is positive, extra resources are available to expand the country's capital stock. A larger capital stock corresponds directly to a higher capacity to produce goods and services. Therefore increasing net national saving today can be an important step toward expanding the productive capacity of the economy for tomorrow.

During the 1990s, net national saving averaged about 5 percent of GDP, down from its 1960s average of nearly 11 percent. Although net national saving was fairly stable during the 1990s, its components varied widely across the decade. Net business saving grew slightly as a fraction of GDP, but there were substantial changes in the contributions of government and personal saving. Personal saving dropped sharply, from a peak of 6.5 percent of GDP in 1992 to just 0.7 percent in 2000. Over the same period, government accounts flipped from a deficit of 4.8 percent of GDP to a surplus of 2.5 percent—a total rise in saving of 7.3 percentage points. Thus, increased government saving roughly offset the decrease in personal saving. Traditionally, personal saving has been an important source of net national saving that finances investment. And because the Federal Government may not be expected to run large, persistent surpluses as an aging population strains its finances, it is imperative that the Nation increase personal saving now in order to expand the economy for the future.

Social Security: Past and Present

Origins of the Current System

The basic institution for retirement security in the United States today was established in the midst of the Great Depression, through the Social Security Act of 1935. Championed by President Franklin Roosevelt as a means of offering “some measure of protection to the average citizen and to his family...against poverty-ridden old age,” Social Security was Roosevelt’s proposal for a national system of retirement security. Ultimately, this proposal became a key part of the Nation’s response to the upheaval of traditional social and economic structures in the early decades of the 20th century.

The secular decline in agricultural employment, on which many Americans had depended for their living, worsened the ill effects of the Great Depression for many of the elderly. The loss of agricultural jobs over several previous decades had forced a shift of employment to the cities. But nonfarm workers had always fared worse than agricultural workers during economic declines, and the pattern persisted during the 1930s. Unemployment in the work force as a whole reached a high of 25 percent in 1932, but unemployment among nonfarm workers peaked at nearly 38 percent. The elderly were hit particularly hard. In 1930, 54 percent of men aged 65 and over were unemployed and looking for work, and another quarter were temporarily laid off without pay.

Aggravating the situation, the stock market crash and subsequent failure of many financial institutions wiped out the limited resources that some older workers had managed to accumulate. Without assets, employment, or traditional support systems, many of the elderly of the 1930s were in dire need of assistance. President Roosevelt sought to provide aid for the aged through his plan for social insurance. Social Security, as envisioned by Roosevelt, addressed the problem through a system in which workers contributed a portion of their earnings while working and, in turn, earned the right to collect benefits upon retirement.

Importantly, Social Security was not implemented as a program for national saving. Although the authors of the Social Security Act of 1935 intended to create a funded system, one that sets aside revenue to meet scheduled future benefits, amendments to the act in 1939 made important changes to provide more immediate relief from the widespread poverty then afflicting the elderly. As a result, Social Security is not today a fully funded system. Rather it is primarily a system for the transfer of income from one generation to the previous one: each generation pays taxes during its working years to support the current generation of retirees. Such a system is called an unfunded, or pay-as-you-go, system.

Although the Social Security system as amended in 1939 addressed the needs of the elderly during the Great Depression, today the United States faces a different challenge. The role of our retirement security institutions in enhancing the ability of relatively fewer workers to support relatively more retirees will be a critical issue as the 21st century progresses. To that end we must consider the effect of Social Security on national saving, the essential ingredient for expanding the economy's productive capacity so that it can support a vastly larger number of retirees.

Social Security and National Saving

To consider how the presence of Social Security affects national saving, one must examine the effects of the current program on two individual components of national saving: government saving and personal saving.

Government Saving

To the extent that Social Security operates as a pure income transfer program, in which taxes collected from current workers are precisely equal to the benefits paid to current retirees, the system itself has no effect on government saving. Thus the effect of Social Security on government saving hinges on how any deviation from annual budget balance in the Social Security program affects overall government budgetary policy.

When Social Security runs a surplus, so that income from payroll taxes and taxes on benefits in a given year exceeds total benefit payments in that year, as is currently the case, the government essentially has two options for the use of those excess funds. The surpluses may be spent, or they may be saved. If the surpluses are used to finance current expenditure beyond the level that would have prevailed in their absence, they do not contribute to government saving. If instead those funds are used to pay down publicly held debt (which represents the accumulation of past government dissaving), government saving increases dollar for dollar with the reduction in the debt. However, the government's ability to save by paying down its publicly held debt is limited by the amount of such debt. If all publicly held debt were to be retired, the only way that the government could continue to save through existing systems would be through investments in non-Federal securities, such as corporate or municipal bonds, or equities. This, however, would raise difficult issues about government interference in equity markets and corporate governance.

Ultimately, the contribution of Social Security to government saving depends on whether non-Social Security surpluses or deficits are affected by the annual balances in the Social Security program. If the presence of Social Security surpluses leads policymakers to increase spending or reduce taxes in the non-Social Security budget, the potential contribution of surpluses to government saving is reduced.

Many discussions of the effect of Social Security surpluses on national saving are confused by misunderstandings about the relationship between the Social Security trust fund and national saving. (Technically, there are separate trust funds for the two major Social Security programs, that for old-age and survivors insurance and that for disability insurance, but for purposes of this discussion we will combine them.) The trust fund is essentially an accounting device for keeping track of annual surpluses in the Social Security portion of the Federal budget. The balance of the trust fund represents the accumulated

value of excess revenue, net of expenses, to the Social Security system in all years that the system has run a surplus, net of accumulated deficits, as well as the interest earned on those surpluses. All Social Security surpluses are credited to the trust fund, regardless of whether they are used to finance non-Social Security spending or reduce debt, and regardless of how the existence of those surpluses affects other government spending. Consequently, the balance in the trust fund is not a measure of the Social Security program's accumulated net contribution to government saving. Rather, it merely represents the upper bound on the saving that could have happened if all Social Security surpluses had been devoted to government saving. Although Social Security has run large surpluses since 1984, these surpluses have in most years been offset by large non-Social Security deficits, suggesting that actual saving through Social Security has been far smaller than the value of the balance of the trust fund.

Personal Saving

To gauge the effect of the current Social Security system on national saving, one must consider the system's effect not only on government saving but also on personal saving. It is difficult to say definitively what personal saving would be, or would have been in the past, in the absence of Social Security, but reasoning and empirical evidence can be useful guides. As discussed previously, careful consideration suggests that Social Security may act as a substitute for retirement saving. Instead of saving, a worker pays taxes on his or her wages and, upon retirement, instead of using past savings to finance consumption, the worker receives a check from the government. In this way Social Security can negatively affect personal—and, consequently, national—saving.

For a number of reasons, however, a rational worker might decide to reduce personal saving less than dollar for dollar with increases in expected Social Security wealth. A worker may underestimate the expected value of Social Security benefits or simply not believe that the scheduled benefits will be forthcoming upon retirement. This is particularly possible in the current climate, when revenue has been projected to fall short of projected benefits. Another possibility is that Social Security affects saving behavior through an effect on retirement behavior (Box 2-2). If Social Security makes retirement an attainable goal and thus prompts workers to plan for an earlier retirement, they may actually save more than they would have in the absence of the program.

Clearly, economic reasoning alone does not lead to an unambiguous conclusion regarding the effect of Social Security on personal saving behavior. Therefore we must rely on empirical analysis to learn about the actual effect of the program on personal saving and, ultimately, on national

Box 2-2. Does Social Security Alter Retirement Behavior?

Careful economic analysis indicates that the current Social Security system does indeed have the potential to alter workers' retirement behavior. Incentives that affect retirement could come through a number of different channels. For some, Social Security provides more retirement wealth than they would have chosen to provide for themselves through their own saving; the resulting benefit windfall in old age could induce their earlier retirement. Also, Social Security adjusts benefits for those who retire and begin receiving benefits before or after Social Security's normal retirement age, currently 65 years and 6 months; if these adjustments deviate from what is actuarially fair, they may create incentives favoring retirement at a particular age. If those who work past 65 do not get an actuarially fair increase in benefits, for example, people might be inclined to retire earlier than otherwise. People with above- and below-average life expectancies will also have varying retirement incentives related to the benefit formula. Social Security may also have affected retirement behavior simply by establishing the social convention that 65 is the "normal" retirement age.

Since rational analysis does not lead to a definite conclusion about how Social Security affects retirement behavior, we must examine empirical retirement patterns in order to understand the ultimate effect of this complex system of incentives. Early retirement has become more common in the United States, as well as in other countries, in recent decades. And a considerable amount of evidence indicates that the relaxation of early retirement rules and the increased availability of benefits at earlier ages in the 1950s and 1960s resulted in these pronounced trends toward earlier retirement. Cross-sectional evidence using only U.S. data has been less clear in establishing a link between Social Security expansions and declines in the average retirement age. Some research suggests that changes in pension wealth have had a much stronger effect on retirement trends than have Social Security changes; this research finds that any Social Security effect accounts for only about 1 percentage point of the 20-percentage-point decrease in the labor force participation rate for males aged 55 to 64 between 1950 and 1989.

saving. Even then the results are less than clear, but in a recent Congressional Budget Office survey, 24 of 28 cross-sectional studies found a negative impact of increases in Social Security wealth on private saving. If Social Security does negatively impact private saving, as much evidence suggests, it may be inhibiting national saving and, consequently, economic growth.

The Future of Social Security

In assessing the role of Social Security as a retirement security institution for the 21st century, two related, yet conceptually distinct, issues must be addressed. The first is the fundamental question about the degree to which government transfers should supplement personal saving for retirement. In the extreme, the essential choice is between a savings-based program in which individuals accumulate assets, and a program that simply transfers income from younger to older generations.

The second issue is that the current Social Security system, which resembles more the latter system than the former, is on a fiscally unsustainable course as a result of the demographic changes discussed earlier: the aging of the population and the consequent projected decline in the ratio of workers to retirees. These changes make it impossible to afford the currently projected rate of benefit growth without large tax increases or other fundamental changes to the system. The following sections deal with each of these issues in turn.

Advantages of Personal Accounts

One of the President's principles for strengthening Social Security is that modernization must include individually controlled, voluntary personal retirement accounts to augment the Social Security safety net. Under such a system, a worker could direct a portion of his or her payroll taxes, or possibly an additional voluntary contribution, into a personal account that he or she would legally own. The worker would then choose, from a variety of options, how the assets in the account are to be invested. Upon retirement, the worker would have access to the accumulated assets, which could be used to purchase an annuity, provide a bequest to heirs, or make withdrawals from as needed. Workers who choose to direct a portion of their existing payroll taxes into private accounts could expect a higher combined level of benefits, because an annuity funded by the personal accounts would have a higher expected value than the benefits from the traditional system that are being partially replaced by the account contributions. Personal accounts would thus represent a voluntary means by which a worker could supplement benefits from the pay-as-you-go portion of Social Security. As such, they could

provide the foundation for a return to individual-based retirement security that takes advantage of the safety net aspects of Social Security and the strengths of individual choice and wealth accumulation.

Although the introduction of personal accounts within Social Security would represent the most significant change in the program since its inception, the idea itself is not new. In President Roosevelt's message to Congress on Social Security on January 17, 1935, he stated that one of his three principles for the program was "voluntary contributory annuities by which individual initiative can increase the annual amounts received in old age." In this light, a system of personal accounts would appear to be the next step in the natural evolution of the program. In addition, many other nations, from the United Kingdom to Australia to former socialist countries like Kazakhstan, have included personal accounts as an important part of their national retirement program.

A Social Security system that includes an element of personal accounts would offer many advantages over the current regime. These include personal ownership of accounts, bequeathability of account assets, better diversification of risk, reduced distortion of work incentives, and the potential for higher national saving. We discuss each in turn.

Ownership

From the perspective of an individual worker, perhaps the most striking difference between personal accounts and the current system is ownership. Under Social Security, a worker's retirement security depends not on the assets that worker possesses, but on the hope that future Congresses will raise taxes on the next generation of workers by a sufficient amount to pay scheduled benefits. In fact, the Supreme Court ruled in *Flemming v. Nestor* (1960) that workers and beneficiaries have no legal ownership claim to their benefits, even after a lifetime of contributing to the system. A personal account, on the other hand, would be the legal property of the worker who contributed to it and whose name it bears. Regardless of the financial situation of the government, a worker would be legally entitled to the assets in his or her account upon retirement.

The security that comes from this ownership, however, is not the only benefit that ownership offers. Asset ownership and wealth accumulation could be a positive new experience for many Americans. In 1998 the median U.S. household owned only \$17,400 worth of financial assets, including sums in retirement accounts. Four out of every nine households saved nothing at all during the year. For many families, contributions to individual Social Security accounts may represent their only chance to build privately held financial assets and wealth. The experience of selecting investments and observing the miracle of compound interest at work might help many workers overcome existing social and informational barriers to asset

ownership. Research has shown, in fact, that the experience of managing a pension account may actually encourage workers to save more outside of their pension than they otherwise would. Accordingly, personal accounts could have an important effect on the personal saving rate.

Studies have suggested a broad range of other benefits from asset ownership as well. Owning assets makes people more oriented toward the future, more likely to take calculated risks, and more likely to participate in the political process. Financial assets have also been found to be associated with positive physical and mental health effects, particularly for those between the ages of 65 and 84. Married couples with property and financial assets are less likely to divorce than couples without assets. Finally, a survey of participants in an experimental program designed to help the poor save and accumulate assets has yielded important information on the benefits of asset ownership. Program participants report feeling more economically secure, are more likely to make education plans for themselves and their children, and are more likely to plan for retirement because of their asset accounts. They also reported that they are more likely to increase their work hours or increase their income in other ways. They are more confident about the future and feel more in control of their lives because they are saving.

Bequeathability and Redistribution

Recent research has shown that Social Security is only mildly progressive and may even be regressive on a lifetime basis, despite an explicitly progressive benefit formula (Box 2-3). One reason for this seeming paradox is that people with higher incomes tend to live longer than those with lower incomes. Because Social Security retirement benefits cease at the death of the insured individual (or the individual's surviving spouse), those with shorter lifespans will earn lower returns on their contributions, all else equal. Additionally, research has indicated that current Social Security arrangements may substantially increase the inequality of the wealth distribution by depressing bequests by low- and moderate-income households who might have accumulated bequeathable assets in the absence of the program. Depending on the degree of annuitization of assets that is required, and on other program design elements, a system that includes personal accounts has the potential to reduce some of the regressive tendencies of the current system. Accountholders who die earlier than the average might be able to pass on to their heirs a portion of the wealth in their personal accounts; this would partly correct for the disadvantage many higher mortality, lower income groups face under Social Security today. The introduction of personal accounts might also provide an opportunity for the creation of a more progressive benefit structure for the pay-as-you-go portion of Social Security.

Box 2-3. The Effect of Social Security on Income Distribution

One of the traditional justifications for a government role in retirement security institutions is the potential to use these institutions as tools for redistribution, especially redistribution based on lifetime income. It is often argued that Social Security is redistributive along a number of different dimensions. However, in large part because of heterogeneity among individuals in marital status and life expectancy, much less redistribution on a lifetime basis occurs under the current system than is widely believed.

Progressivity. The design of the Social Security benefit formula is explicitly progressive at the individual level. When redistribution is considered at the family level, however, the system looks less progressive than the benefit formula seems to imply. There are two reasons for the potential disparity. First, many low-income individuals are members of high-income households; if such a low-income person receives a high return on Social Security, the system will appear redistributive on an individual, but not on a household, basis. Second, the ability to collect benefits on the basis of a spouse's earnings also fosters redistribution to low- or zero-income individuals with high-income spouses. Research has shown that the system hardly redistributes to poor families at all.

Redistribution by marital status. Rates of return are considerably higher for single-earner couples than for dual earners. For medium earners (as defined by the Social Security actuaries) retiring in 2000, for example, the 4.75 percent rate of return for a one-earner couple was very nearly twice that for a two-earner couple. There is also substantial redistribution from single individuals to married couples. A man retiring in 2000 with medium earnings and with a wife who never worked would receive a rate of return on Social Security that exceeded twice the return obtained by an identical man who had never married.

Redistribution by race. Largely because of differences in mortality rates, African Americans receive on average nearly \$21,000 less, on a lifetime basis, from Social Security's retirement program than whites with similar income and marital status, according to recent research. Other research finds that rates of return for African Americans from Social Security are approximately half a percentage point lower than for whites of the same marital status. Survivor benefits that pay benefits to the spouse or the children of deceased workers partly, but not completely, compensate for the negative effect of mortality on returns. The provision of disability insurance through Social Security also improves returns for African Americans, who are more likely than other groups to collect disability benefits.

Diversification of Risk

Another important advantage of adding personal accounts to a pay-as-you-go system is the potential to diversify the risks inherent in such systems. Under the present Social Security system, the ultimate rate of return earned by a participant is subject to political risk. Without structural reform of Social Security, workers and retirees will face significant uncertainty about how future policymakers will alter system revenues and outlays to avoid system insolvency. These actions would directly impact the rate of return earned by participants in the system.

Although funds invested in equities through a personal account can be expected to earn a higher rate of return than funds in a pay-as-you-go system, investment in equities does expose participants to some degree of financial market volatility. However, as long as the market risk associated with equity investment is not perfectly correlated with the demographic and political risks of a pay-as-you-go system, a mixed system of personal accounts and pay-as-you-go benefits offers an opportunity for better diversification than either a pure pay-as-you-go or a pure investment-based system. This diversification could be especially important to low-income workers whose sole source of retirement income is Social Security, and who are consequently less well diversified than wealthier individuals who are able to hold private financial assets in addition to expecting scheduled Social Security benefits.

Labor Supply

A reform of Social Security that includes personal accounts would reduce the economic inefficiency arising from elements of the current Social Security system that distort labor supply. For many workers, including younger workers and secondary earners in a household, the present structure of the benefit formula means that the marginal dollar of Social Security payroll taxes that they pay does nothing to raise their benefits at retirement. When this is the case, that worker's effective marginal tax rate is increased by the full amount of the payroll tax (provided the worker is earning less than the Social Security cap on taxable earnings, which is \$84,900 in 2002). Since a higher marginal tax rate corresponds to a lower return to work, the Social Security payroll tax may discourage work by many low- and middle-income workers. In a system that includes personal accounts, however, the link between current contributions and future income is stronger, and there is more incentive to work than under the current system.

The current Social Security system may also distort labor supply behavior through its effect on retirement age. Growth of assets in personal accounts, however, is governed by the rate of return on those assets rather than by the potentially distortionary rules of a defined-benefit program. Thus workers

with income from personal accounts may be less influenced in their choice of retirement age than if their income from Social Security depended entirely on the particular structure of the Social Security benefit formula.

Higher National Saving

Establishing personal accounts has the potential to raise national saving, thus expanding the capital stock and increasing productive capacity, so that a relatively smaller labor force can support a relatively larger population of beneficiaries. If Social Security payroll taxes were saved in personal accounts rather than used to finance an increase in non-Social Security government spending, national saving would likely be higher. Although it is theoretically possible, within the current system, for the government to save those excess payroll tax revenues, the experience of the last 20 years has shown that, even for laudable reasons, it is difficult to do so. The only truly effective way to preserve a Social Security surplus is to put it safely beyond the grasp of those who would spend it for other purposes, by depositing it into personal accounts. Doing so would also make the rest of the budget more transparent, because any non-Social Security spending in excess of non-Social Security revenue would clearly have to be financed by issuing public debt or increasing non-Social Security revenue.

The degree to which saving in personal accounts would increase national saving would depend in part on whether households changed their other personal saving in response to the accounts. Although ownership of a personal account might dampen other personal saving to some extent, it is unlikely that the effect would be large enough to completely offset the expected increase in national saving. As long as other personal saving were not reduced (and personal borrowing were not increased) one for one with contributions to personal accounts, the net effect of the accounts would likely be to increase national saving (provided that any forgone income tax revenue is less than the increase in personal saving). Since many low-income workers today have very little saving to reduce, overall personal saving should certainly not fall one for one with increases in personal account saving.

International Experience with Personal Accounts

The United States would by no means be the first country to incorporate an element of personal accounts into its social security system. The finances of pay-as-you-go pension systems around the world have come under pressure, due to unachievable benefit commitments and an over-60 population that will rise from 9 to 16 percent of the global population over the next three decades. Finding their pay-as-you-go systems overextended, a growing number of countries have instituted major structural reforms, including

downsizing traditional defined-benefit public pension systems and relying increasingly on a personal account-based system that is fully funded and based on defined contributions. In 1981 Chile became the first country to implement a mandatory, funded system based on personal accounts. Switzerland, the Netherlands, and the United Kingdom also instituted major structural reforms in this direction during the 1980s. After a flurry of reform activity in the 1990s, at least 22 countries have now added funded systems or partially privatized part of the old system. Three more European countries have also advanced proposals. The reformers are a geographically and economically diverse set of nations, including 6 high-income industrial countries, 10 Latin American countries, and 5 former socialist countries. China's autonomous province of Hong Kong has also pursued reform along these lines.

International experience shows that pension reform seems to be one of the most politically difficult reforms to undertake, but also that when a pension reform is actually implemented and people are given a choice, they overwhelmingly choose personal accounts. The case of Uruguay illustrates the popularity of personal accounts in countries that have undertaken reforms, despite the political rhetoric that preceded those changes. In that country, there are 600,000 contributors in the national social security system. Before reform, a number of surveys showed that only 80,000 people would opt for personal accounts. When the system was implemented and people were given a choice, however, more than 400,000 chose personal accounts.

In evaluating America's reform options in light of the experiences of other countries, one should keep in mind the important advantages that this Nation possesses. Indeed, few of the many countries that have converted to personal account-based public pension systems were in as favorable a position to do so as the United States. First and foremost, the United States has the best-developed financial markets in the world, with a wide variety of investment vehicles and about 40 percent of world equity market capitalization. This long and broad experience with financial markets at the institutional level offers a solid foundation for a system of personal accounts. Another institutional advantage is the advanced degree of development of our private pension system. In 2000, 51 percent of all wage and salary workers had some type of private pension coverage at their current job, and almost 80 percent of those eligible participated in defined-contribution plans. This experience with defined-contribution plans means that a sizable portion of the population is already well grounded in the principles necessary for understanding and managing personal accounts. Additionally, the prevalence of these private plans means that much of the basic financial infrastructure needed for personal accounts is already in place.

The Financial Sustainability of Social Security

A system of personal accounts based on individual wealth accumulation has many advantages over alternative methods of financing retirement. Whether or not personal accounts become part of the solution, however, Social Security reform is a necessity. The Social Security system faces a severe, long-term financing shortfall. Put simply, the system does not have a dedicated income stream sufficient to pay the benefits scheduled under current law. According to intermediate projections of the Social Security Administration, by 2016 the system will begin running persistent cash flow deficits; by 2050 the current benefit structure would cost nearly 18 percent of the Nation's payroll, whereas program revenue would be just over 13 percent.

Adverse Demographic Trends

The need for reform arises because the structure of the current system is on a collision course with the changing demographics of our country. In a funded pension system, the resources available to pay retirement benefits depend on the assets put into the system for that purpose and the rate of return those assets earn, not on demographics. Because Social Security is unfunded, however, demographic trends can play an important role in system finances and in determining the rate of return that workers earn on their Social Security contributions. The ability of an unfunded Social Security system to pay benefits to retirees in a given year depends on the size of the taxable wage base in that year. Consequently, demographic trends that decrease the number of workers available to support each beneficiary, referred to as the worker-to-beneficiary ratio, reduce the ability of an unfunded system to pay retirees without raising taxes or reducing benefits. In the United States, lagging birthrates and increasing life expectancies, together with the aging of the baby-boom generation, will put tremendous pressure on the Social Security system.

The baby-boom generation, defined as those Americans born between 1946 and 1964, was a major demographic boon for the United States. In particular, the birth of many new workers-to-be during those years was a major blessing for a pay-as-you-go Social Security system that operates best with a large number of workers for each benefit recipient. The total U.S. fertility rate (roughly speaking, the number of children the average woman would have in her lifetime, based on current births) climbed steadily through the 1940s and 1950s, from 2.2 children per woman in 1940 to a peak of 3.7 in 1957. Unfortunately for Social Security, which depends on the younger generations to finance the retirement of workers in the older generation, fertility rates subsequently fell to pre-baby boom rates. By the mid-1970s, the total fertility rate had fallen by half from its peak, to just 1.8. It presently stands at around 2 children per woman and is not projected to change substantially in the foreseeable future.

These lower birthrates are especially problematic given the aging of the baby-boom generation. Beginning in 2008, the first of the baby boomers will be eligible for early retirement under Social Security rules. By 2026 the youngest boomers will have reached age 62, and most of that generation will have retired and begun to collect Social Security benefits, putting a substantial burden on the system.

Another significant factor in the aging of the population is the fact that, as noted previously, Americans are living longer than ever before. Of the cohort born in 1875—the first to receive Social Security benefits—only 40 percent survived to age 65, and those who did lived an average of 12.7 additional years. In contrast, 69 percent of males born in 1935 lived to age 65, and those who did could expect to survive an additional 16.2 years on average. And among males born in 1985, 84 percent are expected to survive to age 65, and those who do will be able to look forward to an average of 19.1 years of life in old age.

This trend toward increasing longevity, combined with the low birthrate, implies an aging of the overall population. The share of the population over age 65 will increase from 12.4 percent today to 20.9 percent by the 2050s. Moreover, the “oldest old,” those aged 85 and older, will more than double their share of the population, from 1.5 percent today to 3.7 percent in 2050.

The combined effect of these fertility and longevity patterns is to reduce the number of people of working age relative to the number collecting Social Security benefits. Chart 2-3 displays the declining ratio of 20- to 64-year-olds to individuals aged 65 and over. The change in this ratio over time reflects fertility and longevity trends and, together with changes in labor supply and Social Security rules, accounts for the change in the worker-to-beneficiary ratio discussed previously. Today there are approximately 4.8 people of working age for each person 65 or over; by 2030 that ratio will have dropped to 2.8, and by 2075 it will be 2.4. The bottom line is that there will be relatively fewer people of working age to support a growing elderly population. Because Social Security is primarily unfunded in its current form, the declining ratio of young to old foretells serious solvency problems for Social Security.

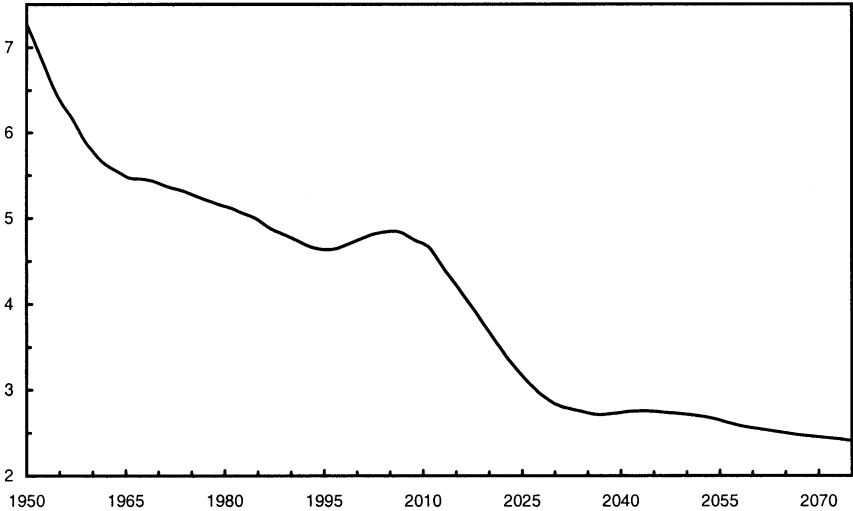
Insolvency on the Horizon

Beginning in 2016, as noted previously, payments to Social Security beneficiaries are projected to exceed revenue to Social Security from payroll taxes and taxes on benefits. The result will be annual cash flow deficits for the system, which are projected to continue indefinitely. Although the trust fund will have a positive balance at that time, allowing Social Security to continue paying full benefits, the Federal Government will be forced to find a way to finance those benefit payments that exceed the revenue generated by payroll and benefit taxation. In that first year of cash deficits, the projected shortfall amounts to \$17.4 billion in 2001 dollars. Just 4 years later, however, the

Chart 2-3 Ratio of Working-Age to Retirement-Age Persons

The declining number of younger, relative to older, Americans is a growing problem for a pay-as-you-go Social Security system.

Ratio



Note: Working-age persons are aged 20-64, and retirement-age persons are aged 65 and over.
Source: Social Security Administration.

annual deficit will have jumped to \$99.3 billion. By 2030 Social Security will face a \$270.8 billion annual cash shortfall, representing over 4 percent of taxable payroll, and deficits will continue to worsen for the foreseeable future. Until the trust fund becomes insolvent in 2038, Social Security will finance these cash deficits by redeeming bonds from the trust fund, but this will put a large strain on the rest of the Federal Government's budget. Financing these cash shortfalls, therefore, requires that the government increase revenue to the system or slow the growth rate of outlays.

Meanwhile, because of the aging of the population, the non-Social Security portion of the Federal budget will face increasing pressure from other sources as well, further complicating the overall fiscal situation. Medicare will demand an increasing share of the Nation's resources, reducing the government's flexibility in addressing Social Security financing issues within the budget. An amount equivalent to 2.3 percent of GDP goes to Medicare today, and the program's claim on GDP is projected to rise to 8.5 percent by 2075. Absent structural reforms, Medicare and Social Security together will consume more than 15 percent of GDP by that year. By comparison, all personal income taxes paid to the Federal Government today amount to only about 9 percent of GDP.

Restoring Fiscal Balance

To solve the serious long-term financing shortfall facing Social Security, some combination of the following two measures is required:

- Future Social Security resources must be increased beyond currently legislated levels, or
- Future Social Security spending growth must be reduced from currently legislated levels.

Every policy proposal to solve the Social Security financing problem, including those that utilize personal accounts, must follow one or both of these two approaches. Thus restoration of fiscal balance to the system will require some combination of a resource increase to support the benefit structure and a reduction in the rate of traditional benefit growth to a level that can be paid by currently legislated tax rates.

Regardless of the path selected, personal accounts would provide participants with the opportunity to increase their expected benefits by investing in a diversified portfolio of assets. Historically, private sector investments have consistently delivered higher returns than government securities over long time horizons. If the future is like the past, personal accounts could provide individuals with higher benefits than in the absence of personal accounts. As such, personal accounts provide an opportunity to increase the expected benefits of participants relative to any comparably funded system that lacks personal accounts, and are therefore an important component of plans to restore fiscal soundness to the Social Security system.

Increases in the system's resources could take a number of forms. One possibility is an increase in the payroll tax, either by an increase in tax rates or by an expansion of the taxable earnings base. For perspective, if taxes were increased each year just enough to cover the contemporaneous benefit shortfall, combined employer and employee Social Security payroll tax rates would need to rise from their current level of 12.4 percent to 14.1 percent by 2020, 16.6 percent by 2030, and 17 percent by 2040. Increasing payroll taxes on this basis would be detrimental to economic growth and ultimately unsustainable, and the President, in enunciating his principles of Social Security reform, has ruled out such an approach. Alternatively, current law benefits could be paid by raising general revenue to support the system, but this would require a comparable income tax increase or a comparable reduction in non-Social Security spending. Yet another possibility is for the government to borrow the necessary funds. Any borrowing, however, would have to be repaid by some future generation through higher taxes or decreased spending. Debt financing alone cannot be a permanent solution in any case, because in the absence of structural reform, the debt could never be repaid, as Social Security's cash shortfalls are projected to continue indefinitely.

An alternative to increasing revenue to pay for currently legislated benefit payments is to place the benefit formula on a more sustainable course. The President has made it clear that benefits for current retirees, and for persons nearing retirement, should not be changed. However, under the existing benefit formula, benefits for future retirees are scheduled to rise substantially above current levels in real terms. One way to achieve fiscal sustainability is to restrain the rate of future benefit growth.

Many specific policy changes could be used to slow the rate of benefit growth. For example, future growth in initial benefits could be indexed by price growth rather than by wage growth in the economy, as now. According to intermediate projections of the Social Security trustees, wage growth is expected to exceed price growth by approximately 1 percentage point a year. Indexing benefits to price inflation would keep benefits fixed at their current real level, significantly reducing future system costs. In fact, according to the Social Security actuaries, price indexing alone would suffice to close the entire 75-year actuarial deficit. This approach would entail no real benefit reductions or tax increases relative to current tax and benefit levels. Another possible change to reduce benefit growth would be to adjust benefit levels in accordance with increases in life expectancy.

Personal Accounts and Fiscal Sustainability

In assessing any reform proposal, it is important to remember that the need for action to restore fiscal sustainability is independent of whether personal accounts are implemented. It would be possible to restore fiscal sustainability without personal accounts, simply by raising taxes or reducing benefit growth, and it would be possible to introduce personal accounts in a way that does not contribute to fiscal sustainability. A well-designed reform package, however, would provide workers with the opportunity to benefit from personal accounts and would, simultaneously, help restore fiscal soundness to the Social Security system.

Many specific design elements in Social Security reform will determine how personal accounts and fiscal sustainability will interact. It is possible to design personal accounts that are wholly separate from the traditional Social Security system; for example, they could be funded entirely by new contributions or from general revenue. In that case the accounts would neither improve nor worsen the underlying fiscal status of the traditional system. On the other hand, many proposals would integrate the two systems by allowing for a redirection of current payroll tax revenue to fund the personal accounts. In this type of proposal, it is appropriate to construct a “benefit offset,” that is, an amount by which a person can choose to have his or her traditional benefit reduced in order to have the opportunity to invest in the personal account. Depending on how this offset is constructed, the decision to choose

a personal account can have implications for system finances. If, on the one hand, the individual is required to forgo a portion of benefits that is actuarially equivalent to the portion that would have been paid with those redirected payroll taxes, the long-run effect of this choice on system finances will be neutral. On the other hand, if the benefit offset deviates from actuarial equivalence, it can have a long-run effect on system finances.

This discussion has focused on the long-run fiscal effects of specific alternative reforms. During a temporary transition period, movement to a system of personal accounts would require additional funds in order to make scheduled payments to current and near-retirees while simultaneously funding the new personal accounts. This is sometimes referred to as a transition cost, but it is more appropriate to think of it as a national economic investment. These funds would not be spent on consumption, but rather saved to finance future retirement benefits through the personal accounts. This prefunding of benefits is the mechanism by which national saving will be increased. Indeed, ultimately, it is only by such a reduction in consumption that saving can be increased.

Baselines for Comparison

As the Nation debates plans to reform Social Security and considers personal accounts as a component of that reform, it is important to keep in mind the appropriateness of the standards by which any proposed reform is assessed. It has become clear that the Social Security system is unsustainable in its present form. As noted above, options for resolving the system's long-range financing issues include increasing system revenue and reducing the rate of growth of system outlays. Because the full benefits scheduled under current law cannot be paid without taking one or the other of these steps, or some combination, it is not appropriate to compare a reformed system with the present, unsustainable system without specifying how "current law" will be brought into fiscal balance. In other words, one set of options for achieving sustainability should be compared with other sets of options for doing so; comparing any set of options for achieving sustainability with the current unsustainable program is neither meaningful economically nor informative to the public.

There are many alternative baselines that one could use in this comparison. One approach is to measure reform proposals against the benefit levels that could feasibly be paid given current Social Security payroll tax rates. In 2040, for example, without tax increases, benefits would have to be 27 percent lower than under current law. Alternatively, if one wishes to use currently scheduled benefits as a basis for comparison, it is necessary to specify the source of the funding required to finance those benefits.

The effectiveness of a particular proposal for reform cannot be judged solely on the basis of tax rates and benefit levels under that proposal, however. The change in the total projected future burden on taxpayers resulting from the reform must also be considered. This total projected burden is the sum of explicit national debt and the present value of the benefits scheduled to be paid under today's primarily pay-as-you-go system. Although the present value of currently scheduled benefit payments to future Social Security recipients can be changed through reform of the system, the value of this implicit burden can be thought of as a form of implicit "debt" on the part of the government. If the current schedule of future benefit payments were binding and were feasible, which it is not, the government would find itself in the situation of paying people alive today about \$10 trillion more in future benefits than it would have collected from them in the form of future payroll taxes. A complete accounting of a Social Security reform's effect on national saving and the country's fiscal situation should recognize the change in this potential burden on the Federal Government.

It is important to understand how any proposed reform would change the combined level of the explicit debt and the implicit burden imposed by scheduled benefits. For example, a change to the current system could make the country as a whole better off by decreasing the total national obligation even while increasing explicit, publicly held debt. This scenario could arise if a transition to a new system with a lower total projected burden were financed by converting a portion of future benefit payments into explicit debt. Under current accounting rules, which document only explicit debt, the Nation would appear to be worse off after such a transition. In reality, however, the overall fiscal health of the Nation might actually have improved. Because of this discrepancy, it is essential that reform proposals clearly specify not only what benefits and taxes would be after reform, but also how the total future burden of the program on future generations would change.

Other Sources of Retirement Security

As the earlier discussion of current sources of retirement income emphasized, Social Security is not the sole source of support for the elderly. Nor is it meant to be. The current average Social Security benefit, for instance, is equal to only about 36 percent of the average worker's wage. Already today, workers need to supplement their Social Security benefits with income from other sources in order to maintain a lifestyle in retirement similar to what they enjoyed while working. With rising out-of-pocket medical expenditures, an increasing number of years spent in retirement, and an unsustainable Social Security system, the need to diversify retirement wealth is imperative

as we move into the future. Personal saving, undertaken both independently and through employer-sponsored pension plans, is an increasingly important element of retirement security.

The role of public policy in ensuring retirement security by no means ends with Social Security. The government can continue to adopt tax policies that reward and encourage the efforts of workers to plan for their own future. Creating a friendly environment for retirement saving requires an awareness of the ways in which the tax structure might encourage or discourage people's efforts to save. The income tax, one of the most basic components of the tax system, may discourage saving by reducing after-tax returns. This is particularly true for capital income, which is often taxed twice: once at the level of the corporation, and once at the individual level. Recognizing this fact, certain mechanisms that reduce the burden of the income tax have been built into the tax system in order to encourage saving for a variety of purposes, but especially for retirement. IRAs and 401(k) plans are the most prominent examples of such tax-preferred vehicles, but there are many less well known arrangements as well.

Employer-Sponsored Pension Plans

One important means by which the government encourages saving for retirement is through provisions in the tax code that grant special tax status to profit-sharing and employer-sponsored pension plans. Generally, contributions made by an employer to a defined-benefit or a defined-contribution plan, including a 401(k) plan, on behalf of an employee are not included in the employee's taxable income. This tax advantage gives employers an incentive to sponsor pension plans for their employees, thus increasing retirement saving. These plans also have the advantage that earnings on invested contributions are not taxed until they are withdrawn, offering participants the possibility of being subject to a lower tax rate in retirement. Moreover, even if the owner's tax rate has not declined, there is an advantage from the deferral of taxes on returns accumulated within the account, effectively lowering the tax rate on such saving.

Employer-sponsored pensions will continue to increase in importance as a source of retirement income, as evidenced by the fact that a substantially larger share of current workers than of current retirees have pension coverage. As noted earlier, the 401(k) plan in particular has become increasingly popular in recent years. In contrast to most other defined-benefit and defined-contribution plans, in which only the employer contributes to the plan, the employer, the employee, or both may make contributions to a 401(k) plan. These plans are expected to account for a growing share of retirement income. By some estimates, assets in such plans could rival or

even exceed total Social Security wealth by the time workers currently in their early 30s retire. Provisions of the Economic Growth and Tax Reform Reconciliation Act (EGTRRA), enacted in 2001, will further encourage this form of saving by increasing the limit on individual contributions to 401(k)-type plans, as well as the limit on an employer's deduction for contributions to certain types of defined-contribution plans. Additionally, workers aged 50 and over will now be eligible to make "catch-up" contributions to their 401(k)-type plans; this will help workers who might not have saved in past years.

Although pension assets represent a large and growing share of retirement wealth, pension coverage remains far from universal. In recent years almost half of retirees lacked pension income or annuities, and 49 percent of those employed lacked a pension plan. With this fact in mind, changes in tax policy and pension law that further encourage all employers to provide plans for their employees should continue to be explored.

The government must also work to expand its outreach to employers, especially small businesses, to encourage retirement plan sponsorship. It should eliminate artificial barriers to employers wishing to provide sensible retirement advice to those who participate in pension plans. Also needed is increased assistance to employers, plan sponsors, service providers, participants, and beneficiaries, to better inform these parties of their responsibilities under the law. This compliance assistance will ultimately lower the cost of investigations, judicial dispute resolution, and plan administration. Reducing such burdens should remain an ongoing Federal goal, because efforts to that end can yield higher retirement income for working Americans.

Individual Saving

Personal saving independent of profit-sharing plans and employer-sponsored pensions is the third important component of retirement security. Public policy has aimed to encourage such saving as well, most notably through IRAs, which allow individuals to save for retirement on a tax-preferred basis. Contributions to traditional IRAs, like those to most employer-sponsored pensions, are tax-deductible under certain conditions, and earnings on investments in these accounts are tax-deferred. Contributions to Roth IRAs are not tax-deductible, but the earnings on these contributions are generally tax-free. IRAs provide an important incentive for individuals, some of whom may not be covered by an employer-sponsored pension plan, to invest for retirement. And research has shown that IRAs are effective in increasing personal saving (Box 2-4). EGTRRA greatly expanded the potential for saving through IRAs by allowing catch-up contributions for those over age 50, raising the annual limit on contributions from \$2,000 in 2001 to \$5,000 by 2008, and indexing that limit to inflation thereafter.

Congress has appropriated increased resources to several Federal agencies to promote retirement saving as well as general financial education. These educational programs should be better coordinated to leverage best practices and resources aimed at communicating the importance of savings, both individually and through employer-sponsored retirement plans. Furthermore, the Federal Government must remain a committed partner with the private sector, both for-profit and nonprofit, to educate Americans about the need and opportunities to save.

Other features of the tax code might also encourage saving for retirement by relieving some of the burden of the income tax system. As one example, medical savings accounts may be a useful mechanism for some people wishing to save in anticipation of possibly large out-of-pocket medical expenses related to old age.

Box 2-4. The Effectiveness of Saving Incentives

How effective are targeted saving incentives such as IRAs and 401(k)s at increasing saving? The answer depends, first, on how much “new” saving these incentives generate, and second, on the cost of achieving that saving, in terms of tax revenue forgone.

The first question can be addressed by considering two possible extremes. One is that all saving in IRAs, for example, is new saving—saving that would not have happened were it not for the tax incentives associated with saving in an IRA. At the other extreme, it could be that all saving in IRAs is saving that would have happened even without the incentive. The question then becomes where, between these two extremes, the actual fraction of new saving lies. This question is widely debated, but estimates suggest that 26 cents of every dollar in IRA contributions represents new saving.

Whatever the amount of new saving is determined to be, is it worth the cost in terms of forgone tax revenue? A useful measure for answering that question is the amount of new saving per dollar of revenue cost. Estimates of this measure have indicated that IRAs need not generate considerable new saving per dollar of lost revenue to generate increases in the capital stock that are “inexpensive” relative to the initial revenue loss. This cost-effectiveness of IRAs results because contributions to IRAs lead to a larger capital stock and faster growth. This faster growth translates into higher corporate revenue and, thus, higher tax revenue that more than makes up for the forgone tax revenue associated with IRA contributions.

Fostering Self-Reliance

The key principle underlying all of America's retirement security institutions should be individual self-reliance in planning for retirement. Personal Social Security accounts, private pension plans, and vehicles for individual saving all aim to encourage and support individuals' efforts to prepare for their own financial future. Pension plans and saving vehicles allow individuals to save for retirement on a tax-preferred basis by reducing obstacles to saving inherent in the income tax system.

In a Social Security system with personal accounts, participants will take a more active role in exercising direct control over their retirement wealth, as participants in defined-contribution pension plans and IRAs already do. Lower income individuals will find in personal accounts a mechanism by which they can play a larger role in their own financial destiny. Meanwhile the defined-benefit element of Social Security will continue to provide a foundation of retirement income for those for whom lower resources represent an obstacle to complete self-reliance in retirement planning.

Meeting the Challenge of Retirement Security

The major challenge facing America's retirement security institutions in the 21st century is how to enable a relatively smaller work force to support a growing elderly population. To meet that challenge, we must fortify all three legs of the retirement stool: individual saving, employer-provided pensions, and Social Security. Today the task at hand is to strengthen each of these institutions to serve our needs tomorrow by encouraging public policy that focuses on individual self-reliance in retirement planning.

Social Security is the retirement institution most urgently in need of rebuilding. Simply put, the system will not take in enough in payroll taxes over the coming years to pay the scheduled level of benefits to retirees. Correcting this problem will require some combination of increasing resources to Social Security and slowing the growth rate of outlays. However, this difficult situation also offers an opportunity to build for the future. Restructuring the current system to include personal accounts could improve Social Security's fiscal situation while giving workers a sense of ownership, an element of choice, and the opportunity to leave something to their heirs. Personal accounts could also increase national saving, helping to grow the economy and support a relatively larger elderly population.

A Social Security system made sustainable is just one component of a complete foundation for retirement security. Personal saving, undertaken both independently and through employer-sponsored pension plans, is also essential for ensuring the financial well-being of future retirees. Employer

pensions have seen considerable growth over the past two decades and should continue to grow. Individual saving outside of these plans, on the other hand, has lagged recently. Tax policy should follow the lead of EGTRRA and continue to develop in ways that encourage, rather than punish, these forms of saving.

Meeting the needs of a growing retired population with a relatively smaller work force is a new challenge for the United States, but it is not by any means an insurmountable one. What lies ahead is clear. What we must do to prepare is also clear. We must reinforce our existing retirement security institutions and use them to begin raising national saving right away. These steps will pave the way for a secure retirement for Americans and a prosperous future for the whole country.

Realizing Gains from Competition

The organization of the firms that contribute to our Nation's economic output is constantly in flux. Some changes in organization are limited to a firm's internal operations, as when firms develop innovative ways to produce an existing good or service, or introduce incentives that encourage workers to be more efficient. Other organizational changes involve changing a firm's size or scope. This might include expanding production or offering new goods or services, to gain a greater share of a market or to broaden the firm's geographic reach. Finally, firms may alter their relationships with other firms that supply them, buy from them, or compete with them. For instance, they might merge to combine operations with a former rival, or outsource some part of their operations to another firm.

Some of these changes may be quite visible to consumers. They may change the names of companies with which consumers have become familiar. They may even affect the types of products available in the market. Other changes may be less visible.

At the same time, the overall composition of the economy is also undergoing constant change. In particular, high-technology industries such as biotechnology and information technology have become a much more prominent part of the economy than they were even a decade ago. Innovations are central to the success of the firms that make up these industries. These innovations have brought us remarkably more powerful computers, more effective drug therapies, and much else.

One might naturally ask what the Federal Government's role in the economy should be in light of these ongoing changes in the organization of firms and the composition of the economy. The vast majority of firms face healthy competition from other firms. A great virtue of this competition is that it yields a number of benefits for consumers without the need for government to intervene in the day-to-day decisions of firms. First, competition keeps prices low. Competition in its various forms discourages any one firm from raising prices above what others would charge for similar goods or services. Second, competition ensures that only those firms that can meet consumer demands at the lowest possible cost will remain viable. Finally, competition encourages innovation in products and services, as well as in production and distribution methods, among other things.

Many of the organizational adjustments that firms undertake are necessary responses to changing conditions, as competition motivates them to

constantly seek ways to lower their costs and improve their products. But in some limited cases these changes in organization may have the effect of reducing the vigor of competition. Recognizing this possibility, since the end of the 19th century all three branches of the Federal Government have contributed to the development of antitrust policy, a particularly important component of competition policy.

Three laws passed by Congress form the statutory basis of antitrust policy in the United States. Together, the Sherman Act of 1890, the Clayton Act of 1914, and the Federal Trade Commission Act of 1914 set forth broad principles forbidding behavior or changes in the organization and relationships of firms that may harm competition. The specific implications of these laws have evolved as Federal courts have interpreted their broad principles in deciding cases brought before them. Two Federal agencies, the Department of Justice and the Federal Trade Commission (FTC), actively enforce these laws. Under the Sherman and Clayton Acts, private individuals and firms may also bring suit against firms they believe are engaged in anticompetitive practices. As the courts consider each new case, they are given an opportunity to further refine their interpretation of these antitrust laws.

Competition policy seeks to prevent behavior and changes in the organization and relationships of firms that may harm competition and therefore consumers. But the fundamental challenge in developing competition policy is to ensure that government measures intended to accomplish this goal do not inadvertently prevent the other, more beneficial behavior and changes that firms undertake. To do so would handicap the ability of firms to lower their costs, improve their products, and thereby benefit consumers and society generally.

This chapter examines the various motivations for changes in the organization of firms, and the resulting implications for competition policy. It begins by focusing on what motivates a firm to combine its assets with those of other firms or to take a financial interest in them. Taking as a starting point the progress that has been made in policies relating to mergers, the chapter then discusses how economic ideas and analysis have been and can continue to be incorporated in the ongoing refinement of competition policy. Next, in view of the increasingly global markets in which firms compete, the chapter addresses how the international nature of competition and of some firms' operations can affect both the motivations for changes in their organization and the impact of other nations' competition policies on our economy. Finally, the chapter addresses the implications for competition policy of the increasingly prominent role of innovation-intensive industries in the economy.

The longstanding core principles of U.S. competition policy remain sound. But competition policy continues to evolve to recognize changes in

modern firm structures, market competition, dynamic forms of competition, and advances in our knowledge of the effects of firm behavior. This evolution is proceeding along several fronts. First, because firms today are engaging not only in mergers, but also in hybrid organizational forms such as partial acquisitions and joint ventures, policy must be sensitive to the efficiency gains these forms of organization create. Second, because firms' activities, and therefore national competition policies, more frequently cross international borders than in the past, inefficient competition policies in any one nation may impose costs on firms and consumers worldwide. The United States is pursuing harmonization of these policies in a way that will spread best-practice and efficient competition policy to all countries. Finally, industries characterized by active innovation and dynamic competition are raising new issues for competition policy, which must respond in ways that foster this innovative activity and maximize the resulting benefits to society.

Motivations for Organizational Change

Firms may change their organization for any of a number of reasons. One of the fundamental forces driving the behavior of firms is the desire to maximize their profits. This leads firms to strive constantly to minimize the costs and maximize the value of the goods and services they produce.

Meanwhile developments in individual markets and in the broader economy are constantly changing the costs associated with each of the various ways that firms can choose to organize their operations. These developments may also alter the business opportunities they face, perhaps opening new markets or affecting the competition they encounter. In the past two decades, some of the most significant of these developments have been improvements in the power and reductions in the costs of information technology; deregulation of certain industries; and the globalization of markets. These or other developments may make it profitable for firms to alter their organization or operations.

The work of Nobel Prize-winning economist Ronald Coase provides a framework for understanding how and why firms might restructure their organizations in response to developments such as these. Coase views a firm's operations, internal and external, as a set of transactions, whether it be obtaining materials for production or arranging for the promotion of the firm's products. To maximize its profits, the firm will seek to minimize the cost of each of these transactions. These costs are influenced in part by whether the transaction is performed within the firm or with another party on the open market. The relative costs of these two options will largely determine which one the firm will choose. When developments in its markets or

in the broader economy change these relative costs, the firm will review these options and may decide to change an internal transaction to an external one, or vice versa. The result is a change in its organizational structure. For instance, a firm may perceive an opportunity to outsource some of its inventory management to another firm that specializes in that task. But if this task needs to be closely integrated with other operations in the firm, outsourcing may become preferable only when communications costs fall below some threshold. In this chapter we address the fact that firms today face more than just two alternatives in choosing how to organize their operations. We highlight some of the alternatives that constitute particularly important developments in the organization of firms and industries for the future.

The Role of Agency Costs in Organizational Change

Agency costs are an important component of costs that a firm can lower by adjusting its organizational structure. They can arise whenever one person or firm (the agent) contracts to perform certain tasks for another (the principal). Differing incentives facing the two parties, coupled with the inability of the principal to costlessly monitor the agent's actions, cause the latter to perform the contracted tasks in a way that does not best serve the principal's interest. Ultimately, a firm's owners (in the case of a corporation, its shareholders) are those most interested in maximizing its profits. Not only are they the residual claimants on the firm's profits, but the value of their shares is affected by expectations of those profits today and in the future. Yet there are many others, both within and outside the firm, whose actions affect the firm's profits but who do not benefit enough from an increase in those profits to make maximizing them their only objective.

For example, the decisions of a firm's chief executive officer (CEO) can clearly have a significant effect on the firm's profits. Although the CEO may be interested in maximizing those profits, he or she may also have other, conflicting objectives: perhaps the CEO would like to increase his or her perquisites by purchasing a company jet, even though that would not be an efficient allocation of the firm's resources. Because the CEO runs the firm's day-to-day operations, the CEO is an agent of the firm's shareholders, and the cost associated with the CEO's pursuit of interests aside from profit maximization is an agency cost. This cost arises from the separation of ownership of the firm from control of it.

Just as they may choose to outsource an operation in order to minimize costs, so, too, may shareholders alter the organization of their firm in order to reduce these agency costs. Certain internal institutional arrangements can serve to better align owner and manager incentives. For publicly traded corporations, a commonly used compensation package for CEOs and other senior managers consists of "pay for performance": executive pay is

determined in part by bonuses based on sales or profits, often coupled with the grant of stock options. When managers own stock or stock options in the company they manage, their interests become more aligned with the shareholders' interests. One study found that, with the recent dramatic increases in such forms of compensation, the average effect of a change in the value of a firm on its CEO's wealth grew by almost a factor of 10 between 1980 and 1998. Clearly, pay for performance has become an increasingly prominent feature of corporate life, suggesting that it may prove a valuable way for shareholders to reduce agency costs.

In addition to the CEO, many other individuals and entities influence a firm's profits, and so a comprehensive definition of agency costs must include costs due to their actions as well. Therefore changes in the organization of firms designed to reduce agency costs may extend well beyond arrangements for compensating managers. For instance, if the actions of a particular supplier can significantly affect a firm's profits, the firm may seek to arrange its relationship with that supplier in a way that aligns the supplier's interests more closely with those of the firm's shareholders. Much as in the case of pay for performance contracts, this may be achieved by having the supplier hold stock in the firm.

Mergers

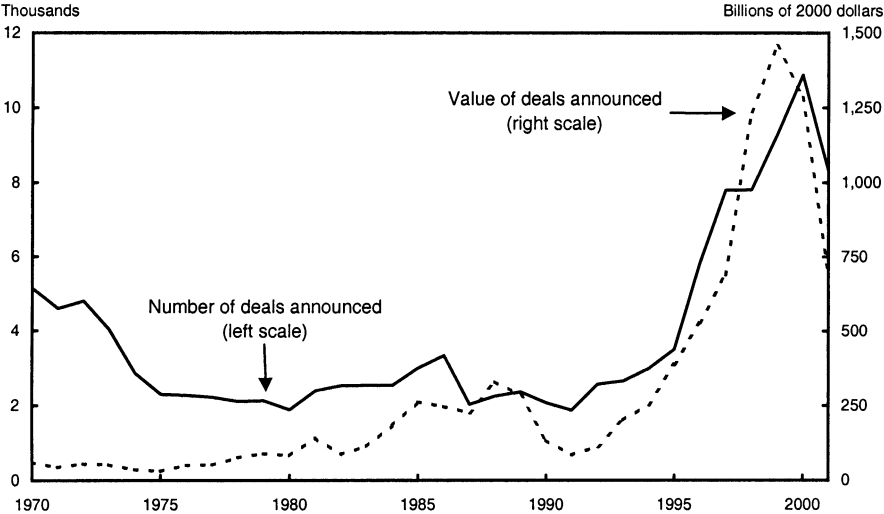
One of the most visible manifestations of changes in the organization of firms is the growing number and value of mergers and acquisitions. During the second half of the 1990s the United States witnessed a remarkable surge in merger activity (Chart 3-1). Indeed, even with the economic slowdown, merger activity in 2001 was well above average levels during the past three decades.

In a significant share of mergers today, one or both parties are firms with operations in more than one country, and many mergers even involve firms with headquarters in different countries. These are often referred to as cross-border mergers. In 2001, 29 percent of all announced mergers and acquisitions in which a U.S.-headquartered firm was a party also involved either a foreign buyer or a foreign seller. This was a markedly higher percentage than was common during much of the 1970s and 1980s (Chart 3-2).

Although general economic theory and empirical research provide a broad framework within which to understand organizational changes across firm boundaries, such as mergers, a substantial body of research has developed that specifically examines the motivations for mergers. The motivations behind each merger are, of course, unique. But some mergers may share certain motivations, and motivations may generally differ across the three broad types of mergers: horizontal, vertical, and conglomerate. Horizontal mergers involve a joining of firms that compete in the same market; vertical

Chart 3-1 Announced Mergers and Acquisitions Involving U.S.-Headquartered Firms

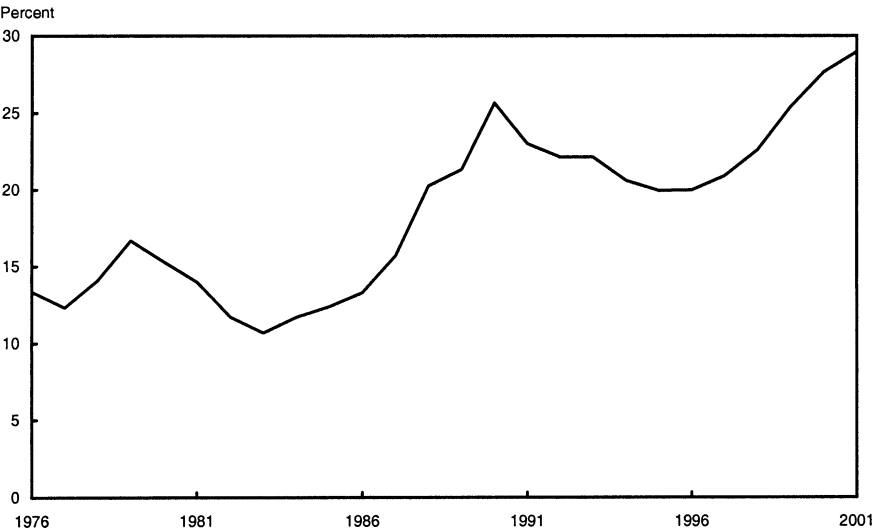
Although mergers and acquisitions increased in number and value during the 1980s, activity since the mid-1990s has far surpassed this earlier wave.



Note: Includes publicly announced mergers and acquisitions of \$1 million and greater that involve at least 10 percent of the target firm's equity. Value is the base equity price offered.
Sources: Department of Commerce (Bureau of Economic Analysis), Mergerstat, and Council of Economic Advisers.

Chart 3-2 Fraction of U.S. Mergers and Acquisitions Involving a Foreign Buyer or Seller

Compared with most of the 1980s, recent years have witnessed a greater proportion of cross-border mergers.



Note: Includes publicly announced mergers and acquisitions of \$1 million and greater that involve at least 10 percent of the target firm's equity. Foreign is defined here as having headquarters outside the United States.
Source: Mergerstat.

mergers occur when a customer buys a supplier, or vice versa; and conglomerate mergers join firms in different businesses. The international nature of cross-border mergers adds another set of potential motivations.

One motivation for mergers is efficiency gains. Two firms may consummate a merger because they expect that the assets of the two firms can be used more efficiently in combination than separately. This might be achieved if merging allows them to lower their costs, improve their products, or expand their operations more effectively than they could as separate entities.

In some cases these efficiencies can be realized through cost savings arising from the increased size of the merged entity, often referred to as economies of scale or scope. This may result from consolidating and spreading certain fixed overhead costs across the combined operations. For instance, economies of scale appeared to be a factor motivating mergers and acquisitions in the food retailing industry during the late 1990s. When two supermarket chains merge, distribution centers made redundant by the merger can be eliminated, and the costs of the remaining distribution centers can be spread over a larger number of supermarkets.

In a horizontal merger, efficiencies might also come from combining the best elements of each firm's operations. One motivation for vertical mergers may be that certain transactions between a supplier and a customer are particularly difficult to arrange between independent firms and can be more efficiently arranged if both parties are part of the same firm. Vertical mergers may also be an efficient method of removing pricing distortions that arise when firms transact with one another in the chain of production, each adding its margin along the way. Elimination of these so-called double margins leads to lower final product prices.

Reduction of agency costs, discussed above, can be another significant source of efficiencies. If a corporation's executives are unwilling to make or incapable of making decisions to increase shareholders' profits, they may be replaced in a merger or acquisition. Or if the firm has assets that a new set of managers could put to higher value use, the firm may be acquired and new, better managers introduced. In some cases, the existing management team may be underperforming because the incentives it faces may be inadequate for it to act in the shareholders' interest, or may even promote behavior that runs counter to their interest. The acquisition or merger of such a firm provides a valuable opportunity for new owners not only to replace management, but also to change the firm's governance structure in order to fix these inadequate or perverse incentives.

Although merger and acquisition activity may sometimes be a response to agency problems, in some settings it may actually be a manifestation of such problems. Some acquisitions may be motivated by a manager's ambition to increase the size of the firm under his or her control, even though the

acquisition is likely to reduce the shareholders' profits. But research also suggests that such poor acquisitions can increase the likelihood that the acquirer itself will become a target for acquisition.

Cross-border mergers can enjoy efficiencies similar to those described above, but the international nature of these transactions introduces another set of potential efficiency gains as well. Just as the opening of world markets to international trade raises productivity, so, too, might a cross-border merger create benefits that no purely domestic reorganization could achieve. These might result, for example, from overcoming barriers to trade that hinder a firm from exporting to another country but not from acquiring production facilities and producing the same goods there. Other efficiency gains from cross-border mergers might come from gaining a better understanding of customers in a foreign market, or from a company with good products acquiring a company with good foreign distribution channels. Alternatively, efficiencies may arise from differences in wages between countries that make it more profitable for firms to locate their labor-intensive operations in countries with abundant unskilled labor, while locating other operations, such as research and management, in countries where skilled labor is relatively plentiful.

Of course, some of these gains may not require mergers, but can be realized simply by establishing new operations overseas. But in some cases, merging with an established firm may be more efficient. Two advantages that mergers can provide are quicker entry into new markets and access to existing proprietary resources and capabilities, such as established brands. A further benefit that a merger or joint venture may provide is the transfer of managerial or technological know-how across national and firm boundaries. The transfer of innovative manufacturing systems may be best achieved through some form of integration. This is discussed in greater depth later in the chapter in the context of the General Motors-Toyota joint venture.

As described above, firms constantly look for potential efficiencies from possible mergers in order to enhance their profitability in a competitive market. Mergers with these motivations have the potential to provide consumers with less expensive and better products or services. But some mergers may reduce competition. This can happen if a merger of competitors allows the merged firm or a collection of remaining firms to raise the prices of the goods or services they sell, or lower the prices they pay for the goods or services they buy from suppliers. In the case of a vertical merger, a firm may be able to reduce the competition it faces by gaining control of either an important supplier to its industry or a significant customer. As in virtually all transactions that come under antitrust scrutiny, this potential to reduce competition may be either a deliberate motivation for, or an inadvertent consequence of, the merger.

Higher prices to consumers as a result of reduced competition are due to what economists call monopoly power, that is, the power of a single seller to affect the market price. Lower prices to input suppliers as a result of reduced competition are due to what economists call monopsony power, that is, the power of a single buyer to affect the market price. Both effects are exercises of market power, and thus a concern of competition policy. Government has a role in preventing those mergers whose adverse effects on competition exceed any benefit from accompanying efficiency gains. The evolving way in which the Federal Government performs this role through its competition policy will be described in more depth later in the chapter.

Other Organizational Forms: Joint Ventures and Partial Equity Stakes

The various possible sources of increased efficiency from mergers, including those that reduce agency costs, can also motivate other forms of organizational change that do not involve complete transfer of both ownership and control. The distribution of ownership and control across parties to an organizational structure affects the parties' incentives and opportunities, their ensuing decisions, and therefore the creation of social value.

Joint Ventures

A joint venture is a business entity created and jointly controlled by two or more separate firms, each of which makes a substantial contribution to the enterprise. Firms may seek to enter a joint venture for any of a number of reasons. Joint ventures may allow firms to combine their complementary skills or assets in a way that improves their ability to accomplish a project. Such a venture may also allow the participants to expand the scale of a project to a size necessary to realize certain cost savings. By avoiding additional costs associated with a full merger, a joint venture may best accomplish the firms' objectives.

One specific type of joint venture, the research joint venture, has its own particular advantages. A joint venture to undertake scientific, technical, or other research may appropriately reward innovation and spread development costs in a setting where the resulting new knowledge, if created by a single firm, would spill over to benefit others. Since in that case no single firm would reap all the benefits of its research, a joint venture may be the most efficient avenue for undertaking it.

But joint ventures might also raise concerns. For example, a production joint venture between horizontal competitors might reduce their ability or incentive to compete independently. Conceivably the participants could

contribute all their manufacturing assets to the joint venture, and their financial stakes in the joint venture could then lead to a reduction in output by the two firms comparable to that in an anticompetitive merger. Even if the joint venture participants retain independent production assets, the joint venture may create the environment for the exchange of competitively sensitive information on prices and costs. This might facilitate an attempt by the firms to raise prices in an anticompetitive manner.

Partial Equity Stakes

A merger or complete acquisition occurs when the ownership of the assets of two firms is combined, for example through one firm's acquisition of 100 percent of the shares of the other, or when two firms exchange all of their shares for those of a new, successor corporation. In contrast, a partial acquisition occurs when one firm takes a partial equity stake in another firm, which remains legally independent.

Partial equity acquisitions, like merger transactions, must be reported to the Department of Justice and the FTC under the 1976 Hart-Scott-Rodino Act if the transaction meets certain conditions. In fiscal 2000, 23 percent of all transactions reported to the two agencies resulted in the acquirer having less than a 50 percent share of the target firm's equity. Although these may be supplemented by later purchases, it suggests that partial purchases are not uncommon.

Partial acquisitions create a form of corporate governance that raises some basic questions about the "ownership" and "control" of one party over another. Partial equity investments by one firm in another can grant the investing firm substantial influence over the other firm. A majority shareholder can be presumed to exercise control, although under some constraints imposed by the duty toward minority shareholders. But research suggests that even ownership of far less than a majority of a company's shares may allow the exercise of control, if the remaining shares are widely dispersed.

PepsiCo, Inc.'s investment in the Pepsi Bottling Group, Inc., is an example of a partial equity stake that involves some control. The Pepsi Bottling Group is the world's largest manufacturer, seller, and distributor of Pepsi-Cola beverages. It has the exclusive right to manufacture, sell, and distribute these beverages in much of the United States and Canada, as well as in Spain, Greece, and Russia. PepsiCo holds the licenses for Pepsi-Cola beverages and is a minority shareholder, although also the largest shareholder, in the Pepsi Bottling Group. There is close coordination between the two businesses, but each remains a legally independent entity whose interests are not legally presumed to align with the other's.

At the other extreme, an individual who buys a few shares in a public company may do so as an investment for retirement or for other purposes.

These small purchases best exemplify so-called passive investments, in that the shareholder has no current plans to gain influence over the firm's conduct or to access certain information about its operations, and there is no good reason to expect such plans to emerge in the future. Likewise, one firm may purchase a small equity stake in another firm without such plans or any realistic potential for such plans to emerge.

A partial acquisition can affect the firms' subsequent decisions through three distinct channels: by altering incentives, altering information, or altering control. Through these channels, an acquisition could have anticompetitive or pro-competitive effects. The potential anticompetitive effects are considered first, because without those effects there is no concern for antitrust policy.

Even if a firm has only a passive investment in another firm, this might, through altering incentives, affect the former's production and pricing decisions. For example, if firm A owns a 5 percent stake in firm B, it will make production and pricing decisions to maximize its own profits plus 5 percent of firm B's profits. The acquirer of a partial equity stake will consequently internalize some of the spillover effects of its actions on the target's profits. This is true whether or not the acquirer can exercise control over the target.

Such a passive investment could have an anticompetitive effect in an imperfectly competitive market if the two firms are direct competitors. If firm A raises its price, for example, the 5 percent stake in firm B could reduce the effect of any loss of customers on firm A's profits because some of the lost customers would begin purchasing from firm B. Firm A would capture part of firm B's increased profits, reducing its overall losses from raising prices. This diminishes firm A's incentives to keep prices at a competitive level. Nonetheless, this concern should not arise if other firms in the market are able to expand their output and win most of the customers that firm A loses when it raises its prices. Thus competition guards against the rise in prices.

The information effect arises from closer unilateral or bilateral communication between the partial acquirer and the target about business operations. For example, if the partial acquirer receives a seat on the target's board of directors, that may become an avenue for improved communication between the firms. This improved communication could facilitate anticompetitive conduct, for example if two competitors attempted to coordinate a rise in prices.

Finally, a partial acquirer may be able to influence the target's business decisions through the control effect. This could have anticompetitive consequences if the two firms are competitors. For example, the acquirer might raise its price and exert its influence so that the target responds by increasing its own price. But these effects can also be prevented if other firms in the market expand their output in response to higher prices.

Partial acquisitions may have socially desirable consequences, operating through these same channels. In particular, partial equity stakes may be

undertaken as part of a larger business relationship, such as a marketing or supply agreement. Such partial equity stakes may align incentives, internalizing spillovers in ways that are socially beneficial. These business relationships may also be cemented by the information and control benefits facilitated by a partial equity stake.

One study examined 402 partial ownership stakes established between 1980 and 1991 in which a nonfinancial corporation held a minimum of 5 percent of the outstanding shares of another firm. Thirty-seven percent of the target firms had explicit business relationships with the corporation holding their shares.

More recent, although preliminary, data suggest that about 5 percent of Fortune 500 nonfinancial companies in 2001 had a corporate blockholder of 5 percent or more of their shares in that year. (This sample examines the Fortune 500 companies, excluding those in finance, insurance, real estate, or retail trade. Companies in which there was a majority shareholder were also excluded.) In this preliminary research, corporate blockholders appear to be more prevalent in certain industries than others. In the rapidly evolving telecommunications sector, for example, about a third of major U.S. corporations had at least one corporate blockholder in 2001.

An example of how partial equity stakes may align the incentives between parties in a business relationship is the 1997 co-production agreement between Walt Disney Company and Pixar. At the time of their co-production agreement, Disney acquired about a 5 percent stake in Pixar. This example is described in Box 3-1.

The potential for a partial equity stake to encourage efficiency gains in the long-term relationship between a supplier and a customer highlights an advantage of this form of organization. In a long-term supply relationship, both customer and supplier may make relationship-specific investments, such as fabricating machine tools to produce a part according to the buyer's specifications. If the buyer's input needs change unexpectedly, it may want rapid delivery of a modified input from its supplier. If the supplier has an equity stake in the customer, and hence a claim to some of the customer's profits, the supplier may have a stronger incentive to meet the customer's request, even if it must incur overtime costs to adjust its machine tools. If the partial equity stake allows one firm to exercise some control over the other firm, the coordination between their operations is likely to be further strengthened.

Box 3-1. A Co-Production Agreement and a Partial Equity Stake: Pixar and Disney

Pixar was formed in 1986. Its first fully computer-animated feature film, “Toy Story,” was released in 1995, also the year of the company’s initial public offering of shares. “Toy Story” was distributed by the Walt Disney Company, under a contract in which Disney also bore all the budgeted production costs. In return, it received a standard distribution fee from Pixar and the vast majority of the film’s revenue, including about 95 percent of box office receipts during the year after its release.

In 1997 Disney and Pixar entered into a co-production agreement to produce and distribute five new computer-animated feature films. Under the agreement, Pixar would produce the films, on an exclusive basis, for distribution by Disney. Disney and Pixar would split production costs and all related receipts in excess of the amount necessary to cover Disney’s distribution costs and an associated distribution fee. The films would also be co-branded.

This agreement was cemented by Disney’s acquisition of a partial equity stake in Pixar. Disney initially acquired 1 million of Pixar’s shares and received warrants to purchase up to an additional 1.5 million shares. At the time, exercising all these warrants would have given Disney about a 5 percent stake in Pixar.

The Pixar-Disney co-production arrangement brought “A Bug’s Life” to the big screen in 1998, and “Monsters, Inc.” in 2001. The alliance benefits both companies and exploits a logical division of labor between the firms. As Pixar’s 2000 10-K filing states, “This agreement allows [Pixar] to focus on the production and creative development of the films while utilizing Disney’s marketing expertise and substantial distribution infrastructure to market and distribute our co-branded feature films and related products.”

An interesting wrinkle is that Disney is not only a partner with Pixar but also a competitor. Pixar notes in its 2000 10-K filing that, under the agreement, Disney directly shares in the profits from their co-branded films, and therefore Pixar believes “that Disney desires such films to be successful.” But the filing also points out that, “Nonetheless, during its long history, Disney has been a very successful producer and distributor of its own animated feature films.”

Thus, although the profit-sharing terms of the agreement give Disney powerful incentives to use its marketing and distribution acumen to further the success of the co-branded films, the partial equity stake plays a complementary role. Through this investment, Disney shares directly in Pixar’s success, and so has additional reasons to foster the collaboration.

Incorporating Economic Insights into Competition Policy

Economists have long studied the implications of changes in the structure and conduct of firms, creating a body of knowledge that encompasses the insights described above. Developments in this body of knowledge provide an important basis for improving the effectiveness of competition policy.

The evolution of U.S. policy relating to horizontal mergers—those between companies that compete for customers in the same market—provides one example of how economic thought has substantially enhanced competition policy in the past two decades. As explained above, a merger between such companies can bring about benefits through reductions in the cost and improvements in the quality of the merging firms' products. But some such mergers have the potential to harm competition. In determining whether to challenge a particular merger, the Department of Justice or the FTC must assess whether the merger threatens to harm competition, and whether the potential benefits of increased efficiencies outweigh any adverse effect the merger could have on competition. To do so, the agencies have developed an analytical framework that allows them to move from a set of observable characteristics of the merging firms and the markets in which they compete to an assessment of the likely competitive effect of the transaction, balanced against any efficiency benefits.

The analytical framework used is important in that it influences the types of characteristics considered in evaluating mergers and related acquisitions, whether the enforcement agencies challenge them, and how they are ultimately viewed by the courts. This framework provides a focus for arguments about the merits of or problems associated with a merger. Finally, an analytical framework that is consistently adhered to increases firms' ability to assess whether a merger they are considering will be challenged, before they embark on the costly process of initiating it.

It is in contributing to the improvement of this analytical framework that developments in economic thought have significantly affected merger policy. This effect is visible in the evolution of the Horizontal Merger Guidelines, a description of this framework that was first established by the Department of Justice in 1968 and periodically revised since then by both the Justice Department and the FTC. Although the need for flexibility in enforcing antitrust law causes these guidelines to be somewhat general in nature, the trend toward an increasing incorporation of a rigorous economic framework is nonetheless still apparent in the periodic revisions to the guidelines. Because the ability to gain the favorable ruling of a judge in an antitrust case affects these agencies' ability to successfully challenge mergers, changes in the

guidelines also to some extent reflect accompanying changes in the judicial interpretation of antitrust law.

Of the various revisions made during the past two decades, the 1982 guidelines and the revisions made to them in 1984 together marked the most dramatic departure from prior guidelines in their incorporation of contemporary economic thought. One significant advance in these revisions was a shift away from a singular focus on market concentration in assessing the effect of a merger. Market concentration is a measure of the extent to which the supply of products and services in a particular market is concentrated among few providers. The earlier focus was consistent with economic thinking, developed in the middle decades of the twentieth century, according to which increases in the concentration of markets harmed competition. As a result, in the 1960s, mergers that raised concentration by increasing a firm's market share to even as little as 5 percent were at risk of being challenged.

The 1982 and 1984 revisions reflected an evolving economic perspective on the effect of concentration on competition in a market. This perspective had been increasingly gaining judicial recognition by the mid-1970s. Theoretical and empirical work had begun to call into question the idea that there is a simple link between a market's concentration and the intensity of competition in that market. By 1982, judicial decisions and enforcement policies had already begun to incorporate the conclusion from economic research that, although high concentration could contribute to reduced competition, by itself it was not sufficient to bring about that outcome. Thus the 1982 and 1984 revisions codified the increasingly accepted view that examining market concentration provides only a useful first step in considering whether a merger raises competitive concerns, and that other factors needed to be present to validate this concern. In line with this view, the revisions described quantitative levels of market concentration and changes therein that would likely cause the Justice Department and the FTC to go on to examine the full set of factors and possibly challenge a merger. The 1984 guidelines also clearly established a level of market concentration below which, "except in extraordinary circumstances," mergers would not be challenged. This "safe harbor" level of market concentration is important in that it reduces the uncertainty that firms considering a merger may have about how the government will respond. Such a clear safe harbor was absent in the 1968 guidelines.

One of the additional factors that the 1980s revisions incorporated as an important consideration in evaluating the intensity of competition in a market was the ease with which new firms could enter that market. Although existing firms in a market are the most visible source of competition for each other, they are not the only source. In considering whether it would be

profitable to raise prices above existing levels, a firm or group of firms must not only consider the response of firms already in the market. They must also consider the possibility that higher prices will encourage other firms to enter the market, adding to competition. Thus, in some cases, even if there are few firms in a market today, the threat of new firms entering tomorrow can provide a strong incentive for incumbent firms to keep prices competitive. In an improvement on the earlier merger guidelines, the 1980s guidelines recognized that a merger could only harm competition if there were reasons to believe that other firms would not or could not enter the market to the extent necessary to keep the merging firms from maintaining prices above premerger levels.

Another substantial advance in the 1984 guidelines, and improved upon since then, was a greater recognition of potential efficiency gains from mergers. Today it is widely accepted among economists that mergers should be evaluated in terms of a tradeoff between any potential adverse impact on competition and their potential enhancement of competition by improving the merging firms' operations. The 1968 guidelines had focused attention almost exclusively on whether a merger could harm competition, with little consideration given to the potential benefits, because these were considered hard to evaluate and often realizable by other means. In contrast, the 1984 guidelines recognized that mergers that might otherwise be challenged may nonetheless be "reasonably necessary to achieve significant net efficiencies." The guidelines set forth a number of types of efficiency improvements that could be considered in assessing the impact of a merger, such as economies of scale. Moreover, the tradeoff often presented by mergers was explicitly recognized in the 1984 guidelines, which state that "a greater level of expected net efficiencies [is needed] the more significant are the competitive risks identified." Improvements in the consideration of these efficiencies, and in other elements of the analytical framework applied to evaluating mergers, continued in later revisions.

Competition Policy, Corporate Governance, and the Mergers of the 1980s and 1990s

In the years leading up to 1982, some elements of the new thinking that would later appear in the revisions to the Horizontal Merger Guidelines had already begun to be incorporated in the Justice Department's and the FTC's enforcement practices, and in the interpretation of antitrust laws by the courts. Nonetheless, the revisions were important in codifying this dramatic adjustment in antitrust policy, which allowed firms greater flexibility during the substantial restructuring of the economy that occurred in the 1980s. In contrast, during the 1960s and much of the 1970s, in line with the 1968

guidelines, Federal policy and judicial decisions relating to horizontal and vertical mergers had been quite restrictive.

During the 1980s the total value of merger activity picked up considerably. In 1988 the total dollar value of mergers and acquisitions was, in real terms, more than four times greater than it had been a decade earlier. Two types of reorganization were prevalent during this period, both of which might have faced greater opposition under the 1968 guidelines. The first involved the merging of two large firms in the same industry, and the second involved the breakup of a conglomerate, in which individual business lines were often sold to firms competing in the same market as the business line they were acquiring. Although such mergers and acquisitions might still be opposed under the revised guidelines if they presented significant concerns about the effects on competition, the improved economic understanding of competition in markets that was reflected in the revisions caused antitrust enforcement policy to be less restrictive toward such mergers. The trend whereby mergers increasingly involved two firms in the same industry continued in the 1990s.

In the 1980s and 1990s, mergers were clustered in particular industries, although the industries in which they were clustered varied over time. This suggests that mergers may have provided an important means for companies to respond to industry-wide shocks such as deregulation, technological innovations, or supply shocks. Between 1988 and 1997, on average, nearly half of annual merger deal volume was in industries adjusting to changing conditions brought about by deregulation. One study of Massachusetts hospitals shows the effect of technological innovation on merger activity. The study found that new drug therapies and improvements in medical procedures were partly responsible for a significant decline in the number of inpatient days from the early 1980s to the mid-1990s. This reduction in the need for hospital beds contributed to a significant consolidation among hospitals during this period, much of which was facilitated by mergers.

Evidence of stock market reactions to merger announcements during the 1980s and 1990s suggests that, on the whole, they created value for the shareholders of the combined firms. Moreover, studies have found that, in the aggregate, the operating performance of merging firms has improved following the merger. But these aggregate results present evidence of only modest gains, the source of which is unclear.

Yet this is to be expected, because mergers have numerous motivations, and, as with all business decisions in a competitive market, not all will yield the success that is hoped for. As a result, more narrow studies of particular industries, particular types of mergers, and even specific mergers can yield a richer understanding of the sources and extent of gains. For instance, detailed examinations of bank mergers during the 1990s found cases of postmerger performance improvements that likely came from a variety of sources,

including opportunities afforded by the merger to expand service offerings and the efforts of a vigorous management team acquiring a laggard bank. Perhaps indicative of larger trends, however, along with uncovering successes, these examinations also revealed some bank mergers with disappointing results.

The important point for competition policy is that, although the overall efficiency consequences of the mergers of the 1980s and 1990s may be debated, there is little evidence that they harmed competition. Thus it appears that thoughtful and adaptive antitrust policy has afforded businesses greater flexibility to respond to changing economic conditions while preventing such responses from significantly harming competition.

The agencies' improved understanding of the sources of possible competitive harm also helped firms structure or restructure their proposed transactions so as to achieve the efficiencies they sought without raising competitive concerns. For example, a 1998 transaction sought to combine two of the Nation's largest grain distribution and trading businesses. The combination had the potential to lower operating and capital costs but might also have depressed the prices farmers received in certain locations for their grain. The parties agreed to divest certain facilities at certain locations, settling the Department of Justice's challenge to the transaction and allowing the acquisition to proceed. Cases such as this one can be seen as a manifestation of an increasingly thoughtful and adaptive competition policy.

The Role of Corporate Governance Changes

For many of the mergers and takeovers of the 1980s that appeared to create social value, changes in corporate governance and ensuing reductions in agency costs often played an important role. In some cases, takeovers led to the breakup of large conglomerates, forcing apart business units that were presumably more valuable on their own or in other companies' hands. Many incumbent managers resisted these restructurings until forced to accept them through the market for corporate control, as takeovers or the threat thereof often led to changes in the organization of firms.

Although many types of mergers and acquisitions may have led to changes in corporate governance, some of the most dramatic changes therein came about as a result of leveraged buyouts (LBOs). Moreover, evidence suggests that LBOs during the 1980s led to significant improvements in the productivity of firms. In an LBO or a management buyout, corporations become closely held companies as their public stock is bought by a group of investors using borrowed money. Consequently, ownership becomes much more concentrated and more tightly connected to control. This new ownership and capital structure creates significantly greater incentives for managers to increase profits as much as possible. One study showed that CEOs of firms involved in LBOs during the 1980s saw their ownership stake rise by more

than a factor of four, thereby making them more interested in increasing the firm's profits. Moreover, the need to service debt issued to finance the buyout provided a disciplining force on management.

Taken together, it was likely that these incentives influenced decisions by some firms to sell off assets that had higher value outside the firm than inside it. Many LBOs did not raise antitrust issues because the initial transaction simply involved changing the ownership of an existing firm, rather than a combination with a competitor. However, some selloffs of business units that followed certain LBOs were to firms in the unit's industry. Therefore, where these selloffs could improve the performance of the firms without affecting competition, the increased flexibility afforded by adjustments to antitrust policy may have been important.

Once the firm's operations were restructured and a new governance structure was put in place, many LBO firms were successfully taken public again. Although LBO activity dwindled in the 1990s, the expansion of pay for performance suggests that mechanisms to align managerial with shareholder interests remain an important, enduring element of corporate governance.

The restructurings of the 1980s provide an example of the importance of adapting competition policy in response to improvements in the understanding of the conditions within industries that may harm or benefit consumers. The ongoing incorporation of these insights into the analytical framework used to guide competition policy has strengthened the effectiveness of antitrust enforcement, while reducing the likelihood that antitrust enforcement will hinder reorganizations whose economic benefits to society would outweigh any potential harm from reduced competition.

Policy Lessons for Promoting Organizational Efficiencies

As noted earlier, organizational change in today's economy takes place not only through mergers but also through other organizational forms such as joint ventures and partial acquisitions. The challenge for antitrust scholarship and public policy is to provide an integrated framework for all these organizational innovations that properly accounts for both competitive and efficiency effects. These types of transactions evoke intertwined issues in corporate governance and competition policy, and so an integrated framework supports sound policymaking. For example, how a given partial equity acquisition is likely to affect the acquirer's relationship with the target depends on more than just the size of the partial equity interest acquired and the nature of any accompanying shareholder agreement, which may, for example, confer the right to appoint representatives to the firm's board of

directors. It also depends on the acquirer's current and likely future plans, and those of other blockholders and the firm's incumbent managers. Even ascertaining that the acquirer will gain control need not imply that the transaction would be anticompetitive; as in merger policy, that depends upon the market environment and on the efficiencies that the transaction would create.

Policy Lessons from Joint Ventures

Joint ventures can lower the costs of producing goods and services and widen consumer choice. But partners in a joint venture may also be actual or potential competitors in the product market. In 1983, for example, General Motors (GM) Corp. and Toyota Motor Corp. agreed to establish a joint venture to produce a subcompact car at a former GM plant in Fremont, California. This venture was later formalized as New United Motor Manufacturing, Inc. (NUMMI). Both partners expected to benefit from the undertaking: GM by adding to its capabilities in producing smaller cars, Toyota from the opportunity to test its production methods in an American environment. It was an unprecedented initiative and generated an extensive, 15-month FTC investigation, which resulted in its approval.

A new organizational innovation, by definition, will not have an established track record for an antitrust agency to review. But such an organization may create genuine, important efficiencies even if those efficiencies are difficult to document at the time of the transaction. For example, a key issue before the FTC was whether the joint venture would enable Toyota to learn how its “lean” production and assembly system would function in an American factory, and enable GM to learn details of the Toyota system that could be applied to raise productivity at its other plants.

If Toyota's manufacturing success was completely embodied in a superior piece of equipment, then merely licensing that equipment to U.S. automakers might have been sufficient to transfer that success to American soil. That type of efficiency gain also would have been relatively easy to document contemporaneously. Yet, as subsequent scholarship has confirmed, Toyota's lean production system is an interrelated set of practices, affecting factory and job design, labor-management relations, relationships with suppliers, and management of inventories. As the FTC majority opinion concluded, “in depth, daily accumulation of knowledge regarding seemingly minor details is a more important source for increased efficiency than a broad but shallow understanding of Japanese methods. Such in depth knowledge appears to be achieved only through the kind of close relationship the [joint] venture will allow.”

Experience shows that the joint venture did lead to productivity improvements. One study indicated that, within a few years, each automobile produced at the NUMMI plant required 19 assembly hours of labor, versus 31 hours at

one of GM's mass production plants in the United States, and 16 hours at one of Toyota's plants in Japan. The productivity of the NUMMI plant was close to that of Toyota's Japanese plant even though NUMMI workers were relatively early in the learning process about lean production, suggesting that this system could indeed be transplanted successfully. Several other welcome developments followed in the wake of the joint venture's early success. Toyota expanded its own production and assembly plant operations in the United States. GM and other U.S. automakers adopted elements of lean production, improving their productivity. And NUMMI expanded. By 1997 the joint venture had produced its 3-millionth vehicle, and in 2001 the Fremont facility was producing three vehicle models.

The broader policy lesson is that joint ventures and other organizational hybrids may create efficiencies in ways that are difficult to prove at the time of the transaction. In evaluating transactions that might also raise anticompetitive concerns, antitrust authorities face the uncertain prospect of improved efficiency as a factor in evaluating the joint venture's likely effect. A new, potentially efficiency-enhancing organization can benefit society in two ways. Society gains direct benefits from the organization. Society also receives the demonstration of the types of efficiencies that such an organization could create. This provides evidence to other firms, and to the antitrust enforcement agencies, about the private and social gains of such organizations. If the new organization proves efficient, other firms may adopt that form. If it does not prove efficient, market forces will motivate the firms to abandon it. In either case, the antitrust agencies will have a broader track record to rely upon when evaluating similar transactions that might raise competitive questions.

The guidelines describing how U.S. enforcement agencies assess mergers or collaborations such as joint ventures indicate that efficiencies arising from them will be considered if they are verifiable and cannot be practically achieved through other means, making them transaction specific. "Verifiable" here means that the parties must substantiate efficiency claims so that the agencies can verify, by reasonable means, their likelihood and magnitude. In these guidelines, certain efficiency claims are viewed as less likely to meet these criteria than are others. For instance, the agencies view improvements attributed to management as less likely to meet the criteria necessary for consideration. But efficiency gains from mergers or joint ventures may be closely tied with managerial improvements, such as combining Toyota management with unionized American workers in NUMMI. Managerial and organizational improvements may indeed be difficult to verify, but given their potential social value, expending the resources necessary to investigate those claims thoroughly is justified. This policy lesson applies to mergers as well as joint ventures.

Legislation indeed exists to encourage efficient joint ventures. In 1984 the National Cooperative Research Act (NCRA) became law, to be followed 9 years later by the National Cooperative Research and Production Act. These two acts encourage research and production joint ventures by codifying antitrust treatment of such ventures. They lowered the maximum penalty that could be assessed in a successful private antitrust lawsuit against any venture that notified the Justice Department at the time of its formation. For all joint ventures, the act also ensured that, in any antitrust challenge, the courts would consider efficiencies arising from the joint venture. This clarified that defendants could exonerate themselves by establishing the benefits of their joint ventures. Since the passage of the NCRA more than 900 research or production ventures have registered with the Justice Department.

Successful research joint ventures may foster innovation and thus bring benefits to society. This and other ways in which economic organization and competition policy promote innovation are elaborated in the section on dynamic competition later in this chapter.

Shaping Policies to Address Partial Equity Stakes

As we have seen, firms make partial equity investments under a variety of conditions, to achieve a variety of ends. The overall effect can be to promote efficiency or reduce competition, depending on the nature of the acquisition and the conditions under which it is made. Partial acquisitions most dramatically confer control, or influence, over the target company when a majority of its outstanding equity is acquired. Acquirers obtain substantial influence in some instances with much smaller stakes, however. Partial acquisitions also give the acquirer a stake in the target firm's future profits. This gives the acquirer an incentive to take those profits into account when making its own business decisions. Finally, a partial acquisition can make it easier for the acquirer to obtain access to the management of the target firm. All these elements can have substantial effects on the relationship between the target and the acquiring firm. Because strong product market competition can depend on the independence of firm actions, all of these aspects of partial acquisitions can raise serious antitrust enforcement concerns. The challenge in shaping policies to address partial equity ownership by corporations lies in distinguishing cases that pose serious threats to product market competition from those that promote efficient cooperation between suppliers. Although some of these issues are fairly new, the challenge is similar to that posed by the analysis of mergers and, of course, joint ventures.

With the emergence of partial acquisitions among major U.S. corporations, the Justice Department and the FTC have created an enforcement record that publicly illustrates some of the concerns these acquisitions can raise. For example, Primestar was formed in 1990 as a joint venture involving five of

the Nation's largest cable television providers and a satellite provider. In 1997 Primestar announced its intention to acquire satellite assets from two other companies. These assets could be used for direct broadcast satellite (DBS) service, which transmits video programming directly from satellites to subscribers' homes and competes for customers with cable television. The cable companies involved in the original joint venture would have maintained a substantial ownership and control stake in the entity resulting from the proposed acquisition. Since the assets in question were the last available that other independent providers of DBS could use or expand into, Primestar's ownership structure raised concerns at the Justice Department during its review of the acquisition. Concerned that the cable companies would exert their influence in Primestar to limit how the acquired assets would be used in competing with cable, the Justice Department challenged the acquisition, which was subsequently abandoned. The determination that this acquisition would have caused competitive harm hinged upon an assessment of how the new entity's governance structure would affect its behavior (Box 3-2).

As the Primestar case illustrates, the government's evaluation of how partial acquisitions are likely to affect competition requires the examination of conditions under which the parties to the transaction compete, as would be the case in the evaluation of a full merger. Only to the extent that competition between cable and DBS benefits consumers, or society generally, would the Primestar acquisition have been likely to have a serious adverse effect on competition. The partial nature of the cable companies' stake in Primestar thus raised questions in addition to, rather than apart from, those that arise in the traditional evaluation of mergers. Also, as in the evaluation of mergers and joint ventures, the Justice Department and the FTC typically consider the evidence on whether each partial acquisition may promote efficiency.

Some of the tools that economists use to analyze efficiency gains derived from vertical relationships generally may prove useful in the analysis of partial acquisitions between suppliers of complementary products. For example, the influence or control that the acquirer may exercise over the target raises the acquirer's incentive to make certain relationship-specific investments. Relationship-specific investments are those that, once made, are much more valuable inside a particular business relationship than outside it, such as fabrication equipment that is specialized to a particular customer's design. The acquirer's control rights make it less likely that the target will later "hold up" the acquirer, and deprive it of its appropriate return on its investment. These control rights are important because it is costly to go to court to try to enforce a written agreement. If one party effectively controls the other party, disputes over the business arrangement may be resolved at lower cost internally. Although the costs of dispute resolution may be

Box 3-2. The Primestar Acquisition

A basic assumption in assessing the competitive implications of a merger is that the merged firms will act in such a way as to maximize the new entity's profits. A firm's owners, however, may also have other objectives. Usually these other objectives are not significant enough to alter the basic assumption. But when a firm's owners clearly have other interests, such as financial stakes in other ventures, these could influence their decisions regarding the firm's actions. In such cases, those assessing a merger must consider how strong those influences might be on an owner and that owner's ability to affect firm decisions in ways that may harm competition.

Primestar was formed in 1990 as a joint venture involving five of the largest cable television providers and a satellite provider. Given that the five cable providers would control almost 98 percent of the voting shares in Primestar after the proposed acquisition, there were concerns about how this would affect its use of the acquired assets. If Primestar used these new assets to compete vigorously with cable for subscribers in order to maximize its profits, under certain assumptions the effect of lost customers on the profits of some owners' cable businesses might outweigh their share of the gains from Primestar improving its subscriber base. As a result, one might suspect that these owners would seek to influence Primestar's actions to reduce its competition with cable.

On the other hand, Primestar's managers and board of directors would have had legal obligations to serve the interests of minority shareholders that would benefit financially from Primestar competing vigorously with cable television, and the board included independent outside directors. Moreover, it appeared that not all the cable providers would have had an incentive to prevent such competition. Thus the composition of Primestar's ownership and governance structure suggested that there might be opposing forces that would seek different outcomes of decisions affecting competition in the consumer market that DBS serves.

The Justice Department analyzed the totality of incentive and governance effects in this case and concluded, on balance, that the transaction would harm competition and consumers. It filed suit to block the acquisition, leading to its abandonment. This case demonstrates that an assessment of a merger or acquisition's competitive implications can require an understanding of how the governance structure of a company allows those with a share in its control, or a financial stake in its operations, to influence decisions affecting the firm's actions.

lowered through a partial or complete equity interest of one party in the other, there are other costs to this integration, such as “influence costs” as agents seek to lobby decisionmakers within the organization. But market forces will lead firms to choose the arrangement that minimizes their total costs.

Another example derives from the lesson from scholarship that, if one firm acquires another outright, the acquirer’s specific investment incentives are strengthened, but the target’s specific investment incentives are weakened. In the context of a corporate acquisition, this means that stakeholders in the target company care much less how that company’s assets are deployed after selling their stakes. Therefore, if a project can best succeed through such investment by both parties, an optimal ownership arrangement may be one in which one party holds a partial equity stake in, rather than completely owning, the other. This raises the investment incentives of the partial owner while not unduly undermining those of the target.

An important challenge in the development of competition policy toward these new corporate governance practices will be to make effective use of these tools in light of the evidence that has emerged on the antitrust concerns that those practices can raise, and the beneficial effects that can result from them. Some progress will arise through the identification of factors that enforcement authorities will increasingly consider in evaluating partial acquisitions, and that parties will increasingly consider when deciding whether to propose them. Other progress will emerge from a clearer understanding of how these practices affect product markets and economic efficiency more generally. With a clearer sense of the general consequences of these transactions, and of the specific factors that can lead those consequences to vary from case to case, we can expect further advances in the development of tools to evaluate these new governance practices.

Policy Toward Vertical Relations

Some tools for the analysis of these governance practices may derive from a well-developed economics literature on vertical relations between independent firms, a subject in which important issues in firm organization and competition policy arise. Firm activities and market transactions often involve a vertical production and distribution chain, such as a relationship between a manufacturer (called in this situation the upstream firm) and a distributor (the downstream firm).

Antitrust law and its enforcement have a long history of influence over these organizational decisions, such as whether a firm owns the retail outlets for its goods or services. For example, the owner of a business format and brand name for a fast-food restaurant concept may also own individual restaurants, or it may enter into a franchise agreement. A franchise agreement is one between two legally independent firms, the franchisor (the owner of

the business format) and the franchisee (in this example the owner of the individual restaurant). The agreement might specify that the franchisee may operate a restaurant at the given location according to the specified format, in exchange for a franchising fee and a royalty rate on the restaurant's sales.

This organizational choice is, in part, a response to various agency costs. In particular, since a franchisee owns the individual restaurant, he or she has incentives to exert certain types of effort to build up the value of that store. Under company ownership, the manager of the restaurant is an employee and, even if paid a bonus wage based on sales, does not have as strong an incentive as a storeowner to invest effort to raise the value of that store. But franchising may exacerbate other agency costs. For example, the owner-operator of the only restaurant on a busy interstate highway may expect to have many one-time customers, and therefore might charge prices that are too high—a decision that may be profitable for that owner but tarnishes the brand name and lowers its nationwide value. In a company-owned restaurant, the manager has less incentive or ability to act in this manner. The fact that both franchise stores and company-owned stores successfully coexist in our economy reflects differences in agency costs in various industries and settings.

These organizational choices can also be influenced by competition policy, which affects the costs of various possible terms of an agreement between independent upstream and downstream firms, such as a franchise agreement. For example, the upstream firm might wish to specify a maximum retail or “resale” price, which would prevent an individual store from taking advantage of its local market position and potentially harming the reputation of the brand name. As the Supreme Court acknowledged in its 1997 *State Oil v. Khan* decision, there are pro-competitive rationales for such vertical restraints, which is why such a pricing provision is now evaluated for its competitive consequences on a case-by-case basis. Before the Supreme Court's decision, however, an attempt to set a maximum resale price in an agreement between legally independent upstream and downstream firms would have been illegal per se. As a result, owners of a business format who were concerned about the possibility of franchisees pricing too high may have instead chosen to own those restaurants or stores outright. That choice would have addressed the pricing issue but increased other agency costs related to effort by restaurant managers. This example shows one way in which competition policy with regard to vertical restraints nowadays takes into account the social benefits that may be created by having transactions organized between two separate firms rather than through common ownership or vertical integration.

Cross-Border Organizational Changes

Competition policy continues to respond to other changes in the organization of economic activity. The GM-Toyota joint venture, for example, presaged something that has become much more prominent since the venture's establishment: changes in firm organization, including mergers, that occur across national boundaries. This section describes some of the challenges that the international nature of these changes presents for antitrust policy, and how the United States is responding.

Multijurisdictional Review

Merger proposals involving or creating multinational enterprises can result in reviews by the antitrust authorities of many nations, often referred to as multijurisdictional review. The United States has managed the issues posed by multijurisdictional review through both bilateral cooperative relationships and multilateral arrangements. This has produced an impressive degree of analytical convergence among the U.S. and other antitrust agencies, resulting in a long line of compatible decisions in transnational mergers. However, some differences remain, and these can have significant consequences. A striking recent example came with the proposed acquisition by General Electric Company (GE) of Honeywell International Inc. Both GE and Honeywell are U.S.-headquartered corporations, but because these multinational enterprises also have significant European sales, the deal was subject to review by antitrust authorities of the European Union.

GE and Honeywell agreed on their merger in October 2000. Although each operates in a number of product lines, a key focus of the case was the complementary goods they produce for the commercial aviation industry. GE is one of three independent global manufacturers of large commercial aircraft engines, and Honeywell makes a number of systems essential for aircraft operation, ranging from landing gear to communications and navigation systems.

After agreeing to some changes to their transaction, including the divestiture of Honeywell's helicopter engine division, the parties received conditional clearance from the Justice Department in May 2001 to proceed with their merger. But the merger could not be consummated until it received clearance from the European Commission and other authorities. The Commission sought additional changes and conditions that were unacceptable to the parties. In July 2001 the Commission rejected the deal, and so the proposed merger did not take place.

The Assistant Attorney General for Antitrust issued this statement after that decision:

Having conducted an extensive investigation of the GE/Honeywell acquisition, the Antitrust Division reached a firm conclusion that the merger, as modified by the remedies we insisted upon, would have been procompetitive and beneficial to consumers. Our conclusion was based on findings, confirmed by customers worldwide, that the combined firm could offer better products and services at more attractive prices than either firm could offer individually. That, in our view, is the essence of competition.

The EU, however, apparently concluded that a more diversified, and thus more competitive, GE could somehow disadvantage other market participants. Consequently, we appear to have reached different results from similar assessments of competitive conditions in the affected markets.

Clear and longstanding U.S. antitrust policy holds that the antitrust laws protect competition, not competitors. Today's EU decision reflects a significant point of divergence.

For years, U.S. and EU competition authorities have enjoyed close and cooperative relations. In fact, there were extensive consultations in this matter throughout the entire process. This matter points to the continuing need for consultation to move toward greater policy convergence.

The European Union's objection to the merger centered around advantages that the combination would yield for the merged firm over its competitors in the markets for aircraft engines, avionics, and other aircraft systems. The Commission found that, among other factors, GE's vertical integration into aircraft leasing through its GECAS subsidiary, along with GE's deep financial resources, would lead inexorably to the merged firm's dominance in markets for certain aircraft systems. In addition, the Commission found that the merger would give the combined GE-Honeywell the ability and the incentive to offer complementary products on more attractive terms than could competitors with narrower product lines. This last category of objections has been termed "range" or "portfolio" effects.

The Commission found that these mechanisms would have the effect of driving the premerger competitors of both GE and Honeywell out of effective participation in their respective markets, presumably leading to higher prices in the long run as the merged firm became unconstrained by competitive pressures. U.S. antitrust authorities, in contrast, found that most of the alleged harms under the Commission's theory flowed from what are normally considered benefits of a merger—efficiencies that lead to lower prices. They

found little evidence that competitors would be unable to respond to any lower prices generated by the merger and thus be driven from the market. Finding more efficient combinations of productive resources that lead to lower costs and lower prices is, as the Assistant Attorney General said, the essence of competition. Blocking mergers that generate such efficiencies risks serious economic harm to consumers and to markets generally.

Elements of International Policy Convergence

Halting efficient multinational mergers destroys value precisely because an integrated, multinational firm can create specific efficiencies. As noted earlier, these may include exploiting economies of scale and scope, and combining central managerial guidance and appropriate pay for performance with the local knowledge of managers in various overseas markets. The European Commission might have been more likely to clear the GE-Honeywell merger if GE had agreed to divest its aircraft leasing subsidiary GECAS. But such a divestiture might have sacrificed efficiencies.

As the GE-Honeywell example indicates, there are some important differences in competition policy between the United States and other nations. But cases that produce such conflicting results have been rare and are likely to remain the exception. Moreover, steps toward appropriate convergence have already taken place, and this Administration is committed to seeking further convergence to promote the spread of sound antitrust policy. The United States should not seek convergence for its own sake, of course, but rather in order to establish certain core principles of sound competition policy across all jurisdictions.

Core Principles of Competition Policy

Competition policy should operate according to explicit guidelines, based on clear economic principles. Economic analysis should be central, because competition policy shapes fundamental economic decisions, such as production, pricing, and the organization of firms. These guidelines should reduce uncertainty by providing an indication to firms as to what kinds of conduct and transactions may bring scrutiny from competition authorities.

Competition policy should be concerned with protecting competition, not competitors, as a means of promoting efficient resource allocation and consumer welfare. There might be rare exceptions, such as certain monopolization cases, in which consumer harm is hard to measure, and then harm to competitors may be examined as an indicator of consumer harm. Indeed, harm to competitors does not play a central role in U.S. merger policy, although it does motivate private antitrust litigation. Since such competitor complaints are often at variance with consumer interests, antitrust

enforcement agencies and courts should view them skeptically. In the European Union the more significant and involved role of competitors in the merger review process has created a perception by some that the Commission's analysis is driven more by effects on competitors than is the case in the United States.

As the International Competition Policy Advisory Committee noted in its final report to the Attorney General in 2000, "Nations should recognize that the interests of the competitors to the merging parties are not necessarily aligned with consumer interests." Indeed, a merger may be opposed by competitors precisely because it would create a more efficient firm, one that will aggressively serve customers better than the existing industry configuration. Blocking such acquisitions deprives the world of an avenue to increased productivity.

The United States and the European Union have already achieved considerable cooperation and substantive convergence. U.S. and EU antitrust authorities have come to similar conclusions about a large number of transatlantic mergers. More work is required, however. The United States has undertaken several steps in bilateral and multilateral forums to facilitate convergence of competition policy to serve efficiency ends.

Bilateral Enforcement Agreements

The United States has entered into bilateral cooperation agreements with several important trading partners—Australia, Brazil, Canada, Germany, Israel, Japan, Mexico, and the European Communities—to facilitate antitrust enforcement. These agreements are implemented by the Justice Department and the FTC, working in cooperation with their counterpart agencies in the other countries.

These agreements typically provide for, among other things, sharing of nonconfidential information, coordination of parallel investigations, and positive comity. Under positive comity one country can request that another investigate possibly anticompetitive practices in its jurisdiction that adversely affect important interests of the country making the request. Such a request does not require the country receiving the request to act, nor does it preclude the country making the request from undertaking its own enforcement. The United States has also entered into one agreement, with Australia, under the International Antitrust Enforcement Assistance Act, which among other things allows the enforcement agencies to share confidential information.

The United States and the European Union have also created a working group to identify and pursue areas of possible further convergence in merger enforcement. Having completed a successful project on remedies in merger cases, the working group has established new task forces to examine conglomerate merger issues and other important substantive and procedural topics.

The International Competition Network

In October 2001 the Department of Justice and the FTC joined with top foreign antitrust officials to launch the International Competition Network (ICN). The ICN will provide a venue for senior antitrust officials from around the world to work on reaching consensus on appropriate procedural and substantive convergence in competition policy enforcement. The ICN will initially focus on multijurisdictional merger review (procedures, substantive analysis, and investigative techniques) and the advocacy role of antitrust authorities in favoring pro-competitive government policies. To facilitate the diffusion of best practices, the ICN will develop nonbinding recommendations for consideration by individual enforcement agencies. The ICN's interim steering group consists of representatives from a cross section of developing and developed countries, including the United States. It will hold its first conference in the early fall of 2002.

The World Trade Organization

The World Trade Organization (WTO) is an international institution in which the United States negotiates agreements with 143 other members to reduce barriers to trade. At the fourth WTO Ministerial Conference in Doha, Qatar, in 2001, members adopted a ministerial declaration. That declaration included a statement that the Working Group on the Interaction between Trade and Competition Policy will focus on the clarification of core principles, modalities for voluntary cooperation, and support for progressive reinforcement of competition institutions in developing countries. The role of the WTO and other international institutions in promoting economic well-being is detailed in Chapter 7.

Benefits of Appropriate Convergence

In some cases, the lack of antitrust harmonization may yield benefits. For example, in an unsettled policy area, in the absence of harmonization, nations might experiment with different competition policies. The world could then learn from these experiences what constitutes best practice in antitrust enforcement in the area in question. The bilateral and multilateral forums into which the United States has entered address this concern by sharing information to promote best practices. This consultation will enable the results of successful policy experiments to be disseminated. Moreover, the United States remains committed to appropriate convergence, in which efficient competition policies are spread worldwide, rather than seeking harmonization for its own sake and potentially promoting less than sound policies.

Dynamic Competition and Antitrust Policy

Through its influence on the development of competition policy over the years, economic analysis has brought a dramatic improvement in the ability of government agencies and the courts to accurately judge the strength of competition in a market. This has enhanced their capacity to distinguish those cases that properly raise concerns about anticompetitive effects from those that might have raised concerns in the past, but should no longer, in light of a better understanding of competitive forces. These changes in antitrust policy are important in that they afford firms greater flexibility to lower costs and improve their products through adjustments to their operations and organization.

But many of these improvements in policy have largely focused on better understanding markets in which firms compete with one another through incremental changes in the prices, quality, and quantity of relatively similar products or services. In some increasingly prominent industries, such as the information technology and pharmaceuticals industries, another important form of competition is taking place. It arises where there is a constant threat of innovations leading to a new or improved product being introduced that is far superior to existing products in a market. This type of competition is sometimes called competition for the market, or dynamic competition.

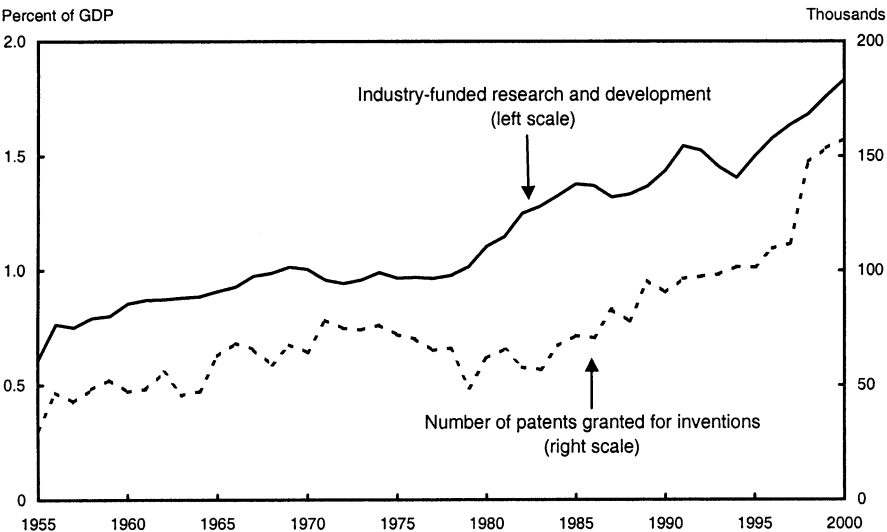
The increasingly important role of innovations in our economy can be seen in a number of indicators of innovative activity. After remaining nearly unchanged during the 1970s, industry's funding of research and development, measured as a share of GDP, grew two-thirds during the following two decades, reaching 1.8 percent of GDP in 2000. The number of patents granted each year by the U.S. Patent and Trademark Office provides some indication of the rate at which patentable innovations are being developed. Since the mid-1980s, the number of patents issued for inventions each year has grown dramatically (Chart 3-3). Although such a change could result from a number of other factors, such as increased incentives to file for a patent based on adjustments to the legal environment, evidence suggests that a burst in innovation is a driving factor behind this rise. Whereas some of the most visible innovations contributing to dynamic competition are technological in nature, such as improvements in the performance of computers, others may involve changes in management or business practices.

The importance of substantial innovations to the economy, as well as the unique form of competition they bring about, was recognized in 1942 by the economist Joseph Schumpeter. He noted that a significant part of the long-term growth of many industries resulted from what he called the “perennial gale of creative destruction.” At the heart of this creative destruction is the introduction of new products or services, technologies, or organizational

forms that lead to dramatic changes in an industry’s structure or costs, or in the quality of its products or services. In Schumpeter’s view, it was periods of creative destruction that brought “power production from the overshot water wheel to the modern power plant... [and] transportation from the mailcoach to the airplane.” Indeed, as he stated, the kind of competition resulting from firms bringing forth these changes or innovations is one that “commands a decisive cost or quality advantage and which strikes not at the margins of the profits... of the existing firms but at their foundations and their very lives.” Because of his early insights, dynamic competition involving the introduction of markedly improved goods or services is often referred to as Schumpeterian competition.

The significance of innovation—and hence of dynamic competition—will vary from market to market: it will be negligible in some and a pervasive force in others. Product improvements are commonly made in virtually all markets. But in markets experiencing the kinds of substantial innovation that Schumpeter addressed, these innovations can be so dramatic or disruptive as to make the products that they improve upon significantly inferior in comparison. The benefits of these innovations to society can be found all around us. Computer processors produced today are, by one measure, more

Chart 3-3 Industry-Funded Research and Development and Patents Granted for Inventions
Both industry-funded research and development and the number of patents granted by the Patent and Trademark Office have grown significantly since 1980.



Sources: Department of Commerce (Bureau of Economic Analysis and Patent and Trademark Office), National Science Foundation, and Council of Economic Advisers.

than 250 times more powerful than those produced in 1980, and more than twice as powerful as those produced in 1999. New drugs have vastly improved our ability to treat various illnesses. Other examples abound.

It has long been recognized that particular incentives are necessary to foster these market-transforming innovations. These innovations are often the result of substantial research and development investments on the part of companies or individuals. Since these investments must be made before it is clear that any profitable innovations will come of them, they are fundamentally risky. Encouraging innovation rests upon an interrelated set of internal and external rewards. The external rewards are those provided in the marketplace to the successful innovating organization. The internal rewards are those provided by the firm, joint venture, or other governance structure. Both economic organization and public policy therefore play significant roles in encouraging innovation.

Sources of Incentives for Innovation

The external risks and rewards facing firms in innovation-intensive industries are highlighted by a preliminary study of firms in the computer software industry between 1990 and 1998, which found that success, as measured by sales growth over this period, was by no means certain. But, compensating for this risk, some firms that did end up being successful were extremely so. At least 10 percent of firms saw sales fall to zero, and at least half experienced negative sales growth over the period. Only 25 percent of firms experienced real annualized sales growth of at least 7 percent during the period. But about 1 percent experienced real annualized growth of greater than 130 percent. This pattern of success highlights the risk involved in investments in these innovation-intensive industries. Therefore firms must have reason to expect that, taking into account the likelihood of failure, the profits from any successful innovations that do result from their efforts will be enough to justify the initial investment.

Intellectual Property Protection

Not only is investing in efforts to develop innovations risky and often expensive, but the innovations that result often produce beneficial knowledge or insights that others can copy at relatively low cost. Furthermore, in the absence of laws to the contrary, knowledge embodied in an innovation can be hard to keep others from using.

For instance, the research and development costs incurred by a firm in determining the correct chemical composition and treatment regime for a particular drug therapy may be substantial. But it may be difficult to keep much of this information out of the hands of competitors that have not

borne any of these costs, yet could use that information to produce the new drug themselves. As a result, competition between the innovator and imitators could keep the price of the drug at the cost of manufacturing it. In such a competitive environment, a firm's profits from its innovation would not suffice to cover its original research and development costs or justify its decision to risk undertaking expensive research efforts that may bear no fruit. Foreseeing this potential outcome, the innovator would have little incentive to embark on the research and development in the first place.

Even if a firm did not face competition from other firms benefiting from the knowledge produced by its innovation, firms or individuals may use aspects of the innovation for other purposes. Given how difficult it can be to keep them from doing this, in the absence of laws to prevent it, the innovator may receive little compensation from those that benefit from its innovation. As a result, the rewards that a firm enjoys from its innovation could fall far short of the benefits that the innovation produces for society. Consequently, in many cases, firms or individuals might not embark on developing an innovation because, although the social benefit from it may be large enough to justify its development costs, the firm or individual could not expect to reap enough of that benefit to justify those costs.

The consequences of this problem were recognized in the U.S. Constitution, which empowered Congress to develop a body of intellectual property laws, including those establishing patents. A patent for an invention confers on an individual or firm (the patentholder) limited rights to exclude others from making, selling, or using the invention without the patentholder's consent. Patents generally are granted for 20 years, and as the rights they provide imply, the patentholder can license to other individuals or firms the right to use its innovation. Patents give a firm the legal power to keep others from using its innovation to create competing products without bearing the cost of the innovation. Licensing provides a means whereby the innovator can receive compensation, in the form of licensing fees, from others that find a beneficial use for the innovation. Thus policy has long recognized that, to encourage innovation, firms must expect that successful innovations will yield a market position that allows them to earn profits adequate to compensate for the risk and cost of their efforts.

Indeed, intellectual property protection often plays an important role in dynamically competitive markets. But it is not the only mechanism that may allow a firm to gain an adequate return on risky investments in developing innovations. Intellectual property laws cannot always provide inventors complete protection against competitors using the knowledge embodied in their inventions without compensation. First, even if they are valuable, not all innovations can be protected by intellectual property law. Second, firms can often "invent around" a patent to create a competing product that,

although similar in value to consumers, is different enough in its composition or features so as not to violate the patent. Although this entails some development costs, these may be substantially reduced by the knowledge gained from studying the original innovator's efforts. On the other hand, some innovations may be difficult enough to imitate that, even without intellectual property protection, the innovator can enjoy a substantial cost or quality advantage over its competitors for some period. In either case, other characteristics of some dynamically competitive industries are important in making it likely that a successful innovation will yield a firm the leading position in a market, and profits that are essential to encourage such innovations.

Economies of Scale

Many industries that may experience dynamic competition are characterized by substantial economies of scale. In such industries, creating a new product entails high fixed costs, such as the costs of research and development and of setting up production and distribution facilities. But once these costs have been incurred, the incremental cost of making each unit of the product is small, indeed sometimes close to zero, and it is often easy to expand production to high levels. In markets with these characteristics, an innovator may be able to introduce its new product and keep production levels high enough to gain substantial market share before others can offer products of competing quality. As a result, economies of scale may allow the innovator to keep its average costs well below that of new entrants offering similar products that have smaller initial market shares. In some cases this advantage may be enough to keep other firms from providing significant competition unless they can offer a product that is notably superior.

Network Effects

Network effects are another mechanism that can help an innovator maintain a market-leading position in many dynamically competitive industries. A product or service is subject to network effects if its value to a consumer increases the more it is used by others. For instance, over the past decade, the number of people using e-mail has grown dramatically, making it a much more valuable means of communication for any individual user today than it was a decade ago. Network effects can also influence the value of some computer software. The more people who use a particular software application, or at least software compatible with it, the more valuable that software is to any individual who wants to share or exchange files with others who use that software. One study of prices of spreadsheet software between 1986 and 1991 found that consumers were willing to pay a significant premium for software that was compatible with Lotus 1-2-3, which was the dominant spreadsheet program during this period.

As more people use a particular good, its value to consumers can also increase because this wider use encourages the production of complementary goods. For instance, as more offices use a particular type of photocopier, businesses offering repair services and spare parts for that copier may become more common, making the copier even more attractive to offices.

As a result of these network effects, the value that consumers attach to a product that is already widely used may be substantially greater than the value they place on a relatively similar product that is used by fewer people. For instance, a manufacturer may introduce a new copier that offers performance largely similar to that of the market leader. But if the new copier is built in such a way that users cannot draw from the same service and spare parts network, it may be less valuable than the incumbent product. Thus, if a firm can quickly gain market share after introducing a new innovation, network effects can play an important role in helping the firm maintain that market leadership in the face of competition from new entrants offering similar products. This, in turn, increases its ability to reap the profits that are necessary for it to earn an adequate return on its risky investment.

Many have expressed concern that network effects can give such substantial advantages to incumbent products that new firms with potentially superior products are unable to compete. In theory, this could happen, but it does not happen necessarily. If a new product is clearly superior to the leading product, whether network effects are large enough to keep the new product from successfully competing will depend on the value of those effects compared with the net advantages it offers after taking into account the cost of switching to it. But, of course, measuring either of these—the value of the network effects or that of the new product's superior features—is difficult.

Although there have been cases where a new product took over a market-leading position from one that presumably enjoyed network effects, conclusive evidence that network effects have prevented the widespread adoption of a markedly superior product has not yet been found. For example, one common case put forward to argue that network effects can hinder the entry of superior products is that of the QWERTY keyboard, the familiar, century-old keyboard arrangement that virtually all typewriters used and that most computer terminals use today. In the 1980s a study suggested that a keyboard arrangement called the Dvorak keyboard, introduced in the 1930s by August Dvorak, was superior to QWERTY but had failed to gain market share because of the network effects that the already-established QWERTY enjoyed. Yet a more recent study raises significant doubts about claims that the Dvorak keyboard was superior. For instance, the most dramatic claims of its superiority are traceable to research by Dvorak himself, who stood to gain financially from the patented keyboard's success. Examination of his research revealed that experiments comparing keyboards

often failed to account for differences in the ability and experience of participating typists. The best-documented experiments, as well as recent ergonomic studies, suggest little or no advantage for the Dvorak keyboard. This highlights that generalizations cannot be made about the significance of network effects in deterring the entry of superior products into a market. Their impact must be judged on a case-by-case basis.

Fostering Innovation Through Organizational Structure

Although the prospect of gaining a market-leading position can encourage firms to innovate, firms can reap the benefits of innovation through other means as well. As was mentioned above, the benefits of innovation are often shared by many. Licensing agreements offer one means by which a firm can capture some of these spillovers. But such arrangements are an imperfect way of ensuring that innovators benefit from the spillover effects of their innovations while also encouraging additional beneficial uses of the innovation by others. As noted earlier, addressing this spillover problem is one motivation for a research joint venture among firms that expect to mutually gain from an innovation. Moreover, firms that develop new innovations subject to network effects will benefit from the production of complementary products that enhance those network effects. Partial equity stakes may provide a useful mechanism to foster the development of these complementary products.

Even when conducted within a single firm, successful research requires appropriate effort from multiple parties. This includes not only the work of research scientists and engineers, but also efforts by managers to craft an organizational structure that attracts and rewards such personnel appropriately. Thus, successful innovating firms must address various agency costs in product discovery and development, to align the interests of these various participants with the interests of the firm.

For example, one study indicates that research programs in pharmaceutical companies that encourage publication by their scientists experience higher rates of drug discovery. Whereas stock options are often the focus of discussions about means of resolving agency costs, this example makes clear that incentives must be carefully tailored to the desired objective. In this case, keeping a firm's researchers closely connected to leading-edge developments in fundamental science may provide a critical advantage in developing commercially valuable drugs. Thus, just as firms can use stock options as an incentive for managers to pursue shareholders' interests, so, too, they can create incentives for researchers to be connected to developments at the leading edge of their science, by making a researcher's standing in the greater

scientific community a significant factor in promotion decisions. A further study suggests that these firms provide a balanced system of incentives: those firms that use a scientist's publication record as a positive factor in promotion are also more aggressive in rewarding research teams that produce important patents. This reward structure helps direct scientists' efforts to engage in both basic and applied research, culminating in successful drug discoveries.

Decisionmaking at all levels of a firm can play an important role in determining its success in introducing substantial new innovations. A study of the computer hard disk drive industry found that established firms often had the technological know-how to develop what would turn out to be the next disruptive technology in their market, such as the 3.5-inch disk drive. In fact, they were sometimes among the first to develop them. But new entrants were always the leaders in commercializing the disruptive technologies examined in this study.

In this industry, the failure of incumbents to lead in commercializing disruptive innovations was often traced to decisionmaking that focused on the needs of their established market, failing to promote new technologies whose initial applications fell outside that market. Yet it would be these technologies that would eventually develop to become the leader in the established market. Thus the organizational structure and incentives faced by managers of established firms played a more important role than technological know-how in their failure to lead the commercialization of disruptive innovations. Of course, innovation benefits society whether it arises from established or from entrant firms, but in either case, successful innovation requires good organization.

Dynamic Competition as Repeated Innovations

All the factors we have examined—the market-transforming nature of some innovations, the presence of intellectual property protection, the potential for economies of scale, and the presence of network effects—provide explanations for why a firm can gain a market-leading position and earn high profits after introducing an innovation. But what makes a market subject to dynamic competition is the fact that the very same factors can allow another firm, with an even greater innovation, to take much or all of the market away from the leading firm. Indeed, as Joseph Schumpeter commented, the competition provided by new innovations “acts not only when in being but also when it is merely an ever-present threat. It disciplines before it attacks. The businessman feels himself to be in a competitive situation even if he is alone in his field.”

One example of a market where dynamic competition prevails today is that for personal digital assistants (PDAs). Apple Computer, Inc., made substantial investments to develop the Newton, the first handheld PDA,

which it introduced in 1993. This product did not succeed, but by 1996 at least six firms had operating systems for handheld PDAs either in development or already available to consumers. The Palm Operating System soon emerged as the preferred PDA, with a 73 percent market share in 1998. Although the innovations embodied in its products have made Palm a leader in this market, it is losing market share to new PDAs.

This example demonstrates a number of the elements often found in markets undergoing rapid innovation. First, firms that make substantial upfront investments in product development do not always experience the success necessary to gain an adequate return on those investments. Second, significant innovations can make a product the clear leader in a market at a particular point in time. Finally, even these innovative market leaders face challenges from later innovations by other firms that have the potential to make the leader's product obsolete. Therefore a potential innovator must believe that, if it gains a market-leading position through innovation, the resulting profits will be adequate to justify the development costs, given not only the possibility of failure but also the likelihood that future innovations will make any market leadership short-lived. Box 3-3 describes another market in which dynamic competition has been particularly intense.

Implications of Dynamic Competition for Competition Policy

Competition policy also has a role to play in markets characterized by dynamic competition. Markets experiencing rapid or substantial innovation can still be subject to conditions or behavior by firms that hinder competition. For instance, price fixing among firms will harm competition even in industries undergoing dramatic innovation. Other behavior may have more ambiguous implications for competition, dynamic or otherwise. Therefore the antitrust agencies will continue to scrutinize behavior by firms in these markets. Since the lawfulness of certain actions by a firm depends, in part, on the degree of competition in the firm's market, the ability to properly assess all types of competition is essential. Consequently, the analytical framework used to assess competition must encompass its potentially dynamic dimension. This involves recognizing the shortcomings of traditional methods for assessing competition when applied to markets undergoing rapid innovation, and developing new methods for determining how significant dynamic competition is in a particular market.

Highlighting the importance of developing and applying such methods is the fact that markets characterized by significant dynamic competition may not appear competitive through the lens of some common tools of traditional competition policy. Thus continuing adjustments in competition

Box 3-3. Dynamic Competition in the Market for Prescription Anti-Ulcer Drugs

The dramatic nature of innovations in the drug industry can give a firm that introduces a new drug significant market share. But subsequent, equally dramatic innovations by competitors can make this market leadership short-lived. Such leapfrog leadership is one characteristic of markets subject to dynamic competition.

As an example, in 1977 SmithKline introduced the first anti-ulcer prescription drug, Tagamet. Just 6 years later, however, Glaxo plc introduced a competing drug called Zantac. Compared with Tagamet, Zantac had fewer adverse interactions with other drugs and needed to be taken only twice rather than four times a day. Within a year, on a revenue basis, Zantac had gained more than a quarter of the market for prescription anti-ulcer drugs, and by 1989 that share had risen to more than half while Tagamet's had fallen to about a quarter (Chart 3-4).

In 1989 Merck & Co., Inc., introduced a drug developed by Astra AB called Prilosec, the first of a new class of anti-ulcer drugs called proton pump inhibitors. The new drug had to be taken only once a day. Also, studies have shown that it heals a greater percentage of patients than Zantac does in a 4-week period. By 1998 Prilosec accounted for about half of total sales revenue for prescription anti-ulcer drugs, while Zantac's share of sales revenue had fallen to about 5 percent. (In the wake of mergers and other developments, the names of the firms that sell all three drugs have changed.)

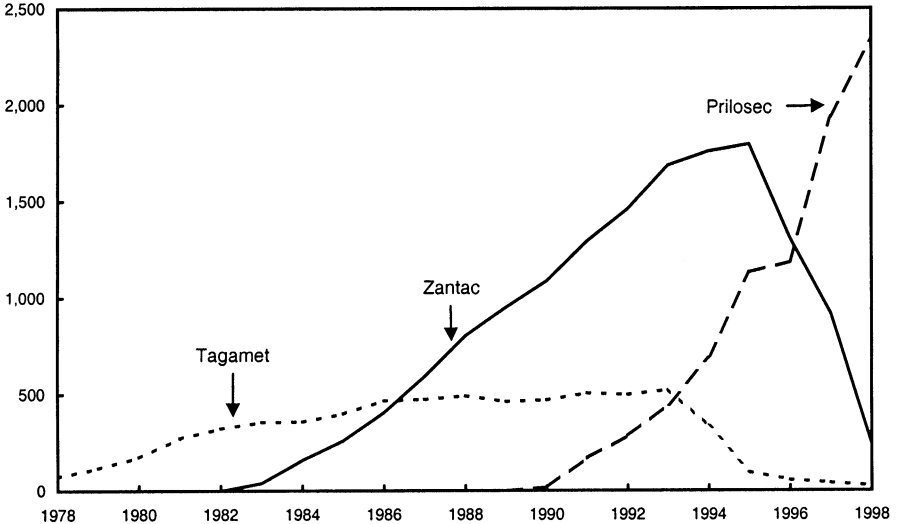
This example demonstrates the rapid rate of innovation in the drug industry and how it can quickly render obsolete even highly innovative drugs that companies have spent hundreds of millions of dollars developing. In such a competitive environment, patents play an essential role in encouraging firms to spend the huge resources needed to develop ideas and products that competitors could easily copy in the absence of legal protection.

This example also shows that, even with a patent, a firm can see its market share taken away by another firm that develops an even better drug for the same illness or condition. In this example, Prilosec was introduced into the market well before Zantac's patent expired. Given the substantial upfront investments in drug research and development, companies will be motivated to develop drugs only if successful drugs can achieve high profits and capture a leading market share in the relatively short time before new innovations emerge. In the drug industry, substantial market share can easily be lost in just a few years.

Chart 3-4 Sales Revenue of Selected Prescription Anti-Ulcer Drugs

The rise and fall of sales revenue for these three anti-ulcer drugs reflects how dynamic competition can lead to substantial success, but can also make that success short-lived.

Millions of current dollars



Source: IMS Health Incorporated.

policy are needed to avoid incorrect conclusions. Likewise, continuing adjustments are needed to correctly identify markets in which high profits and market leadership cannot be explained by the ongoing nature or pace of innovation, suggesting that the market may indeed not be competitive.

As noted in the discussion of merger policy above, a market's degree of concentration is typically used as a screening mechanism to evaluate competition in that market. Although finding that a market is highly concentrated does not by itself suffice to conclude that competition is limited, finding that it is not highly concentrated usually does suffice to allay any such concern. Thus measures of concentration provide a useful screen, because many markets may not be concentrated enough to warrant further investigation.

However, given the significant role of innovation in markets characterized by dynamic competition, it is common to see one leading firm that, through innovation, has for the time being created a superior product. Although such a market would be highly concentrated, there may in fact be substantial dynamic competition in the market, with new innovations emerging to threaten the leading firm's position. Consequently, because many markets undergoing rapid innovation will have a high measured concentration, such measurements may not be as useful a screening device if dynamic competition is the primary form of competition in that industry. In light of this

shortcoming, the development of effective screening mechanisms to evaluate dynamic competition may be a useful supplement to concentration measures. Such screening mechanisms could allow businesses in innovative industries to better predict the responses of antitrust agencies to their actions, just as the safe harbor provisions relating to concentration measures did in the 1980s.

In assessing competition in a market, antitrust agencies and the courts also examine whether the threat of entry by a firm into that market would be both likely to occur and sufficient to counteract any ability of existing firms to exercise significant market power. However, for it to be adequate to assuage concerns, entry in response to such behavior must generally be able to take place within a period of 2 years, essentially ensuring that the incumbent firm or firms' ability to profitably raise prices is only that durable. As the length of patents indicates, firms may need substantially more than 2 years for profits to provide an adequate return on their research and development investments. Moreover, in a typical assessment of the impact of a merger on competition, the threat of entry can be viewed as adequate to counteract anticompetitive price increases if it would prevent the merging firms from keeping prices significantly above premerger levels. But as Schumpeter pointed out, even if they may take longer than a few years to emerge, innovations in dynamically competitive markets may not only reduce incumbents' profits that are above competitive levels, but indeed threaten the very viability of incumbent companies. Such competition surely threatens the durability of a firm's market power.

Some common tools of antitrust policy may thus be less complete and informative in dynamically competitive markets than in other situations. But just as the antitrust agencies improved on simple concentration measures in assessing competition during the late 1970s and early 1980s, so, too, the existing toolkit can be further augmented to deal with dynamic competition. The central role of innovation in these markets suggests the kind of information that is useful in assessing this type of competition.

In general, antitrust enforcement must continue the effort to understand the patterns, nature, and pace of innovation in a given market. In established industries, the antitrust agencies and the courts can examine firm and industry history to assess the significance of innovative activities. These activities would include research and development expenditures and complementary investments in production or distribution that would have much less value if the product they support lost its market to a competitor's innovation. The risky investments associated with developing innovations go well beyond research and development to include all investments that future innovations could render obsolete.

An industry's history can also provide indications of the fragility of market leadership to substantial innovations in that industry. For instance, the history of innovations in the market for prescription anti-ulcer drugs, reviewed in Box 3-3, suggests that the threat of future innovations will remain an important competitive force. Where such threats are important, one might conclude that the industry is dynamically competitive.

Brand-new industries, of course, lack such a history. Nonetheless, antitrust officials should still endeavor to assess the importance of innovative activity in these markets, and thus the potential significance of dynamic competition. For both new and old markets, the potential for competition from developments in other rapidly innovating fields should also be considered—even if the technologies of the respective fields are fundamentally different—as long as the application of those technologies is converging. For instance, vascular grafts are used today to repair and replace diseased or damaged blood vessels. But any assessment of competition in that market must take into account the potential for substantial innovations in other invasive procedures or in drug therapies that could either reduce the incidence of diseased or damaged blood vessels or provide alternative treatments. In both new and established industries, we must encourage dynamic competition and the benefits of innovation it secures, by updating competition policy appropriately.

Such updating has already taken place with respect to the scope of intellectual property protection and the effect it might have on other firms' abilities to innovate. Although intellectual property protection is important to encourage firms to innovate, it can also be used in ways that hinder the development of future, and potentially competing, innovations by other firms. The FTC and the Justice Department have addressed this possibility in guidelines that recognize the interaction between intellectual property law and antitrust law. These guidelines encourage the development of new technologies and the improvement of existing ones, while seeking to preserve the desired incentives underlying the creation of intellectual property.

Conclusion

Antitrust policy has contributed greatly to the economy by fostering competition and allowing the efficient adaptation of markets to new opportunities. This chapter has showcased some recent changes in the organization of economic activity and market competition and outlined the adjustments that competition policy is making in response.

First, corporate governance and structure continue to evolve, as the rapid pace of merger activity proceeds and hybrid organizational forms such as joint ventures and partial equity stakes continue to be established.

Competition policy should be sensitive to the efficiencies that new structures have brought and can continue to bring to society. Since a large source of these efficiencies may be rooted in managerial and organizational improvements, it is worthwhile for the enforcement agencies to investigate such factors thoroughly.

Second, the growth of multinational enterprises and cross-border mergers will continue to make more goods and services available to consumers at lower cost. But possible anticompetitive concerns arising out of such mergers can now result in reviews by antitrust authorities from many nations. The application of inefficient competition policies worldwide could harm U.S. interests. The United States is working to narrow divergences in countries' competition law and policy through cooperation with other national antitrust authorities, under a number of bilateral cooperation agreements. Through the creation of the International Competition Network, the United States has joined with other nations to facilitate procedural and substantive convergence.

Finally, competition policy in the United States and abroad must address the greater prominence of markets characterized by dynamic competition. Competition policy should take into account that characteristics, such as high profits and substantial market share, that might warrant concern about competition in some markets may mask vigorous dynamic competition among firms in innovation-intensive markets.

Promoting Health Care Quality and Access

Health care is one of the largest sectors of the American economy, and one of the most vibrant. Biomedical research has led to dramatic advances in our understanding of the human genome, basic biology, and mechanisms of disease, and in our ability to diagnose and treat illness. More researchers from the United States have been awarded Nobel prizes in medicine in the past 40 years than from all other countries combined. Innovative diagnostic and imaging tools have improved our understanding of diseases and our ability to identify illnesses quickly, accurately, and painlessly. Novel drugs, devices, and techniques have dramatically improved the treatment of a wide range of illnesses. New information systems, including those relying on the Internet, allow health care providers to work more effectively with their patients to manage illnesses and avoid complications. These advances testify to the success of our health care system in encouraging discovery and innovation. Coupled with a strong tradition of dedicated, professional care, they hold great potential for further improvements in the health of Americans.

Evidence from biomedical, epidemiological, and economic studies confirms that these technological advances have made Americans far better off. An American born in 1990 can expect to live 7 years longer than an American born in 1950. The mortality rate from coronary heart disease, the Nation's leading killer, has declined by 40 percent since 1980, both because of reductions in the incidence of serious heart events like heart attacks and because of better outcomes when those events occur. Among seniors, rates of disability have declined by more than 20 percent in the past two decades. Many complex factors have undoubtedly contributed to these improvements. For example, better scientific understanding of diseases has enabled Americans to make lifestyle changes, such as quitting smoking, to reduce their risk, and improvements in economic conditions and public health have enabled more people to avoid environmental health risks. But a growing body of research indicates that medical technology played a starring role in these dramatic improvements.

Thanks to these innovations, the number, scope, and quality of available medical treatments have risen dramatically. These improvements in medical treatment, rather than rising prices or other causes, have been the single most important contributor to growth in medical expenditure. In large part as a result of the expanding capabilities of medical care, the United States now spends 13.4 percent of its GDP on health care, and this figure is predicted to

rise to 15.9 percent by 2010. There is growing evidence that, on average, the health improvements resulting from newer, better, and more intensive treatments have been well worth the added cost. But there is also growing evidence that substantial opportunities remain both to reduce costs and to achieve greater health improvements through more effective use of medical services—that is, to improve the value, or output per dollar spent, of our health care system. Even though the American health care system provides high-quality care overall, too often Americans receive neither the best care nor the best care for the money. Whether lower value care results from the underuse of basic preventive services, the overuse of medical procedures in patients unlikely to benefit from them, or the misuse of treatments resulting in preventable complications, there is tremendous potential to improve the value of health care in the United States.

With rising health care costs have come rising concerns about the affordability of health care. Many health care expenses are unpredictable, and serious illnesses have the potential to place households in financial peril. Insurance is a standard solution: in a well-functioning insurance market, individuals pool their risks, trading unpredictable and potentially large expenses for much smaller, more certain expenses in the form of insurance premiums and copayments. Yet about one in six Americans lacks any kind of health insurance, and many more Americans are concerned about the value of available health insurance plans. Providing high-value health insurance is not easy. Generous, first-dollar insurance does provide protection against the high costs of medical treatment, but by eliminating incentives to weigh the costs of medical care against its expected benefits, it also contributes to the overuse and the misuse of medical care.

Health care also differs from many other goods and services in that Americans generally believe that basic health care should be available to all members of society, even those with little or no ability to pay. Public support in the form of assistance with health insurance and health care costs helps achieve this goal and accounts for well over \$400 billion annually in Federal expenditure and forgone tax revenue. In the past, advocates for expanding government health insurance programs such as Medicare and Medicaid to address the problem of uninsurance have maintained that “guarantees” of coverage, plus government regulation of prices for covered services, could provide high-value health care services. But government health care plans have faced enormous difficulties in keeping up with innovations in medical practice and in providing high-quality, innovative care. Medicare still does not cover prescription drugs, and Medicare beneficiaries must increasingly rely on supplemental private insurance to provide acceptable coverage. Many Medicaid plans, facing rapid cost increases and very low provider participation rates under the traditional approach of regulated fee-for-service insurance,

are adopting alternative strategies to provide coverage. Other major industrialized nations with larger public health insurance programs, such as France, Germany, Japan, Switzerland, and the United Kingdom, are also experiencing rapid growth in expenditure and problems with the provision of high-quality care.

Private health insurance also has faced difficulties in supporting high-value health care. In the early 1990s, advocates of managed care believed that plans combining insurance with new financial and other incentives for health care providers to control costs could result in higher value care. But although managed care did contribute to a slowdown in medical cost growth in the mid-1990s, public uncertainty about the quality of care in managed care plans has increased, and this uncertainty has been accompanied by a return of rapid cost increases in private insurance. Many Americans are not satisfied with the cost and quality of the public and private health care coverage options now available to them.

Another important obstacle to high-value care is the quality of information available in markets for medical care. In most market settings, consumers' purchase decisions are based on good information on the value of the products they buy. But in health care the lack of good information on the success of different treatments—in terms of the best outcome per dollar—means that individuals and families have difficulty making informed decisions, and insurance companies are not rewarded for altering their coverage to encourage high-value care. Thus strategies to improve the value of care include supporting the development of better information for patients and providers on high-quality, high-value treatments.

In the face of these various problems, many have concluded that American health care policy is again at a crossroads, with fresh policy approaches needed to support innovative health care in the future. New policy directions are being proposed, a consistent theme of which is the encouragement of patient-centered care—care that puts the needs and values of the patient foremost and makes the patient the primary clinical and economic decision-maker, in partnership with dedicated health care professionals. Patient-centered care requires more flexibility and innovation in health care coverage; it also places more responsibility on the patient—and less reliance on third-party payers and government regulators—to avoid wasteful costs. To encourage the development and use of such innovative coverage options, competitive choices among health insurance plans and among health care providers are more important than ever. In turn, effective competition to help all Americans get the care that best meets their needs requires innovative, market-oriented health care policies.

To achieve more patient-centered health care by encouraging innovations in the financing and delivery of services in this dynamic sector of the economy, the Administration is pursuing three broad objectives:

- *Develop flexible, market-based approaches to providing health care coverage for all Americans.* Markets respond more rapidly than bureaucracies to the changing technology and new innovations in products and services that characterize the American health care system. Market flexibility and competition are essential if medical treatment decisions are to reflect patients' individual needs and personal preferences and are to be based on the best available evidence on benefits and costs. Important obstacles to innovation in health care coverage must be addressed, such as the potential for competing plans to reduce costs by designing benefits to attract healthier enrollees rather than by providing more efficient care for all persons regardless of their health risks. But these obstacles must be addressed through health care policies that increase rather than reduce insurance coverage rates. Competition need not threaten the quality of care received by those with the least ability to pay; rather, government support and oversight can be better directed to ensure that all Americans are able to participate effectively in a competitive health care system.
- *Support efforts by health care providers and patients to improve the quality and efficiency of care.* The incentives provided by a truly competitive system of health insurance coverage choices are an essential foundation for a high-quality, efficient health care system for the 21st century. But other policy changes are also needed to create an environment for medical practice that encourages high-quality, efficient care. Government and private health care purchasers can also help patients and providers develop and use better information on the quality of care, improving the ability of patients to identify high-quality providers and plans and helping providers deliver better care. Improving the environment for medical practice also includes reforming the litigation systems dealing with medical liability and reducing regulatory barriers to innovations in health care delivery.
- *Provide better support for biomedical research.* Outstanding basic research and path-breaking biomedical innovations have already had enormous payoffs, generating long-term public benefits. Because of the high returns on these investments, Federal support for biomedical and other scientific research should be enhanced. At the same time, the Federal Government can expand and improve the knowledge base for medical practice, by supporting projects that analyze which treatments work best for whom, how they can be delivered safely, and which health care providers are doing the best job for their patients.

The remainder of this chapter explores each of these critical issues for improving the quality and value of health care in more detail. As treatment options continue to multiply and costs continue to increase, improvements in the value of health care would make Americans more willing to purchase coverage for themselves and to pay the taxes required to subsidize it for those who need additional assistance.

Encouraging Flexible, Innovative, and Broadly Available Health Care Coverage

Recent Trends in Health Care Costs and Coverage

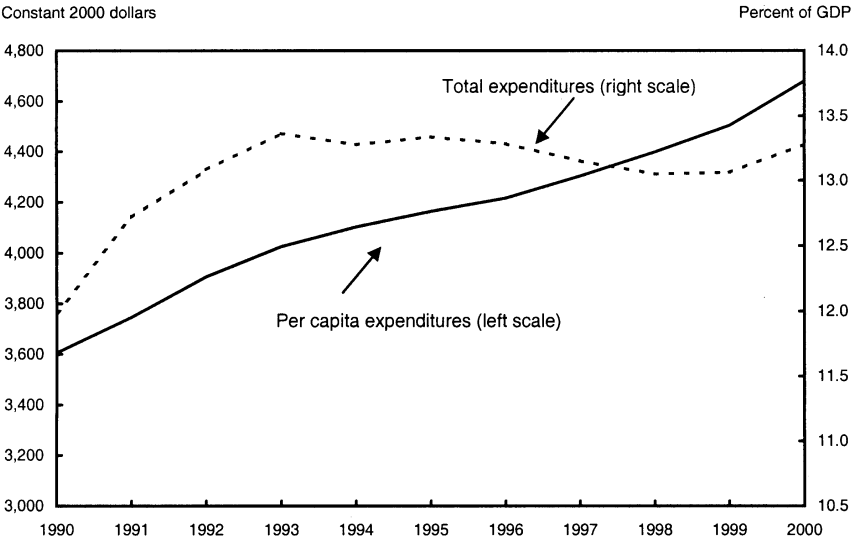
Health care spending grew rapidly during the past decade, from \$916.5 billion in 1990 to \$1,311.1 billion in 2000, or more than 3.6 percent a year on average (2.6 percent a year in per capita terms; Chart 4-1). Home health care expenses and drugs were the fastest growing categories of this expenditure (Chart 4-2). The real, constant-dollar cost of private health insurance increased by 4.9 percent a year between 1984 and 1999. Since the 1980s, health care benefits have also increased substantially as a share of total compensation for workers. Growth in health care costs is projected to accelerate, with total expenditure predicted to account for 16 percent of GDP by 2010. Over the longer term, forecasts predict that health care spending will become even more predominant in the economy, continuing a 60-year economic trend and reaching as much as 38 percent of GDP under conservative assumptions.

Rising costs of private health insurance in the 1980s and early 1990s led to the emergence of managed care in private health insurance plans. Managed care seemed to offer a solution to a fundamental health care dilemma. Its small copayments and low out-of-pocket limits protected individuals from substantial out-of-pocket health care costs. At the same time, its cost control mechanisms—including capitated payments, preferred provider networks, preapproval and utilization review requirements, and restricted formularies discouraged the use of some discretionary medical services whose benefits were likely to be low relative to their cost. In traditional fee-for-service health insurance, in contrast, third-party insurance made patients and providers less sensitive to the value of medical services per dollar spent.

In the mid-1990s, managed care succeeded temporarily in limiting cost increases, largely by negotiating lower payments to providers for specific services, and by discouraging utilization of some medical services and avoiding some costly complications of inappropriate treatment. Thus, for a

Chart 4-1 **Health Care Expenditures**

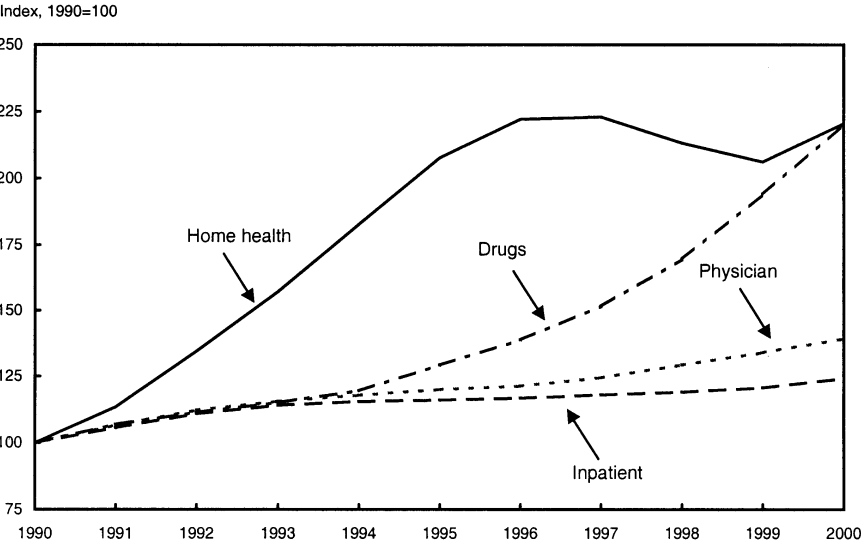
Health care expenditures grew substantially during the 1990s, both per capita and as a share of GDP.



Source: Department of Health and Human Services (Centers for Medicare and Medicaid Services).

Chart 4-2 **Expenditures on Components of Health Care**

Home health and drugs were the fastest growing components of health care expenditures during the 1990s.



Source: Department of Health and Human Services (Centers for Medicare and Medicaid Services).

while, managed care by and large achieved its primary goal: bringing the rise in insurance premiums under control without compromising quality of care. Today, however, with the perception that managed care has often focused more on reducing costs than improving quality, many of the managed care approaches to controlling cost increases may be reaching their limits: providers are negotiating more effectively with health plans, patients are pressing for greater choice of providers, restrictions on treatment choices are being challenged in courts and legislatures, and few additional easy targets for reducing costs remain (Box 4-1). As a result, premiums for private health insurance are again rising rapidly.

Public health care spending has grown rapidly as well, so that government-sponsored health insurance plans are facing cost increases that seem difficult for taxpayers to sustain. Federal, State, and local governments have long been involved in the financing, provision, and regulation of health care services. The Federal Government directly spends over \$200 billion annually for the Medicare program, which provides health insurance for nearly all elderly and disabled Americans, and over \$100 billion annually for Medicaid, the joint Federal-State program that provides health insurance for low-income and medically needy populations. Federal Medicaid funds are matched by almost

Box 4-1. Managed Care: Good, Bad, or Somewhere in Between?

The managed care option is an important one for many Americans. The vast majority of nonelderly Americans with private insurance are now enrolled in some form of managed care, representing a sea change in health insurance coverage over the past decade. The reputation of managed care organizations has suffered in recent years, however, and the widespread perception, based largely on anecdotal cases, is that care is worse. To what extent does research on the performance of managed care plans bear out this perception? Not surprisingly, the picture is mixed.

A large number of studies that have looked at quality of care have found no significant differences between health maintenance organizations (HMOs) and fee-for-service plans. Along some dimensions, such as the routine management of chronic illnesses and the provision of preventive care, HMOs tend to perform better. Many managed care programs are better able to implement systematic monitoring of quality of care, particularly for chronic and preventive care. In one study, for example, only 35 percent of women in fee-for-service plans received scheduled mammograms, whereas 55 percent in managed care plans did. In addition, because they have been able to negotiate

continued on next page...

Box 4-1—*continued*

lower prices from their network providers and for their formulary drugs, many HMOs have been able to offer more comprehensive benefits, such as lower copayments on prescriptions. In turn, this may contribute to better adherence to recommended drug therapies and other treatments among patients in HMOs.

However, certain studies have found better performance in fee-for-service plans in particular instances, especially those involving more costly management of patients with complex illnesses. Although they do not make a compelling general case against HMOs, these studies provide some cautionary evidence that particular attention should be directed toward ensuring that plans have good incentives to care for patients with predictably costly diseases. This can be accomplished through public policies that discourage risk selection and that provide good information on quality of care for people to use in choosing plans.

Private insurance markets have already responded to such concerns. For example, HMOs with closed networks are not the most popular or the fastest-growing form of managed care coverage today. Over the past 5 years, employee enrollment in preferred provider and point-of-service plans has increased from 42 percent to 70 percent, while enrollment in traditional HMOs has decreased from 31 percent to less than 23 percent. Overall, the vast majority of enrollees are in some form of coordinated care. The major exception to this trend is the Medicare program, which has a low rate of HMO enrollment (because of significant payment and regulatory problems) and has had considerable difficulty making preferred provider organizations, point-of-service plans, and other nonnetwork managed care plans available.

\$80 billion in State and local contributions. The Federal Government also provides approximately \$100 billion a year in tax exclusions to support private health insurance for workers who receive coverage through their employers.

Historically, the Medicare and Medicaid programs have been government-run, fee-for-service insurance plans. They have controlled growth in costs through tight price controls and restricted coverage. For example, Medicare's government-run plan does not cover prescription drugs or widely used disease management programs that assist beneficiaries with chronic illnesses. This is in part because the introduction of new benefits in government-run programs tends to require either extensive rulemaking or new legislation, and in part because of policy concerns about the potential costs of these benefits. Access to treatment may also be restricted when physicians refuse to

participate in a program, because of either administrative complexities or (in the case of Medicaid) low fee-for-service reimbursement rates in many States. The combination of tight price controls and restrictions on access to treatment is likely to make it even more difficult for government-run health insurance plans to keep up with treatment innovations in the future.

Despite these efforts to control costs, annual Federal Medicare expenditure (in constant 2000 dollars) increased from almost \$141 billion to \$215 billion between 1990 and 2000, and combined Federal and State Medicaid spending almost tripled, rising from \$95 billion to \$202 billion. The faster growth in Medicaid spending resulted from expansions of eligible populations, including new coverage through the State Children's Health Insurance Program (SCHIP), and from more rapid growth for certain benefits, including outpatient prescription drug coverage for some recipients and long-term care services—benefits not included in Medicare. Both Medicare and Medicaid are expected to continue to grow rapidly relative to Federal budget resources. Over just the next 10 years, Medicare spending is expected to double, as is Medicaid and SCHIP spending. Medicare has dedicated payroll tax financing for its hospital insurance (Medicare Part A) benefits, but the 2001 Medicare trustees' report projects that by 2016 the system will begin to spend more than its tax revenues bring in, and that by 2029 the program will become insolvent, unable to pay these benefits. Furthermore, these hospital insurance benefits account for only a portion of Medicare expenditure. Supplemental medical insurance (Medicare Part B) expenditure is financed primarily by general revenue. Without program changes, by 2030 Medicare is projected to account for 4.1 percent of GDP and 21.9 percent of Federal revenue, and Federal Medicaid payments are projected to equal 2.4 percent of GDP and absorb 12.8 percent of Federal revenue. Medicaid and SCHIP are also creating growing budgetary pressures for States: already the programs account for around 20 percent of aggregate State spending.

Although still high, the proportion of the population covered by health insurance has generally been falling as health care costs have been rising. This rise in the uninsured population has occurred despite the substantial eligibility expansions for Medicaid and SCHIP and despite the growing share of Americans eligible for Medicare. In the absence of new policy directions, a further decline in the number of Americans with access to health insurance is a serious risk, as a result of loss of jobs or reductions in benefits, even if further expansions of eligibility for government programs occur. These trends, considered in more detail below, provide important lessons for encouraging competitive innovations in health care coverage, whether in private insurance markets or in public programs.

Addressing Barriers to Effective Competition in Health Insurance

In most sectors of our economy, competitive private markets coupled with good information work well to improve the welfare of Americans. Tight government regulation and extensive direct government financing are not needed. The health care market has traditionally been regarded as different, however, for several reasons. Among these are potential inefficiencies resulting from adverse selection and moral hazard; an insufficiency of information available to patients, health providers, and insurers; and societal concerns about access barriers for lower income or disadvantaged Americans. Some have argued that these problems create fundamental obstacles to competitive approaches to health care delivery, requiring extensive Federal involvement in regulation and financing.

Tighter regulation and increased Federal oversight, however, are likely to lead to the same kinds of inefficiencies and stagnation seen in other highly regulated industries. Even Medicare, which has primarily consisted of government-provided fee-for-service insurance for elderly and disabled Americans, has long included some competitive private health plan options. To preserve and improve health insurance options for all Americans, the Federal Government can encourage policy reforms that improve the functioning of health care markets, building on steps already being taken by public and private payers.

A crucial obstacle to the effective functioning of competitive markets for health insurance is the problem of adverse selection. Adverse selection occurs when people who expect to incur significant health expenses sign up for more generous, less restrictive health plans in greater numbers than do healthier people. Because these more generous plans attract patients with higher medical costs, premiums for those plans are driven even higher, making the plan even less attractive to healthy individuals, in a classic “death spiral.”

Careful policy design, however, can help prevent problems associated with adverse selection. Many large employers, including many States and the Federal Government, have adopted a variety of competitive systems that offer choices to the populations they cover. The following steps can reduce selection problems:

- *Introduce benefit standards.* In the absence of any benefit standards, insurance plans could attract a healthier mix of enrollees by reducing benefits and insurance premiums, potentially undermining the insurance protection offered and driving up the costs of competing plans that have less healthy enrollees. By contrast, broad, flexible standards—such as requiring catastrophic protection and some coverage for all common health problems—have encouraged stable competition among a variety of types of plans in the Federal employees’

system and other successful competitive choice systems used by large private employers. However, specific coverage mandates—such as inflexible restrictions on copayments or required coverage for particular types of medical services—may not only exacerbate adverse selection, by causing more individuals to drop coverage entirely, but also unduly inhibit innovations in coverage.

- *Adjust premiums for risk.* Some purchasers implicitly or explicitly require additional contributions for the plan choices of higher cost enrollees. For example, plan payments might be adjusted based on age, sex, and certain health characteristics (Box 4-2). Medicare is currently expanding its risk adjustment factors to include a range of chronic health conditions.
- *Limit enrollment periods.* Employer plan choice systems generally allow plan changes only during a once-a-year “open enrollment” period, except in special circumstances. The limited lock-in period reduces the likelihood that people will enroll in an inexpensive plan with limited benefits and then switch to a more generous plan just when treatment is needed for a health problem.
- *Provide limited additional subsidies for higher cost plans.* In some competitive choice systems, employer contributions are set equal to a flat amount. In contrast, in the Federal employees’ program and many other employer purchasing groups, employer contributions increase with the health plan’s cost over some range of plan choices, reducing adverse selection pressures. Recent proposals for improving competition in Medicare and for providing assistance for purchasing private coverage in the form of refundable tax credits would provide partial subsidies for additional expenses, up to a cap.
- *Introduce health care accounts.* Dedicated accounts that provide a tax-favored “buffer” in the event of significant health expenses can make plans with nontrivial out-of-pocket payments more attractive to workers who perceive themselves as having a higher risk of significant expenses. This may reduce the extent to which high-risk individuals tend to choose more generous plans, and at the same time give individuals more control over their care.

There is now considerable evidence that the savings from efficiency gains due to the adoption of competitive systems in large purchasing groups are generally more than adequate to support even costly steps to control adverse selection. Such steps can include providing some limited or partial subsidies to help sustain the higher cost plans that some of the covered populations prefer.

For insurance markets involving small firms and individuals without access to group coverage, adverse selection problems can be more severe. To varying degrees, States permit providers in the market for individual insurance to rate

Box 4-2. The Need for Good Risk Adjustment

Price competition in insurance markets can be a powerful force for efficiency, but it must be used carefully if it is to result in better care for patients. Consider, for example, a large firm that offers its workers a menu of insurance plans. If the firm pays the insurer a flat, or “capitated,” fee for each enrollee, insurers offering these plans will have an opportunity to increase their profits by enrolling only the healthiest patients, since they will tend to have the lowest medical spending. In this situation the financial incentive for the insurer is not to provide high-quality, high-value care, but simply to identify and enroll healthy patients. The same issue arises in Medicare or Medicaid, when enrollees choose a managed care plan and the plan receives a capitated payment from the government for providing care.

Public or private plan sponsors can correct this incentive through risk adjustment, that is, adjusting their payments to the insurers on the basis of risk. Insurers need to be paid more to cover enrollees with higher expected medical spending, to remove the incentive for “cream skimming.” Instead, plans will have an incentive to improve the quality of care so as to attract all patients.

The best practices for risk adjustment continue to evolve. Although it is very difficult to predict an individual’s future medical spending, researchers are developing more effective techniques for doing so. Moreover, there is growing evidence that many medical expenses are not predictable and that, in the vast majority of cases, very high expenditures, when they occur, do not persist for many years. Some types of predictable expenses do not reliably or uniformly influence health plan or provider choices.

Medicare and Medicaid have played an important role in the development of effective risk adjustment techniques. For example, Medicare is developing a system of risk adjustment that relies on detailed diagnostic information collected from both inpatient and outpatient sources. As risk adjustment techniques continue to improve, health plans will increasingly have to compete for enrollees on the basis of the quality of care they provide.

each individual on the basis of his or her medical risks and past medical expenditure. The practice of underwriting is not controversial for many lines of insurance, such as automobile and home coverage, where differences in claims are largely the result of voluntary individual behaviors such as driving habits. In health care, however, a significant part of an individual’s disease risk is outside his or her control. To reduce the extent to which high-risk individuals face higher premiums, and to improve the availability of certain

health insurance benefits, States and the Federal Government have imposed a range of restrictions on insurance underwriting practices as well as coverage mandates on nongroup (and in many cases on group) health insurance plans. The 1996 Health Insurance Portability and Accountability Act imposes some Federal requirements on insurance offered by private insurers, so that individuals who change jobs but wish to continue their health coverage face only limited underwriting restrictions in doing so. Some States impose more significant restrictions on insurance underwriting practices, in the form of guaranteed issue and community rating requirements.

Such restrictions tend to reduce insurance premiums for high-risk individuals but increase them for lower risk individuals; they may also encourage individuals to wait until they have a significant health problem before enrolling. The result may be less insurance coverage and only limited reductions in premiums for chronically ill individuals, as healthier individuals choose to forgo coverage entirely rather than pay higher premiums. Thus it is an empirical question to what extent the benefits of making coverage more available for high-risk individuals outweigh the costs of higher average premiums and insurance rates. Stringent underwriting restrictions in individual insurance markets, such as guaranteed issue and community rating, may severely limit the availability of individual insurance and lead to very high premiums. Thus coverage mandates and underwriting restrictions should be undertaken only after careful analysis of their impact on health insurance premiums and coverage rates. Although limited restrictions on underwriting practices and coverage mandates may incrementally increase the availability of more generous coverage, even these policies are likely to increase the average cost of health insurance, and thus to have some adverse effects on health insurance coverage rates.

An alternative to tighter regulation is to take steps to lower health insurance costs and thus encourage broader participation. Voluntary purchasing groups and association health plans, which allow individuals or small groups to band together to purchase insurance, are a promising approach. Supported by standards to ensure financial solvency and group membership based on factors other than health, these purchasing groups have the potential to achieve economies of scale in negotiating lower rates with participating insurers, and may be able to set up a competitive choice system that would otherwise be very difficult for individuals and small groups to manage. In addition, they may be able to reduce the relatively high fixed costs associated with enrolling a group. (Many of the administrative costs of health plans are largely independent of group size, whereas some costs, such as underwriting, are higher for smaller groups or for individuals.) Each purchasing group can also adopt strategies used by large employers to encourage competition and manage adverse selection.

Some local regions as well as some States such as California have set up and then privatized insurance purchasing cooperatives for small businesses. Many experts have suggested that States, which have considerable experience with competitive purchasing groups for their employees and (in a growing number of cases) for their Medicaid and SCHIP plans, would also be effective sponsors of individual purchasing groups. In addition, some private companies have set up voluntary programs for small agricultural groups, and many “affinity group” insurance plans are available for individuals: for example, many professional associations and college alumni associations offer insurance programs. The early experience of such groups in generating lower premiums through competition and economies of scale, and their effect on risk segmentation in health insurance markets, have been mixed. Some purchasing groups have been unable to obtain health insurance premiums that were significantly better than those available from independent insurance brokers. However, many group purchasing arrangements and association plans have attracted large enrollments and have been able to keep premiums stable and competitive without selectively excluding high-risk participants. Steps to encourage the development of purchasing groups, such as providing them the same exemptions from complex and variable State coverage mandates available to large employers while creating clear mechanisms to ensure solvency, are likely to make these options more widely available.

The market for individual health insurance would also be improved if the same kinds of subsidies that have worked well in employer group markets were available. As described in more detail below, subsidies such as a refundable tax credit would significantly lower premiums, thereby reducing adverse selection because a larger number of healthy individuals would take up coverage. In addition, 29 States have significantly improved the functioning of their individual and small-group markets by setting up high-risk pools. These pools provide the opportunity for hard-to-insure individuals to purchase subsidized coverage in a special purchasing group. Typically, the pools are funded by broad-based fees, for example an add-on to health insurance premiums or fees. The eligibility, subsidies, and funding mechanisms vary from State to State, contributing to differences in the stability of the pools, in their effect on health insurance costs for chronically ill people, and in their ability to address adverse selection problems in the State’s individual health insurance market.

Alternatively, innovative approaches by independent insurance brokers aimed at reducing the loading or transactions costs for individuals and small groups seeking insurance may also lower costs and expand participation. For example, online insurance “clearinghouses” allow small firms and individuals to obtain competitive rate quotes quickly from a large number of insurers. This improves price competition and can help reduce signup costs (for example, through a standardized online application procedure).

A further concern about competition in the health care system involves poor information. In addition to the problems of adverse selection already discussed, patients, providers, public policymakers, and taxpayers often have to make major decisions about medical treatments, regulations, and financing choices with only limited information. The obvious solution is to develop better information on treatments and on health system performance. Helping patients to understand their choices not only empowers them to choose the care they want but also leads to better decisions and, in some cases, reduced costs.

Finally, health care financing and regulation can and should reflect and reinforce the foundation of professional norms and ethics underlying the American health care system. Physicians, nurses, and other health professionals have a long tradition of caring deeply for patients and of working closely with them to provide the care that is in their best interests. Too often, however, these health professionals must work in a regulatory and economic environment that fails to encourage high-quality, efficient care. As these barriers are overcome, leading to fewer errors and more effective treatments, more Americans will find participation in health plans worthwhile. This important issue is addressed in the next section.

Increasing Health Insurance Coverage

Clearly, innovative approaches are needed now more than ever to help keep up-to-date health insurance available to workers and temporarily unemployed Americans and their families, and beyond that, to increase rates of health insurance coverage. To encourage such innovations, public policies should encourage a broad range of coverage options. Some of the most promising approaches to increasing coverage provide support for purchasing health insurance and health care services while easily adapting to changing circumstances and patient needs. Policy studies indicate that several principles are important:

- *Recognize existing support.* Tax exemptions for employer contributions to private health insurance are an important contributor to the stability of employer-sponsored health insurance plans. Although a concern is that unlimited tax exemptions may create an incentive to purchase very costly health care coverage, this form of subsidization does make health insurance more affordable for employees and contributes to very low rates of uninsurance—around 5 percent—for workers who are offered employer-sponsored coverage.
- *Focus new Federal support on those most likely to be uninsured.* Some groups currently receive little or no assistance with their health insurance costs. Most notably, workers who must purchase individual coverage because their employer does not offer health insurance

generally receive no tax subsidies for health insurance at all. Many small employers and employers of low-wage workers do not offer health insurance. This lack of subsidization is a major reason why individuals in families with incomes less than twice the poverty line have very high uninsurance rates, around 25 percent, and account for a majority of the uninsured. Researchers have found that unemployed workers are three times more likely than employed workers to be uninsured. Often these workers are eligible to continue their former employer's coverage temporarily through COBRA (or are covered under "mini-COBRA" laws in 38 States that expand COBRA to smaller employers), but usually they must pay the full cost of their insurance. (COBRA refers to provisions under the Consolidated Omnibus Budget Reconciliation Act of 1986.) Those ineligible for COBRA, and those whose former firm no longer exists or no longer offers health insurance, also receive no tax subsidies. Unemployed workers are likely to regain coverage on finding a new job and generally are not without insurance for long periods. Hence, temporary assistance for involuntarily unemployed workers would also be relatively likely to reduce uninsurance rates. In contrast, because insurance coverage rates are already high among the many workers with employer-based coverage, any new or expanded Federal assistance to them beyond existing tax subsidies would be more likely to crowd out existing private contributions. That is, such assistance might encourage workers who would otherwise have kept their private coverage to obtain coverage under the new Federal program instead, and thus save money even if the coverage is not as good. Such assistance might also decrease the incentive for employers to offer health benefits in the first place. New support would thus improve the incomes of the affected workers but would have a relatively modest effect on health insurance coverage.

- *Design any new assistance to maximize uptake by those without coverage.* Many uninsured Americans have little income tax liability and are likely to work in firms with other workers without substantial tax liability. Thus tax incentives that are valuable only to individuals and families with substantial income tax liabilities (such as income tax deductions) do little to encourage coverage. In contrast, refundable tax credits would provide valuable assistance. In addition, because many uninsured households have few liquid assets such as personal savings with which to pay health care bills, tax credits must generally be available at the time health insurance is actually purchased (that is, they should be "advanceable"). For the same reason, credits should not be subject to a significant risk of additional "reconciliation" payments at the end of the year.

- *Encourage a broad range of coverage options.* Minimum standards for coverage, such as protection against catastrophic health care expenses, are important both to ensure that the policy chosen actually covers the significant financial risks and to discourage inappropriate health plan strategies for risk selection. But the fact that many new approaches to delivering care are under development and becoming more widespread now means that specific mandates and restrictions on sources of coverage are especially likely to foreclose valuable innovations in health insurance, limit the attractiveness of available coverage options, and increase uninsurance.

As important as the goal of expanded health insurance coverage is, it is also important to remember that increasing health insurance coverage is a means to an end: effective medical treatment of all Americans, where the definition of “effective” depends importantly on the preferences and unique circumstances of each patient. As the next two sections describe in more detail, both public programs and private health insurance plans have considerable room for improvement in meeting this goal. Public policies should seek not only to increase health insurance coverage rates, but also to increase the value of health insurance that is provided, by promoting opportunities for individual choice and responsibility.

Innovative Tax Incentives for Increasing Private Health Insurance Coverage

A wide range of proposals focus on refundable, advanceable, nonreconcilable tax credits to reduce uninsurance rates. Refundable credits have the same dollar value regardless of taxable income. Advanceability means that the credit is available when eligible individuals are actually purchasing insurance; they need not wait for a refund until the following year when they file their tax return. Nonreconcilability means that, when the advance credit is awarded, eligible individuals need not worry about retroactively losing benefits at the end of the year, for example if their income turns out to be higher than expected.

Under the Administration’s proposed health insurance tax credit, which phases out with income, an individual’s income in the previous tax year would be used to determine eligibility for the advanceable credit. Those who qualify would receive certificates that could be used like cash to purchase coverage, so that the eligible individual need only pay the difference between the plan premium and the tax credit. Because the previous year’s income is already known, no eligible individual would be afraid to use the credit for fear of turning out to be ineligible because of too-high income at the end of the year. The refundability of the tax credit would augment the ability of lower and moderate-income individuals to purchase private health insurance,

giving them improved access to competing plans. The resulting broader participation in private health insurance markets would reduce pressures for adverse selection.

The Administration's tax credit would be available to people purchasing private health insurance coverage outside of plans offered by their employer or their spouse's employer. That is, working and unemployed people who do not already have tax-subsidized, employer-provided insurance would be eligible. Similar Congressional proposals would also make assistance available for purchasing COBRA coverage. These groups currently have the lowest takeup of available private coverage, because they are not currently subsidized. As a result, these proposals should achieve large net increases in coverage per dollar of program costs.

The generosity of the credit would also influence the cost-effectiveness of the expansion of coverage. A very generous credit would obviously induce more people to take up coverage but, depending on its design, might also draw more workers away from current employer coverage. The result would be a relatively expensive incentive with relatively less net effect on coverage. Recent studies of insurance markets and worker decisions about taking up coverage suggest that a capped credit of around \$1,000 for individuals and \$2,000 for families strikes a reasonable balance. A credit in that range would cover half or more of the cost of a reasonably comprehensive health insurance plan—one that provides preventive coverage and major-medical protection—for most of the uninsured, yet would not be so generous as to substantially crowd out employer-sponsored health insurance. Although many studies indicate that such a credit would provide enough of a subsidy to have a major impact on coverage, particularly for younger, healthier individuals, a potential problem is that it would cover a much lower percentage of the premium for individuals over 50 and those with chronic illnesses, for whom rates in the individual market are considerably higher. However, the additional policy steps described previously, such as additional subsidies through risk adjustment and high-risk pools, or expanded availability of voluntary purchasing groups, would help markets for non-employer-sponsored health insurance function better for these groups.

Some health policy experts and Members of Congress have proposed a broader based refundable tax credit—one that would also provide significant new subsidies to all workers with employer-provided coverage. Because so many workers have employer coverage already, however, a tax credit for employer coverage would have a far greater budgetary impact, and a much larger share of its costs would go toward existing rather than new health insurance coverage. To limit the additional budgetary costs, many experts have proposed a gradual transition from the current tax exemption to a system of tax subsidies for employer coverage that relies more on credits.

Although such a transition would probably encourage lower cost employer coverage and increase the takeup of employer coverage by lower income workers, it could have a significant impact on current employer plans, union negotiations, and other issues affecting worker compensation.

Clearly, the proposed tax credits would not cover the full costs of very generous, “first dollar” health insurance plans. Yet there are many reasons why such expensive coverage may not make good economic sense in any case. First, minimal copayments lead to moral hazard in health care spending: because the marginal cost to the patient of health care services is so low under such plans, a disconnect emerges between cost and value in health care decisions, contributing to rising health care costs and patient frustration. In the future, assuming that health care costs continue to rise rapidly, such policies will be even less sustainable. Second, reliance on minimal copayments in both private managed care and government health insurance plans has led to significant regulatory intrusions and price controls, which adversely affect doctor-patient decisionmaking. However well intentioned as an approach to limiting cost increases, such intrusions may make it more difficult for patients to get appropriate treatment.

On the other hand, many families do not have sufficient liquid assets to absorb even a few thousand dollars in health costs without sudden, major disruptions in their other household spending. To encourage saving for such contingencies, some innovative proposals have been developed. Some of these would help families set aside a “buffer” account to absorb such costs, for example by relaxing the carryover limitation on flexible spending accounts or the restrictions on medical savings accounts. Currently, many employers allow employees to set aside predetermined dollar amounts on a tax-free basis in such accounts to be used for health care or child care expenses. However, employees in these arrangements must spend all of their allocated dollars annually, and so cannot accumulate assets to be used in the event of a serious illness in the following year. This use-it-or-lose-it requirement contributes to unnecessary year-end medical spending. If at least some of the account balances could be rolled over to future years, workers could build up a rainy-day health account by making relatively painless, regular, tax-deferred contributions to interest-bearing accounts.

Such permanent flexible saving accounts would be similar to 401(k) retirement accounts, which have quite high rates of enrollment even among the lowest income eligible groups. The combination of flexible accounts with a tax credit or existing tax subsidies would make a reasonably priced health insurance policy very attractive—the premium would be relatively low, and the potential for some out-of-pocket spending would not be a deterrent to choosing such a plan. In fact, combinations of individual health accounts with insurance plans that provide protection against substantial expenses as

well as freedom from traditional restrictions on managed care coverage are now being offered by some employers, including the members of the Pacific Business Group on Health. But the absence of needed tax incentives may limit the attractiveness of these forms of insurance. For example, employee out-of-pocket spending in these innovative plans is not tax-deductible, and tax-favored contributions to flexible savings accounts cannot be rolled over from year to year. Expanding the availability of health accounts by addressing these concerns would reduce financial barriers to access while encouraging promising innovations in private health insurance.

Increasing Coverage in Public Health Insurance Programs: Medicaid and SCHIP

Public health insurance programs can also benefit from innovative approaches to expanding coverage. For example, even though SCHIP has encouraged most States to provide coverage for children in lower income families (those with incomes up to or approaching 200 percent of the poverty level), one-fifth of such children remain uninsured, compared with only 7 percent of children in families with incomes over 200 percent of the poverty line. Innovative expansions of public health insurance coverage for lower income households thus remain a high priority. Particularly needed are expansions that would make private health plans used by higher income families more affordable to the growing number of working families covered through these programs. In addition, employer-provided private health insurance coverage is much less widespread among lower income than among higher income households; therefore expansions of public health insurance coverage are less likely to crowd out existing coverage, leading to greater net reductions in the number of uninsured as spending in the government health insurance programs rises. (See Chapter 5 for further discussion of the crowding out of private programs.)

Many States have exercised options available under current law as well as implemented specific Medicaid and SCHIP “waivers” to cover the parents of eligible low-income children, because some evidence suggests that parents are more likely to take up coverage for their entire family than to enroll in children-only coverage. Some States have also implemented waivers to extend coverage to childless adults with low incomes, in the expectation that broader coverage for all low-income persons will strengthen the State’s health care infrastructure. However, efforts to expand coverage are impeded by the complex structure of Medicaid and SCHIP, which require States to deal with multiple funding streams and administrative requirements even to provide coverage for a single low-income family. In addition, Medicaid’s detailed and outdated statutory requirements mean that virtually all States must frequently go through the Federal waiver process to update their program.

Although dramatic progress has been made in clearing a backlog of plan amendments and waiver applications, resulting in eligibility being extended to 1.4 million additional individuals and coverage expanded for 4.1 million, a more promising approach would emphasize the flexibility of program design that has proved effective in SCHIP. This could be coupled with heightened but reasonable accountability requirements, to permit objective evaluations based on better evidence of whether State program changes that are intended to increase coverage and improve quality of care for program beneficiaries actually achieve their goals.

Finally, many States are now providing coverage under Medicaid and SCHIP through competing private insurance plans, suggesting that the combination of public funding and competitive private provision of health insurance coverage is an effective strategy for encouraging innovation in health care delivery for low-income populations while controlling costs. This topic is covered in more detail in Chapter 5.

A Coordinated Safety Net for the Uninsured: Funding for Community Health Centers

Even with expanded subsidies for private and public insurance, most research predicts that a substantial share of currently uninsured Americans would remain uninsured. For this reason, and because proposals to expand health insurance coverage will take some time to implement, the Administration has also developed initiatives to improve the availability and coordination of medical services for those without coverage. This has been done by increasing the flexibility of State and local governments to provide access for low-income residents through integrated community health center (CHC) programs. The mission of CHCs is to provide care to underserved populations, including populations that have proved difficult to reach through private or public insurance. To accomplish this, local CHCs have developed innovative approaches that build on unique community features and resources, and have collaborated with other public, private, and academic programs.

For example, the Centers for Medicare and Medicaid Services (the agency formerly known as the Health Care Financing Administration) have partnered with the Institute for Healthcare Improvement (a nonprofit organization) and with specific CHCs around the Nation to improve health care for low-income individuals with chronic illnesses such as diabetes, asthma, and cardiovascular disease. The Clinica Campesina Family Health Centers in Lafayette, Colorado, the Lawndale Christian Health Center in Chicago, and CareSouth Carolina have developed programs adapted to their populations and have achieved measurable improvements in diabetes care—including the patient self-management efforts so central to successful treatment of chronic illnesses.

CHCs have also developed innovative approaches through community partnerships and collaborative funding strategies. For example, Grace Hill Neighborhood Health Centers in St. Louis provide services in two public housing projects and to the homeless in 16 sites through a combination of Federal funding as a CHC, special Federal expansion funds, and contracts with the city, the county, and other CHCs. Grace Hill has also developed vital information management systems, including registries of individuals with chronic illnesses, relevant tracking reports to providers, and automatic reminders to patients of needed preventive and follow-up tests. Because of their community roots and their ability to focus on the distinctive needs of their patient population, CHCs can provide a quality of care that rises well above what might be implied by the term “safety net.”

Making Medicare Coverage More Flexible and Efficient

One of the most obvious examples of the difficulty of keeping up to date with innovations in health care delivery is the Medicare program's lack of a prescription drug benefit. More than one-quarter of Medicare beneficiaries have no prescription drug insurance at all, despite the fact that diseases are increasingly being treated with drugs rather than through hospital or clinic care. This lack of prescription drug benefits among Medicare enrollees has had adverse health consequences. In one study the use of cholesterol-lowering drugs, an essential component of care for many individuals with coronary heart disease, was 27 percent for appropriate elderly Medicare enrollees with supplemental, employer-provided plans providing drug coverage, but only 4 percent for those with no drug coverage at all. Innovative drug use for the treatment of ulcers costs \$500 per patient but can save as much as \$28,000 by avoiding the need for a prolonged hospitalization.

Lack of prescription drug coverage is only one element of the undesirable economic effects of Medicare's outdated coverage. As health care capabilities have risen over time, the benefits and the costs of changes in treatment have been particularly great for seniors and persons with disabilities. But because Medicare benefits have not kept pace, Medicare beneficiaries spend on average over \$3,100 a year out of pocket on major medical care, and this spending is rising much faster than inflation. Medicare beneficiaries also face a significantly higher risk than other insured groups of very high out-of-pocket expenses.

Because beneficiaries have inadequate options for making this spending more predictable, they can find it very difficult to budget their often-fixed retirement income effectively. Much of the private prescription drug coverage available to seniors today includes spending caps, and many seniors do not

have the opportunity to purchase prescription drug coverage that protects them from high drug expenses at a reasonable premium. Moreover, seniors without good drug coverage are much more likely to pay full retail prices for medications, in contrast to the significantly lower prices available from manufacturer rebates and pharmacy discounts to virtually all other Americans with modern health insurance. Even for covered benefits, supplemental private “Medigap” insurance that fills in substantial copayments and coverage limits is virtually essential, because Medicare includes no stop-loss protection, and the copayments are large. For example, the copayment required for a hospital episode is over \$800, and that for many major outpatient procedures is almost \$100. Physician services generally have copayments of 20 percent. Fewer than half of all seniors obtain coverage through Medicaid or a supplemental insurance policy offered by a past employer as a retirement benefit. Because of these coverage gaps, one-quarter of beneficiaries purchase individual Medigap plans, which must conform to standards developed over a decade ago that require first-dollar coverage in order to get reasonably complete protection against high expenses. Consequently, premiums for individual Medigap policies are substantial, accounting for a significantly larger share of the out-of-pocket expenses of the average Medicare beneficiary than prescription drugs, and they have been increasing rapidly: premiums for the most popular standardized Medigap plans rose more than 20 percent between 1997 and 2000. In addition to being costly for seniors, such first-dollar coverage results in billions of dollars of additional utilization in the Medicare program each year.

The coverage gaps in Medicare’s required benefit package, and the rising cost of the supplemental coverage that is essential to fill those gaps, are among the reasons why many Medicare beneficiaries prefer private insurance plans. Such plans, which can compete for beneficiaries through the Medicare+Choice program, typically have been able to offer more comprehensive coverage, including prescription drugs, for far less than the combined Medicare plus Medigap premiums that beneficiaries must pay in the traditional, government-run Medicare plan. (These premiums now exceed \$150 a month and are often much higher.) However, after several years of rapid growth, enrollment in private plans has begun to drop significantly. An important contributing factor is the “minimum update” for private health plan payments imposed by the Balanced Budget Act beginning in 1998 for most areas in the country with high private plan enrollment. Because the payment updates are now limited to 2 percent a year at a time when private health insurance and Medicare costs are growing much more rapidly, Medicare’s contributions to private plan premiums in these areas are diverging from the costs of providing coverage. Poor prospects for reimbursement, coupled with the Medicare+Choice program’s substantial

regulatory burdens and the requirement that the private plans provide coverage that actuarially meets or exceeds Medicare's unique and uneven benefit structure, have led a number of private plans to pull out of the program. Those that remain have instituted substantial increases in premiums and copayments. Meanwhile the options that have proved most popular with nonelderly Americans—preferred provider plans and point-of-service plans, which provide a balance between the savings possible in tight managed care networks and the flexibility of treatment options in broader indemnity plans—are virtually nonexistent in Medicare. As a result, Medicare beneficiaries are headed toward having few options beyond a single outdated benefit package, at a time when the Medicare program desperately needs innovation in coverage to improve quality and reduce costs.

By contrast, employees of many private firms and of the Federal and State governments, as well as many Medicaid and SCHIP beneficiaries, are able to choose from a variety of health plans that offer a range of options in terms of breadth of coverage networks and out-of-pocket payments. In turn, competitive choice provides incentives for health plans to reduce costs and adopt innovations in benefits or in health care delivery that beneficiaries find worthwhile. For example, the Federal Employees Health Benefits (FEHB) program has long offered a range of reliable choices to all Federal employees in the country, a work force with diverse health needs and circumstances that has participants in virtually every urban and rural zip code nationwide (Box 4-3). FEHB has accomplished this by providing a level of support for premiums that is tied to the average cost of the plans chosen by employees. Employees can reduce their health care costs if they choose a less expensive plan, because a portion of the plan's cost savings is passed on in the form of lower premiums. Conversely, much of the additional cost of more expensive plans is also passed on, so that employees who choose a more costly plan face correspondingly higher premiums. All participating plans must meet the FEHB benefit standards and must provide information to beneficiaries about coverage networks and performance on a growing set of quality measures.

Analogous proposals have been developed in recent years for improving Medicare's coverage options, building on the proposals considered by the National Bipartisan Commission on the Future of Medicare in 1999, the criticisms of those proposals, and subsequent ideas from members of both political parties. One key concept in these recent proposals is that of preserving Medicare's promise of a defined set of benefits while encouraging competition between the traditional Medicare plan and private health plans in how those benefits are provided. As in the FEHB system, beneficiaries would pay more for plans that used a more costly approach to provide Medicare's required benefits, and would pay less for plans that adopted a less costly approach.

Box 4-3. Federal Employee Health Insurance Plans

The Federal Employees Health Benefits program covers 9 million Federal civilian employees and their dependents. The program allows employees to choose from a menu of plans, including 11 fee-for-service plans that are available to Federal employees in any part of the country. Employees in most areas also have the option of enrolling in a managed care plan such as a health maintenance organization or a point-of-service plan. For example, Federal workers in the Washington, D.C., area have a menu of 7 different managed care plans from which to choose in addition to the 11 nationally available fee-for-service plans.

Plans are required to offer a package of minimum benefits but may differ with respect to the generosity of copayments, deductibles, and other benefits. The government pays about two-thirds of the average cost of coverage, with workers contributing the rest. Since 1999 the government's share has been calculated using a "fair share" formula that maintains a consistent contribution from the government regardless of the plan chosen, so that the employee bears the marginal cost of choosing a more generous plan. Workers who prefer generous benefits are free to choose them, while workers who choose more cost-conscious plans benefit from their lower cost.

The FEHB program provides a wide variety of coverage choices to accommodate the preferences of a large work force that is diverse both geographically and in terms of its health care needs. At the same time, FEHB plans as a whole have experienced stable premium growth that ensures that the program will remain on a sound financial footing. The experience of the FEHB program shows how empowering consumers to make insurance choices can result in coverage that is both secure and flexible.

Some critics of the commission's proposal have argued that any such reforms would force seniors into private plans, because the cost of the traditional Medicare plan would be higher. But that is not necessarily true. For example, the so-called Breaux-Frist II proposal could not lead to higher premiums than under current law in the traditional Medicare plan. This is because the traditional plan premium would continue to be determined as it is now, but beneficiaries would face lower premiums if they chose a private plan with lower costs than the traditional plan, and would face higher premiums if they chose a private plan with higher costs.

Obviously, the Breaux-Frist II approach would work best in areas where the traditional plan is the dominant plan. In areas where a large share of

beneficiaries have enrolled in private plans, and where performance measures indicate that these beneficiaries are receiving at least as good care as those in traditional Medicare, using the traditional plan or any particular nonrepresentative plan as the reference point for Medicare's support for beneficiary premiums would be both inappropriate and potentially costly for the government or for beneficiaries. Instead, the FEHB approach of tying the government's support for health insurance costs to the average cost of the plans that beneficiaries actually choose is a better way of ensuring that savings from providing Medicare's defined set of benefits accrue to both beneficiaries and taxpayers.

Last year the President proposed a framework that would provide Medicare beneficiaries with better health insurance options, similar to those available to Federal employees. Under this proposal, plans would be allowed to bid to provide Medicare's required benefits at a competitive price. Beneficiaries who elect a less costly option would be able to keep most of the savings, so that some beneficiaries might pay no premium at all. Moreover, the President proposed using the savings from greater efficiency in providing Medicare's current benefits to support further benefit improvements, including better coverage for preventive care and stop-loss protection. The President proposed to implement these benefit improvements while retaining the option for current and near-retirees to stay in the current Medicare system with no changes in benefits if they prefer it.

In addition to providing reliable, modern health plan options and better benefits for Medicare beneficiaries, the Administration has proposed a subsidized prescription drug benefit in the context of Medicare modernization, to help protect seniors from high drug expenses and to give those with limited means additional assistance to pay for needed medications. Both Democrats and Republicans generally agree that any new drug benefit in the traditional plan should not adopt the traditional approach to delivering care, that is, direct fee-for-service government provision with complex coverage rules and price controls. There is broad agreement that such a bureaucratic approach would significantly reduce the availability of innovative drug therapy for seniors. Instead the drug benefit should give all seniors the opportunity to choose among plans that use some or all of the tools widely utilized in private pharmacy plans to lower drug costs and improve the quality of care—tools that include competitive formularies to generate lower manufacturer prices, pharmacy counseling, prescription monitoring, and disease management programs.

The Administration has also proposed a Medicare-endorsed prescription drug card plan that would provide immediate assistance to beneficiaries without drug coverage. The drug card plan would not be a drug benefit, nor would it be intended as a substitute for one. Instead it would provide access

to pharmacy programs that use private sector tools like those just mentioned to reduce drug costs and to improve the quality of the pharmacy services available to beneficiaries. The drug discount card would be a step toward an effective, competitive prescription drug benefit under Medicare by giving both beneficiaries and the Medicare program some much-needed direct experience with the private sector tools that are widely used in prescription drug benefit plans today. It would also provide immediate assistance to beneficiaries in obtaining lower cost prescriptions until the drug benefit is implemented.

Better Support for High-Quality, Efficient Care

Our current system of financing and regulating health care providers is not geared toward recognizing and rewarding high-quality, efficient care. For example, when poor surgical protocols result in infection, readmissions, and additional surgical work, Medicare pays more, not less, to the hospital and health care providers responsible. In contrast, some private payers have begun to pay higher quality providers more, and one can envision further reforms in this direction, while still using risk adjustment and the other tools described in the previous section to reward appropriate care for patients with more complex health problems.

This section highlights some of the clear opportunities to improve the quality of health care, as well as the promising public and private initiatives that have begun to do so. Recent private sector initiatives have encouraged hospitals to improve patient safety through the use of computerized record-keeping and other measures, efforts that should be reinforced at the Federal level. Government support for research and provision of information to health care providers about the quality of their care, and about pathways to improving care, is another element in improving the health care system. Reforming the legal system so that it encourages rather than discourages collaboration and sharing of information among health providers is also a key building block in improving the quality of clinical care.

Shortfalls in the Quality of Care

Two influential reports from the Institute of Medicine have called attention to the serious problem of medical errors. The Institute estimated that as many as 50,000 to 100,000 deaths each year may be attributable to medical errors; even if these estimates are too high, as some analysts have suggested, many avoidable deaths do occur. However, improving quality is more than

the reduction of errors, or *misuse* of treatments. In the terminology of the Institute of Medicine reports, the sources of poor quality include both the *underuse* of procedures or treatments whose effectiveness has been demonstrated, and the *overuse* of treatments with unclear or harmful effects.

Many procedures or diagnoses are widely understood to provide benefits to nearly every person who receives them, yet are underused in practice. Examples include screening for breast and colorectal cancer in high-risk populations, annual blood tests for people with diabetes, and the use of aspirin and, when appropriate, beta blocker drugs for patients with recent heart attacks. One study of Medicare recipients, in 1997, found that fewer than two-thirds of patients who had experienced a heart attack and had no contraindications to beta blockers were taking them on discharge from the hospital. In some States that rate of use was as low as 30 percent. A similar study indicated that many Americans who could benefit from the newly developed cholesterol-lowering drugs do not receive them. Indeed, failure to use effective treatments has been estimated to result in 18,000 avoidable early deaths among heart attack patients in a year.

Whereas some procedures are underused, others are overused. One-fifth of all antibiotics prescribed in 1992 (12 million prescriptions) were used to treat common colds and other viral respiratory tract infections, despite the ineffectiveness (and potential long-run harm) of antibiotics for such illnesses. A study of coronary angioplasty concluded that the procedure was clearly medically appropriate in fewer than one-third of cases; the remainder were either of uncertain benefit (54 percent) or inappropriate (14 percent). Despite important technological advances in imaging methods for the detection of appendicitis (such as computerized tomography and ultrasonography), one recent study showed no improvement in rates of unnecessary surgery.

Reducing overuse of procedures is clearly beneficial for taxpayers, who save money, and for patients, who avoid unnecessary interventions and their resulting side effects. The potential savings from this reform are substantial. One estimate suggests that as much as 20 percent of the Medicare budget could be saved by reducing the overuse of care, particularly among patients with long-term chronic illnesses. Although such savings might be offset by increased use of valuable, underutilized interventions, the net effect of these improvements in care would be much better value for the health care dollar.

Health care costs are also increased by the misuse of treatments. For example, a patient undergoing surgery may receive the wrong medication, and as a result experience complications that result in longer illness, permanent disability, or death. One study estimated that as many as 27,000 avoidable deaths each year are due to the misuse of medications. Such errors are probably most common among seniors, who take many more prescription drugs than other insured Americans but are less likely to have

prescription drug coverage that assists them with medication management. Even technological advances can be undone by low-technology failures related to poorly coordinated care, inadequate follow-up, and resulting incomplete recovery. Investing in methods to reduce medical errors would reduce suffering, disability, and death—and the associated costs.

Disparities in the Health Care System

Not everyone with a given disease receives the same level of care. The quality problems discussed above may be greater for low-income and minority populations. For example, among women covered by Medicare, 74 percent of white women living in high-income areas received influenza immunizations, whereas only 51 percent of African American women living in low-income areas did. Rates of surgery for heart attacks are lower among African Americans than among whites, although there is substantial controversy about the causes of such differences. Indeed, one recent study showed that overuse of this surgery—that is, its inappropriate use in cases where the risks outweigh the potential benefits—was actually higher among whites than African Americans.

These differences in utilization and quality across large geographic areas have been documented in other cases as well. A recent study showed a remarkable degree of variation across States—from 44 to 80 percent—in the appropriate use of an effective pharmaceutical treatment (beta blockers) for patients who have had heart attacks. There are also wide differences across regions with regard to overall spending and utilization (Box 4-4). It is intriguing that areas with the highest levels of health care expenditure per capita are not necessarily those with the best measured quality of care. In other words, improving quality does not necessarily result in higher Medicare expenditure. Many cities in the United States experience relatively high quality and low costs.

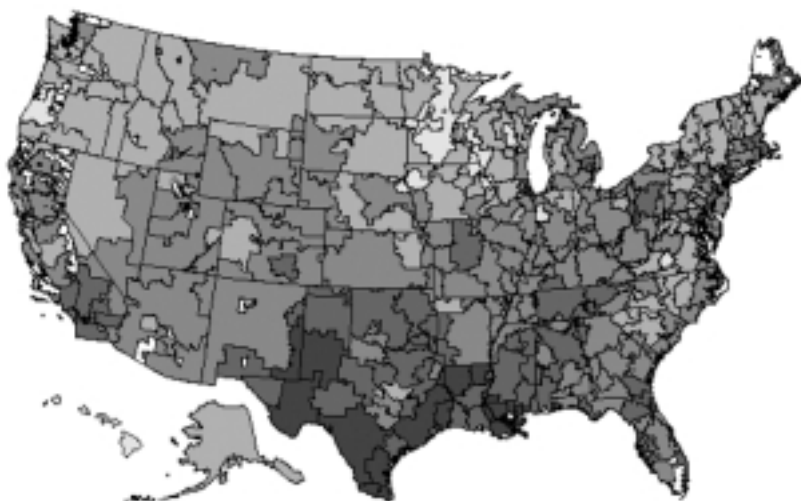
The prescription for reducing disparities is clear in the case of overuse and underuse of health care. Better quality care means encouraging much more utilization of services that are often not used in patients for whom they are clearly beneficial—and this holds true for all races, both sexes, and all regions. Better quality care also means moving toward zero utilization rates for inappropriate, procedures that have no documented benefits for any race or either sex. Where there are a range of reasonable treatment options, patient preferences are particularly important; for example, in the treatment of prostate cancer in men or breast cancer in women, the “right” level of care should depend heavily on those preferences. The reforms in health care coverage described in the previous section would help create an environment that rewards valuable innovations in communicating the benefits, risks, and costs of treatment options to patients to help guide their decisions.

Box 4-4. The Puzzle of Geographic Variations in Medicare Expenditure

Despite the Federal nature of the Medicare program, there are remarkable geographic differences in the level of Medicare expenditure per capita. The *Dartmouth Atlas of Healthcare*, using Medicare claims data under an agreement with the Centers for Medicare and Medicaid Services, has documented net spending per capita in 1996 among Medicare enrollees in 306 separate areas of the United States. Even after correcting for differences in age, sex, and racial composition, spending per capita differs widely, ranging from \$7,800 in Miami to only \$3,700 in Minneapolis. Only a small part of these differences can be explained by variations in underlying illness levels.

The map below, reprinted from the atlas, shows the corrected patterns of geographical variation in spending. The darkest areas are those where spending per capita ranges from \$5,698 to \$8,862, and the lightest areas those where the range is from \$3,117 to \$4,178. (Some areas are inhabited by too few seniors to allow spending to be measured accurately.)

The disparities in health care utilization highlighted here translate into large disparities in Medicare benefits across regions and States. One study showed that average lifetime Medicare expenditure for a typical 65-year-old may differ by as much as \$50,000 depending on the State of residence. At the same time, quality of care appears to be similar in low- and high-utilization regions. These differences suggest that better information on the effectiveness of different styles of medical practice, possibly coupled with better incentives to encourage efficient care, could result in substantial cost savings for Medicare without any adverse consequences for patient health.



Source: *Dartmouth Atlas of Healthcare* © 1999. Reproduced with permission.

Empowering Providers to Improve Quality of Care

Improving quality saves lives and can save money. No one disagrees with the objective of improved quality; the problem is creating an environment for medical practice that gets results. A variety of new and innovative approaches developed at both the local and the Federal level hold the promise of improving how care is delivered. (Many of these are described in the recent Institute of Medicine reports on quality of care.)

A number of private sector quality initiatives have involved aspects of health care where success can be measured objectively. For example, a collaborative quality improvement program for the intensive care unit at LDS Hospital in Salt Lake City, Utah, improved outcomes for its patients while also lowering costs by almost 30 percent. Similarly, the Northern New England Cardiovascular Disease Study Group developed a working group that enabled cardiac surgeons to reduce the complications of surgery at each stage of the procedure and to reduce postoperative mortality by 24 percent. Each of these successful programs set the goal of studying well-defined interventions in specific populations, using clear, objective measures of success. Initiatives are currently under way to develop evidence on the overall benefits of implementing quality improvement measures across an entire hospital system.

All of these efforts, and many others around the country, have gotten off the ground as a result of provider initiatives in the face of many institutional, regulatory, and financial obstacles. An enormous amount of research, including the series of studies by the Institute of Medicine, has concluded that high-quality care can best be achieved in an environment that emphasizes and rewards continuous quality improvement. The complexity of health care delivery means that there are generally tremendous opportunities to improve the coordination of care, reduce communication problems, and eliminate many avoidable mistakes and complications that occur despite the best of provider and patient intentions. Most of these quality improvement opportunities are “low-tech”: problems that are not so hard to solve technically, if health care providers can openly discuss and work together to respond to the root causes of errors, near-misses, and concerns expressed by patients and colleagues. Applying the lessons learned from many other highly complex technical systems, such as nuclear reactors, is a promising direction for reducing health care errors.

The growing evidence on quality improvements indicates that hospitals and doctors would undoubtedly benefit from such local, collaborative efforts to improve quality. But there are many obstacles to success today. Under the current system of medical liability, this type of open discussion is widely viewed as carrying substantial financial risks of malpractice exposure. Leading analysts of quality improvement have called for modifications in

medical liability laws so that the collection and sharing of information to avoid errors and improve quality are not impeded. Another obstacle is financial: under fee-for-service systems like those used in Medicare and many State Medicaid programs, providers that improve quality receive less reimbursement, because follow-up visits and admissions for complications are fewer.

As noted previously, research on how medical treatments can be used more safely and effectively in a wide variety of actual medical practice settings is an important element of the Federal Government's biomedical research portfolio. In addition, many Federal programs, activities, and laws can support providers who want to work together to improve care. Today the Medicare quality improvement organizations (QIOs, formerly known as peer review organizations) provide some important but limited support for efforts by local groups of hospitals, physicians, and some other providers to identify, assess, and improve certain aspects of health care quality. QIOs provide some protection from malpractice liability for their quality improvement activities. But liability protections should be broadened to include new information generated beyond the standard medical and administrative records, through quality and safety improvement activities, whether or not they are actively sponsored by QIOs.

The Administration is also developing regulatory standards for health care information systems, to implement legislation on administrative, clinical, and privacy standards enacted by Congress in the Health Insurance Portability and Accountability Act. These standards have the potential to improve health care quality, because consistent and up-to-date information standards, coupled with privacy rules that inspire patient confidence, will lead to more effective use of health care information. Health care providers will incur significant costs to come into compliance with the regulations. However, well-designed and timely standards can provide the lead time and guidance required to minimize compliance costs. Indeed, many health care providers have for years faced disincentives to upgrade their information systems until the content of the regulations becomes clear.

Empowering Patients to Make Informed Health Care Choices

As noted above, encouraging high-quality, efficient care requires meaningful and reliable choices of health plans and providers for well-informed patients. Within health plans, information about alternatives is increasingly important for helping patients work with their providers to make the best possible choices about specific illnesses such as heart disease, breast cancer, back pain, and prostate cancer. Researchers are beginning to understand the central role that patient preferences and choices can play in improved and cost-effective care of chronic illnesses, including late life care

decisions. Research is also leading to better and more reliable measures of the quality of health plans and providers, in terms of both clinical processes and outcomes of care as well as overall satisfaction.

Informed Decisionmaking: Better Choices, Higher Value Care

Many diseases have no single “best” cure or treatment. Instead there are a variety of ways to treat the disease, each with associated risks, benefits, and costs. For example, women with breast cancer often face the choice of mastectomy or a combination of breast-sparing surgery followed by radiation therapy. Both options carry similar implications for survival for many patients. But each has quite different implications for the patient in terms of physical impact and the duration of treatment required, and many patients have strong preferences about how they want to be treated.

Prostate cancer provides another example. There are tradeoffs regarding screening for prostate cancer using the current prostate-specific antigen (PSA) tests. Because the cancer grows so slowly, with as much as a 10-year lag between detection and clinical importance, the use of PSA tests among older men, who are likely to die of a different cause, should depend on the patient’s preferences, weighing his concern about the unpredictable course of the cancer against the unfortunate side effects of treatment, such as incontinence and impotence. These are decisions that the physician cannot make alone.

Many health care providers are implementing changes to enhance the ability of patients to participate in clinical decisions. At the Spine Center of the Dartmouth Hitchcock Medical Center in Lebanon, New Hampshire, patients with lower back pain fill out computerized evaluation forms regarding their goals and preferences when they arrive, so that the staff is prepared to address their concerns regarding treatment for their spine-related illness. The risks and benefits of treatment options, including surgery, are explained using a video featuring summaries of the clinical evidence as well as balanced discussions by patients who have experienced each of the different options. Following the implementation of this informed decisionmaking approach, surgical rates for herniated discs fell by 30 percent, whereas those for spinal stenosis (the squeezing of nerves emanating from the spinal cord) rose by 10 percent. These changes in surgical rates move in the direction indicated in the medical literature, which suggests that the former procedure is overused and the latter underused. Thus the program appears to have provided patients with quality information to assist them in making educated decisions, thereby improving their well-being while reducing overall costs.

This patient-centered approach to evaluating health care outcomes also provides a valuable framework for judging differences in treatment rates by race or sex for specific “preference sensitive” diseases. The important message is not that treatment choices should be the same across all subgroups of the

population. Rather, when several alternative treatments are available, patient preferences (rather than race or geography) should govern choices. For example, preferences for elective hip and knee surgery vary by sex, even among patients for whom the treatment is deemed medically appropriate. Less is known about differences in preferences by racial identity, although differences in preferences between whites and African Americans regarding end-of-life care have been noted.

Better Public Information on the Performance of Health Care Providers

A growing number of private health care purchasers are supporting informed decisionmaking by their employees by making measures of quality available on their health plan choices and, in some cases, on particular health care providers. These include clinical measures of plan performance such as those now widely used by the National Commission on Quality Assurance (for example, rates of appropriate treatment for diabetes and immunization rates) as well as patient-focused measures such as those developed by the Foundation for Accountability (FACCT). The Federal Government also has a particularly important role to play through supporting the development of appropriate information to help patients and providers identify and reward high-quality care. The Medicare, Medicaid, and Federal employee insurance systems hold information on literally millions of health care subscribers who are among the heaviest users of the health care system. With appropriate privacy protections, clinical studies using the data systems of these very large health insurance programs could augment data from private payers, allowing the construction of more comprehensive and accurate measures of plan quality, and potentially of provider quality as well. Indeed, the Federal Government has collaborated with private organizations in the development and use of patient satisfaction measures (Consumer Assessment of Health Plans, or CAHPS, measures). It is also a key player in the National Quality Forum, a public-private approach to endorsing reportable quality measures that are supported by experts, consumers, and other major stakeholders.

The process of identifying appropriate measures for public reporting is a difficult yet important one, because the measures endorsed must be valid indicators of quality if they are to encourage better health care decisions. Because patients are not allocated randomly to health plans or providers, measures are potentially biased by differences in case mix and may thus require adjustment for risk, so that they truly reflect differences in performance rather than differences in the health of the patient groups treated. In addition, medical information systems are imperfect, and some quality measures may not be captured adequately. Finally, because many important medical outcomes (including death following surgery) are relatively rare

events, some measures may incorrectly attribute bad luck to poor quality care. (For a more detailed discussion of performance measurement issues, see Chapter 5.) Quality measures that are themselves of poor quality may be worse than no measures, if they discourage providers from taking difficult cases or if they can be manipulated to improve measured performance. Thus, many quality and safety measures are better used on a confidential basis, as part of the internal quality improvement programs described in the previous section. As measurement methods and data systems have improved, however, a growing number of quality measures have been developed and are becoming widely used for public reporting by employers, States, and the Federal Government.

In addition, as mentioned above, some private purchasers now reward better measured performance with higher reimbursement, at least to a limited extent. Some insurers and purchasers include an incentive payment for achieving high scores on certain validated quality measures. Others have begun to use quality measures to influence their selective contracting with providers. For example, the Leapfrog Group, a consortium of more than 80 Fortune 500 corporations and other large institutions, has developed guidelines for contracting with hospitals by establishing a growing set of specific performance standards. The initial recommended measures for contracting include high numbers of certain surgical procedures (because hospitals that perform a higher volume of many complex procedures achieve better results), the use of computerized recordkeeping (because computerization helps reduce medical errors and misuse of care), and the direction of intensive care units by physicians specializing in intensive care.

Fulfilling the Promise of Medical Research

Developing an economic and institutional environment that encourages continued technological advances is a critical goal for the coming decades. As part of this environment, direct Federal support for an increasingly broad range of biomedical and related research is essential. The value of this research is evident in the medical progress witnessed over the past several decades. In large part because of active support by the National Institutes of Health and other Federal agencies, biomedical knowledge has grown rapidly, encompassing dramatic advances in understanding basic biological processes, identifying the pathology of specific diseases, and developing effective treatments. The decoding of human genome through public and private support is but one recent example of pioneering research that will lead to innovative prevention and treatment approaches.

The Benefits of Biomedical Research

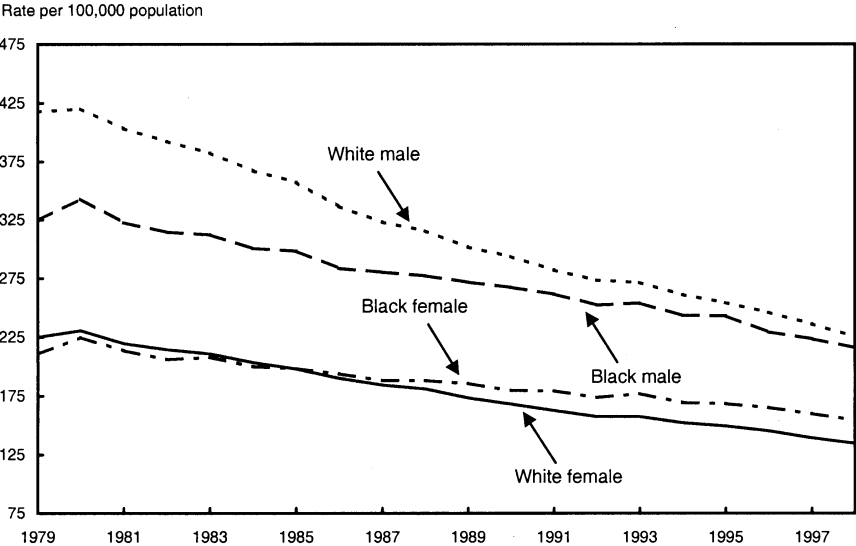
The past several decades have seen remarkable gains in longevity and reductions in disability. One of the most striking examples of technological progress in the treatment of illness is that for coronary heart disease (CHD). Since 1970, mortality from CHD has been declining between 2 and 4 percentage points a year on average, with overall rates falling by about 40 percent since 1980 (Chart 4-3). Although primary prevention has been an important contributor, most advances in cardiovascular health care are due either to innovations in mechanical treatments to improve blood flow to the heart (such as bypass surgery, newer and less invasive angioplasty procedures, and special wire stents to help hold diseased vessels open) or to pharmacological treatments (such as beta blockers and antihypertensive drugs to reduce the heart's work load, and thrombolytic "clot busters" to open up blocked vessels during a heart attack).

These improvements have not come without cost, which raises the critical question, in light of generally rising expenditure on medical care, of whether the increased costs are worth it. The answer, at least in the case of heart attacks, appears to be yes. One recent study concluded that the improvements in survival after a heart attack more than compensated for the increased financial costs. In this case, the money was well spent. Even though annual expenditure on cholesterol-lowering drugs is well into the billions of dollars, they have been proved to be highly cost-effective for many patients and have contributed to the improved life expectancy and better functioning of Americans today.

Such examples are not limited to heart disease. Chart 4-4 displays the rapid improvement in 3-year survival rates following the onset of an opportunistic infection signaling AIDS infection. Even though the new treatments developed to prevent AIDS complications are quite costly and have many side effects, these survival improvements suggest they are well worth the cost. As another example, new medications for depression have similar efficacy with fewer side effects, resulting in better adherence to treatment, better real-world effectiveness, and a reduction in the net cost of a remission. In addition, the availability and ease of use of these medications have contributed to a doubling in the rate of treatment of depression, increasing the economic benefits. Medical advances are doing more than just keeping increasingly frail elderly people alive: a recent study suggests that rates of disability among the elderly population have actually declined in recent years, probably because of avoided complications and better supportive care for chronic illnesses. We should remain aware of the distinction between long life and long, healthy life, but for the present, advances in medical technology seem to be accomplishing both.

Chart 4-3 **Mortality Rates for Coronary Heart Disease**

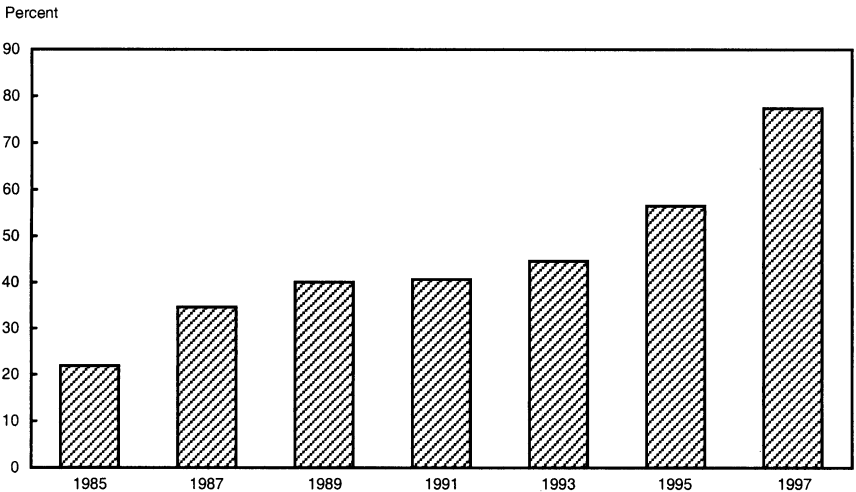
Age-adjusted mortality from coronary heart disease declined steadily for all segments of the population during the past two decades.



Source: Department of Health and Human Services (National Center for Health Statistics).

Chart 4-4 **Survival Rate After AIDS-Defining Infection**

Two-year survival rates after an AIDS-defining infection improved markedly over the 1985-97 period, especially in the 1990s.



Source: Lisa M. Lee, et al, "Survival After AIDS Diagnosis in Adolescents and Adults During the Treatment Era, United States, 1984-1997," *Journal of the American Medical Association*, 2001. Reproduced with permission of the author.

These studies are part of a growing body of evidence that, for a wide range of diseases, the additional money spent on treatment is more than offset by savings in direct and indirect costs of the illnesses themselves. Indirect costs include lost productivity and, especially, poor health, which people are clearly willing to pay to avoid. Stated differently, because the quality-adjusted cost of treating many diseases has fallen, health care has become more productive over time, even as absolute costs are rising with greater use of more intensive treatments.

Many Unanswered Questions About Existing Medical Treatments

Although these gains are impressive, there is still much to learn. Cardiovascular disease is the success story of modern medicine: a plethora of articles have demonstrated the value of different treatments compared either in isolation (drug treatment versus invasive cardiac surgery, for example) or in combination. Thus conclusions about rising productivity for cardiovascular care are the best documented, with literally thousands of clinical trials and epidemiological studies. Yet even in this area, substantial opportunities for further productivity improvements appear to exist. For example, in one recent study a large share of the treatments for coronary artery disease performed were judged to be of uncertain value based on medical expert reviews. Other examples of opportunities to improve the quality of cardiovascular care were discussed in the previous section. The situation is even cloudier in the treatment of other chronic diseases, where the evidence-based science is much sparser; here physicians have a less extensive knowledge base to draw upon. For example, on chronic lower back pain—an extremely common condition—no evidence is yet available from large randomized trials on the benefits of surgery versus medical management and supportive care, although one trial is currently under way. It is also more difficult to determine the effectiveness of many screening and preventive treatments. Better diagnostic methods often result in the identification of earlier or less severe illness that would have been overlooked before. Thus when previously “subclinical” cases with relatively good outcomes are added to the population diagnosed with the illness, survival rates may appear to improve, even if treatment methods have not (Box 4-5). In addition, clinical trials of preventive treatments are often prohibitively expensive, because they require very large enrolled populations and take many years for effects to be detected with confidence.

Furthermore, the effectiveness of specific treatments often varies substantially across population subgroups. For example, it is just now being understood that the effectiveness of cholesterol-lowering drugs depends significantly on the characteristics of the patient. As we develop a clearer

Box 4-5. Survival Rates and Mortality Rates

Survival rates for breast cancer have risen dramatically. Whereas in 1950-54 the 5-year survival rate was only 60 percent, by 1989-95 it had risen to 86 percent. This improvement is in part the result of important technological innovations in the treatment of breast cancer; nonetheless, these 5-year survival rates probably overstate the actual gains. The reason is that the detection of breast cancer has also improved dramatically: current technology is able to detect much smaller nodes than could be identified before, which may or may not develop into cancerous sites. Thus, improved 5-year survival rates reflect several phenomena. First, more women are being diagnosed, some of whom might not have developed clinically significant cancer during their lifetime. Second, more diagnoses are occurring at an earlier stage of the disease; this means a higher likelihood of surviving 5 years after the initial diagnosis, independent of improved treatment. Third, treatment is actually producing better outcomes. Unfortunately, most of the measured gain in survival has occurred because more women have been diagnosed at an earlier stage of the disease.

The story for prostate cancer is similar. Older men are increasingly aware of the risk of prostate cancer, and the use of PSA tests to detect the disease has expanded rapidly. This has led to a 190 percent increase in the rate (per thousand men in the population) diagnosed with prostate disease, and survival rates have improved from 43 percent in 1950-54 to 93 percent in 1989-95. Unfortunately, the number of deaths due to prostate cancer per 100,000 men in the population (that is, the mortality rate) during this same period actually rose. Again, the improvement in survival rates primarily reflects earlier diagnosis rather than significant improvements in treatment.

Because of this discrepancy between 5-year survival rates and mortality rates, there is controversy among clinicians and medical researchers about the benefits of universal screening for prostate cancer, particularly for older men. The reason is that prostate cancer typically grows quite slowly; the median time between detection of prostate cancer through the PSA test and the ability to detect it clinically is about 10 years. Men may have prostate cancer, be entirely unaware of it, and die of something entirely different. Both prostate cancer and breast cancer hold promise for substantial technological breakthroughs that would reduce mortality rates, just as they have for coronary heart disease. Until that time, management of the disease can benefit from a better understanding of the treatment options available to patients.

understanding of the genetic and molecular mechanisms of diseases, treatments are likely to become even more tailored to individual circumstances. All of these examples suggest that better scientific knowledge, including more information from both randomized clinical trials and large-population studies of actual practices, can lead to substantial productivity improvements through more efficient use of the many medical treatments available today. These improvements in productivity can be facilitated by developing systems to disseminate information about the value of different interventions—their benefits, risks, and costs—and by developing better electronic health records with effective privacy protections. Providing patients with better information about the true value of different treatments, coupled with stronger incentives for patients and providers to use approaches of demonstrated value, will help ensure value and productivity in health care in future years.

The Role of the Federal Government in Supporting Research

The impressive improvements in the health of Americans over the past several decades have not occurred in a vacuum, but arose because of work—much of it collaborative—by government, private, and charitable organizations in support of basic research, clinical testing, and product development. The health care system of the future will need to preserve and encourage this product development, through direct support for research with potentially broad applications, and through the protection of patent rights, to help turn promising new research insights into treatments approved for clinical use. The government can also provide critical support for improving our knowledge of how to use existing medical treatments even more effectively. Follow-up clinical trials often find that medical treatments that are beneficial for the average patient in a population may have no beneficial effects for some subgroups and may even cause them harm. There may be insufficient private incentives to explore which of the many types of patients—younger, older, sicker, healthier—with a given clinical problem actually benefit from a treatment, yet this understanding may have important implications for the best treatment decisions for individual patients and for the costs of public and private health insurance programs.

In addition, research on the underuse, overuse, and misuse of treatments has benefits that extend across all who pay for health care, and as a result, individual payers may underinvest in research to improve health care quality and safety. Thus the Federal Government should provide support for research using population data on health system performance and public health. This should include support for medical information and privacy standards that allow clinical data to be pooled for research and public health purposes.

Conclusion: Fulfilling the Potential of 21st-Century Health Care

The American health care system stands at a critical juncture. The gains in medical productivity of the last 40 years have been tremendous; the next 40 years have the potential to bring even more valuable advances. Promoting flexible, market-oriented care that responds to the diverse needs of patients is increasingly crucial to improving the well-being of all Americans. But health care costs are also rising rapidly, and enormous opportunities exist to increase the value of health care and improve health insurance coverage. Addressing these fundamental problems and fulfilling the potential of our health care system will require innovative Federal policies to help Americans get the care that best meets their needs, and to create an environment that rewards high-quality, efficient care. To meet this challenge, Federal policy must rely on market mechanisms to encourage our health care system to identify and reward high-value treatments, while reducing wasteful spending on treatments of little value. It must harness the benefits of competition for the well-being of all Americans.

Flexibility to respond to rapid changes in medical treatments and the changing needs of patients is crucial. A bureaucratic system that fails to respond to patient needs or that is slow to embrace new technological developments is not the appropriate foundation for the future of American health care. Nor is a health care system that creates perverse incentives, rewarding the underuse of effective treatments and the overuse of ineffective ones while penalizing providers who seek to practice cost-effective care. Instead the Federal Government should improve coverage options in public programs like Medicaid and Medicare. It should ensure that Americans with limited means or high health care needs have the opportunity to participate in mainstream health plans, through refundable tax credits and strategies to increase participation in health insurance markets. It should support both biomedical research and health services research, to improve our understanding of disease, develop new treatments, and improve the quality and value of health services. It should encourage the development of better information on the quality and outcomes of care. And it should support an environment for medical practice that encourages high-quality, efficient care that meets patient needs. The need to empower patient choice and enhance market-oriented incentives calls for government policies that move away from detailed top-down regulation and one-size-fits-all government-run programs, and toward ensuring that all Americans have innovative health care options.

These changes in our current system are likely to affect both patients and providers. As the health care sector continues to grow, it becomes increasingly important to encourage new medical options that are worth the cost to consumers. Economic theory suggests that those critical decisions should generally be made by those with the best information and the most direct stake in using that information appropriately: the patient and his or her medical providers, not government or insurance plan bureaucrats. But economic theory also suggests that the ability to make these decisions should be paired with responsibility for their consequences, both for health and for medical costs.

Decisions about health care and health care systems, for both providers and consumers, require not only good information but also financial responsibility. Medical providers have a responsibility, as well, to assist patients by examining their own practices through the unflinching analysis of errors when they occur, and by reexamining long-held beliefs about the standard of care in light of new evidence about treatment effectiveness and costs. Already, case studies of both private payers and public plans around the country indicate what these efforts can achieve. Public policy should encourage these promising trends.

Finally, the Administration's overall economic policy is a critical factor in improving our ability to provide high-quality care. Rapid economic growth in the mid- to late 1990s helped keep the rise in health care costs roughly in line with growth in Americans' earnings. Uninsurance rates declined in 1999 and 2000, in large part because of the increased takeup of private, employer-provided health insurance, which, thanks to productivity increases, was becoming relatively less expensive as a share of compensation. Encouraging rapid economic growth not only will help keep private health insurance more affordable; it will also provide a growing revenue base for Medicare and other Federal programs.

Economic growth is not enough, however. A growing body of research, confirmed by many examples from the public and the private sectors, suggests that we can do a much better job of allocating medical care resources both efficiently and equitably. Providing competitive choices for all Americans, and meaningful individual participation in those choices, is the best way to encourage needed innovations in health care coverage and health care delivery. Improving the information available to guide choices, taking steps to help individual patients and providers use that information effectively to provide patient-centered care, and making a range of additional policy changes that create an environment of medical practice that encourages innovation and high-quality care will help ensure that health care remains one of the most dynamic and productive sectors of our economy.

Redesigning Federalism for the 21st Century

The Nation's federal system is one of the great strengths of the American economy. Federalism gives States and localities the freedom to provide services that best meet the needs of their diverse populations. It puts citizens closer to their government, and thus in a better position to monitor and control how their tax dollars are spent, and it creates competition between jurisdictions, which drives innovation.

The Federal Government plays a crucial role in the effectiveness of this system. It is important for the Federal Government to seek a framework for competition and accountability that avoids burdensome rules and regulations, which undermine the competitive advantages of State and local governments. Rigid dictates from Washington about how public goods and services are provided preclude innovation and dull competition. Creating a flexible institutional structure that will allow the efficient provision of public goods, by focusing on achieving goals and freeing up innovation, is an important goal of this Administration. In this way the Federal Government can improve the quality and efficiency of public programs and increase their responsiveness to public needs.

The advantages of this results-oriented, flexible approach are evident in many programs and at all levels of government. First, when the focus is on results, such as student achievement, rather than on process, such as how schools spend money, States, localities, and private organizations are empowered to choose, from a wider menu, the most effective means to these ends in their area and for their population. Second, flexibility allows more institutions to become involved in providing these services. As long as all are evaluated on the basis of results, governments, nonprofit organizations, faith-based organizations, and others can compete on an equal footing, while using different methods. The resulting laboratory of methods allows more effective ideas and organizations to win out over less effective ones, creating the potential for more and better services for a given amount of spending.

This chapter examines both the promise and the challenges of federalism, focusing on three specific areas of program design in systems of flexible accountability: education, welfare, and health insurance for those with low incomes. In education, this Administration believes that the competition provided by choice is the best tool available to improve quality, with public, private, and charter schools vying with each other to provide the best education most efficiently. When the right institutions are in place, parents can

hold school systems accountable for results. Similarly, taxpayers must be able to hold the providers of safety net programs, like welfare and Medicaid, accountable for the quality of services they provide and the resources they use to provide them. By tying payments to social service providers to the results that they achieve, and by allowing private nonprofit providers to compete on an equal footing with government providers, the same market discipline that drives innovation and efficiency in the private sector can be brought to bear on these programs as well.

Institutional Design in a Federal System

The preeminent means for providing goods and services in the U.S. economy is private markets. The fundamental strength of the market system is that consumers are able to evaluate the quality and price of a variety of goods and services that they might purchase, and are free to make decisions about which vendors to patronize. Competition among providers promotes efficiency, which means goods and services of the highest quality at the lowest cost.

In those circumstances where markets do not work efficiently, there may be an avenue for governments to improve overall economic performance. An example is the provision of public goods. Public goods are those goods and services that, in contrast to conventional private goods, provide benefits for society beyond those enjoyed by any individual consumer. For example, there is no single “consumer” of a cleaner environment; as discussed in Chapter 6, environmental protection is therefore a public good. Similarly, each member of the population gets the benefits of safer streets, or a better informed electorate, or a public park. Here the collective nature of the benefits flowing from the good or service makes it difficult or impossible for private providers to make any single consumer pay for it. To ensure the availability of these public goods, the government may arrange for their production, provision, and financing.

The long federalist tradition in the United States is a tremendous resource for governments seeking to meet this challenge. A neighborhood park, for example, is a local public good, shared by the citizens of a local area, not the Nation as a whole. Getting the “right” amount of these local public goods in every locality would be an insurmountable task for a central government. Instead, State, county, city, and town officials, who are closer than their Washington counterparts to the needs and desires of their electorates, are better positioned to be responsible for these goods. Moreover, there is a natural check on their actions: residents voting at the ballot box—or with their feet, by moving elsewhere—force local governments to compete. Just as

private firms compete in markets for private goods, so, too, governments can compete in terms of the quality, price, and quantity of the services they provide, and this fosters innovation and efficiency. This marketplace for government services constitutes a more efficient means by which to provide these services in our society.

Although there might be a clear role for governments in providing local public goods, it is not immediately obvious that it is efficient for the public sector to produce a particular public good or service. Instead the government could choose how much to provide but rely on the private sector to undertake actual production. If minimizing costs is the only objective, complete reliance on competitive private sector production will likely be efficient. In other circumstances, however, competition could foster an excessive focus on cost reduction to the detriment of achieving results of the desired quality. (This is especially likely when it is difficult to write contracts that comprehensively specify the level of quality to be achieved.) Strictly public provision, on the other hand, might promote a focus on high-quality results without due consideration of the efficient use of public resources. Which is the better method of production depends on how difficult it is to observe the quality of the services provided, the degree to which cost reductions affect the level of quality, and the potential for innovation in producing the services.

Thus, although competition between jurisdictions generally promotes the efficient provision of public goods and services that are tailored to the diverse needs of their citizens, it is neither always necessary nor desirable that those jurisdictions themselves produce those goods and services. The focus of public spending should be on efficiency: on the quality of results achieved for every dollar spent.

One way to produce public goods more efficiently is to let private firms compete for public contracts. Some municipalities contract out services such as trash collection to private vendors through competitive bidding. There is no reason, however, that such competition should be restricted to the for-profit sector. Indeed, government agencies can promote competition through outside contracts for staffing, limited reliance on exclusive grants and contracts, and opening competition for grants and contracts to faith- and community-based organizations. In each of these cases, it falls to the government responsible for providing the service to monitor the quality of services provided and to ensure, through whatever contracting means are available, that services being purchased with public funds live up to public expectations and requirements. Competition between governments can then lead to the right public goods and services being provided with the greatest efficiency.

In practice, several complex issues arise in a federalist approach. First, by its nature, competition among governments offers no guarantee of equal outcomes: competing jurisdictions may differ greatly in the resources at

their disposal to finance government services, and thus in the amounts and the variety of services that they can offer. Although these differences may reflect differences in the tastes of households across jurisdictions—and thus show that the government marketplace is working—they may run counter to a desire for greater equality. In these and other circumstances, the Federal Government may choose to provide funds to State and local governments in a way that makes outcomes more equal. That is, it may seek to alter the result of the federalist system. This may be desirable in itself, but often the Federal Government has chosen to dictate the use of these funds. Such mandates are at odds with the goal of encouraging State and local governments to respond flexibly to the desires of their constituents.

The history of federalism is to a large extent a history of the struggle to achieve an optimal balance between allowing flexibility for State and local governments and maintaining accountability for the use of Federal funds. The New Deal of the 1930s and the Great Society of the 1960s consolidated in the Federal Government much authority for the programs they created, and Federal spending increased from 3.4 percent to 19.3 percent of GDP between 1930 and 1970. Then, in the mid-1970s, the “New Federalism” sought to increase efficiency in the federalist system and to devolve program control to States and localities, while introducing such innovations as Community Development Block Grants and general revenue sharing. In the late 1970s, the Federal Government sought to expand its authority over these block grants. Ninety-two new categorical grant programs were instituted from 1975 to 1980. (Categorical grants are those that must be spent on a designated population, and they may involve Federal matching of State funds.) In the 1980s, the tide once again turned toward decentralization: 77 programs were consolidated into 9 block grants. Much like the 1970s decentralization, this movement was thereafter partially reversed as more constraints were placed on the block grants, and previously scaled-back regulations again became more cumbersome. The major federalist initiative of the 1990s was the partial decentralization of welfare. These swings highlight the tension between the desire for assurances that Federal funds will be spent productively to advance program objectives, and the desire to take advantage of the efficiencies generated when local agencies have the resources and the freedom to innovate and to cater programs to local populations.

These two goals need not be at odds. Federal micromanaging of resources and processes achieves neither. By focusing instead on setting standards for results—not dictating actions—and rewarding providers for achieving goals, the Federal Government can give local governments more control over the use of funds without sacrificing progress toward national goals. This focus on outputs is a key piece of the infrastructure for an efficient federalist system, one that centers attention on what is delivered to the final consumer and puts in place incentives to identify and measure desired results. This

Administration has signaled its commitment to such systems through its vision for Federal, State, local, and private partnerships across all areas of public spending.

Fostering Partnerships, Competition, and Accountability

Organizations, be they public or private, that accept Federal funds in return for providing a service must agree to provide that service in a manner that meets Federal standards and goals. It is desirable, however, that they do so with the minimum interference possible. In activities where measuring results is difficult, it is harder to hold providers accountable. In some cases the data currently available are insufficient for this task. However, it is important to recognize that the existence of good data on program outcomes is in large part determined by the measures used to evaluate the programs. Developing a system of accountability based on well-measured output will promote the collection and analysis of this important information. This Administration seeks to create an institutional framework that will encourage the development of measurable standards to which all providers of public services—Federal and local, public and private—can be held accountable, and then to allow these providers themselves to find the best way to meet those standards.

Leveling the playing field for governments, nonprofit providers, and for-profit providers, and thereby encouraging the free entry of all providers, promotes market efficiency just as in the private sector. This is a desirable goal, and not an entirely new phenomenon. Market forces already bear on for-profits, but they do on nonprofits as well, when they compete for private donations. In a 1998 survey, 75 percent of respondents said that whether or not a charity used their time and money efficiently affected their choice of charities. Allowing private providers to compete with public agencies to provide services in areas such as welfare, and evaluating all providers based on achieving program goals, are ways of expanding this market discipline to public providers. However, several institutional and logistical barriers currently inhibit this kind of competition. For example, although the Charitable Choice provision of the 1996 welfare reform legislation was intended to allow faith-based organizations to compete on an equal footing with other organizations to provide welfare services, preexisting laws and regulations in many States still prevent them from participating. This Administration is committed to eliminating these barriers.

Despite these impediments, many State governments are already forging new partnerships with private organizations for the provision of high-quality

public services through performance contracting in social services. Performance contracts usually include output targets and may make the size of payments contingent on meeting those targets. States have long used performance standards in their budgeting processes. For example, under Texas's approach to performance measurement, agencies are required to include 6-year strategic plans in their budget requests. Each plan must specify the agency's goals, objectives, outcome measures, strategies, and efficiency measures. Pennsylvania has included performance measurement in its program budgeting for over 25 years. As of 1997, 31 States had legislated some form of performance-based budgeting requirements, and 16 had implemented such measures through guidelines and instructions.

Although such provisions have long been standard in municipal service contracts such as those for garbage disposal, they are relatively new in social service contracting. In the municipal services sector, results may be more easily defined and codified in contracts: for example, where and how often trash will be collected. However, the quantities and the quality of social services desired can be much harder to specify and to observe, making contracts more difficult to write. Recipients may not have the expertise to evaluate the quality of the services they are receiving, and they may not have the option of changing service providers if dissatisfied. In such circumstances, the contracting agency must provide oversight to ensure that adequate services are provided. Creative solutions have been devised for some of these problems; for example, providers can be required to meet a professional or industry standard, potentially simplifying contracts. The Federal Government could make performance contracting easier for States by developing generic contracts for commonly used social services, which interested States could then adapt to their particular needs.

These public-private partnerships illustrate some of the advantages and some of the difficulties of designing programs with flexibility and accountability in a federal system. These issues are explored below in the realms of education, welfare, and Medicaid.

Elementary and Secondary Education

Unlike many other publicly financed services, primary and secondary education has historically been under the control of local governments, with educators accountable to local taxpayers. Taken at face value, this suggests that the forces of competition should already be at work to promote high-quality, efficient provision of public education. To some extent, taxpayers have the ability to control the quantity, quality, and price of education by “voting with their feet”: if the local school district fails to perform adequately, they can move elsewhere. In some jurisdictions, citizens vote directly on

property taxes, or even on school budgets. Parents may also remove their children from the public school system altogether by placing them in private schools or home schooling them.

These mechanisms are more effective, however, when parents can accurately evaluate the quality of local schools. When they cannot, or when local alternatives to poor-quality schools do not exist and moving is prohibitively expensive, effective competition is limited. Also, given the broader social benefits of a well-educated work force, some redistribution may be necessary to ensure that all children have access to an adequate education. Thus, even though State and local governments retain the primary responsibility for educating the Nation's children, and face competitive pressures in doing so, the Federal Government can still serve a vital role in further lowering barriers to local competition.

This Administration seeks to create and strengthen the institutions that allow local education markets to work, that let school districts cater to the diverse needs of their populations, that empower parents to choose what is best for their children, and that ensure that no child is left behind. An efficient and effective market for education, much like any other market, requires freely available information and incentives for good performance. Tests are a key component of this framework. This Administration believes that once this information and these incentives are in place, competition among schools is the best way for parents to make sure their children receive the best education possible. School choice empowers them to do so. To ensure that adequate options are available for all children, the Federal Government can provide supplemental resources for the education of low-income children and children with special needs. However, these subsidies must be designed so that they do not interfere with the incentives for schools and school districts to spend efficiently.

The No Child Left Behind Act, proposed by the President, passed by Congress, and signed into law on January 8, 2002, addresses each of these goals. It is a major step toward improving the quality and efficiency of the schooling available to America's children. The rest of this section discusses in detail the principles that underlie this legislation.

Setting Standards and Measuring Progress

In the provision of education, accountability hinges on the development of adequate measures of results. In the long run, important measures of the success of education are the well-being, self-sufficiency, and productivity in adult life of today's schoolchildren. As a practical matter, however, it is difficult to evaluate schools based on their pupils' accomplishments 10 or 20 years later. For this reason, tests are a fundamental building block for school accountability. This Administration believes that well-designed tests are

among the most valuable tools for evaluating school performance, giving early feedback about the success or failure of programs, educational reforms, teachers, and students alike. They augment the other information parents need to evaluate their children's schools. The No Child Left Behind Act makes available school-by-school report cards, which include data on test results, to help parents make the best decisions for their children.

Although the Federal Government provides substantial funding to States for education, State and local governments themselves contribute the lion's share—over 90 percent—of the funds for public elementary and secondary schools. Consistent with its focus on results, this Administration believes that States should have the freedom to design tests that provide parents with the tools they need to evaluate local school systems, and the No Child Left Behind Act specifies that each State be evaluated based on the test of its choice. At the same time, however, a key aspect of good testing is comparability: the ability to compare schools within districts, and districts within a State. The tests that States choose must be consistent enough so that parents can use them to evaluate their children's education and make well-informed choices. The National Assessment of Educational Progress (NAEP), a nationally representative test designed to evaluate America's students and schools, is also a useful tool for evaluating student progress at the national and the State level. The Federal Government has provided funds through the No Child Left Behind Act for some of every State's fourth and eighth grade students to participate in the NAEP.

Designing good tests is only the first step in strengthening school accountability and enhancing competitive efficiencies in education. Tests serve two goals: to create incentives for students, teachers, and schools to excel, and to trigger appropriate consequences for failure. When schools fail, parents should have the choice to move their children to better schools. To this end, the No Child Left Behind Act makes Federal education funding conditional upon local school districts and States taking defined steps to improve schools that fail to make adequate yearly progress, as determined by testing.

Expanding Options

Once clear, measurable results have been defined, competition can be a strong motivating force for improving schools. This competition can come from several sources, including other public schools, charter schools, and private (including parochial) schools. School charters and the contracts of educational management organizations (EMOs are private enterprises that run charter schools and contract with school districts to operate individual public schools) can be reviewed before renewal, and if measures of their results are publicized, parents and school districts alike will be able to

evaluate their performance. The No Child Left Behind Act supports school competition (through the creation of charter schools, for example), which can improve school quality and increase the choices available to parents.

There are currently some 2,400 charter schools operating in 37 States and the District of Columbia, and the number is growing rapidly. The performance-based competition for students that charter schools exert puts pressure on all schools to excel. Indeed, research shows that competition from charter schools forces traditional public schools to respond and improve. Many school districts are also experimenting with outsourcing education to EMOs, which brings the benefits of market competition to public education. Some studies of EMOs suggest that they perform well relative to their public school counterparts. Competition from private schools can have a similar effect: one study found that such competition significantly increased the performance of public schools in the same area. Another study found that competition among public schools seems to both increase achievement and lower costs.

The No Child Left Behind Act also ensures that parents in school districts receiving funds under Title I of the Elementary and Secondary Education Act (ESEA) will have the option of moving their child to another public school in their district if the child's school has failed to make adequate yearly progress (as defined by the State) for 2 or more consecutive years, except where that option is prohibited by State law. Students in schools that fail for 3 straight years can receive funds to obtain supplemental educational services, such as tutoring, after-school services, and summer school programs. These options would benefit students in thousands of schools that have already been identified as failing under current law. Finally, if a school fails to make adequate yearly progress for 5 consecutive years, it will face restructuring as a condition for the State in which it is located to continue to receive Title I funds. Such restructuring by the State or locality may take forms such as conversion to a charter school, contracting with an EMO, or complete reconstitution of the school. Furthermore, any school district receiving any funds under ESEA must provide parents with the option of moving their child to another public school if the child has been the victim of a violent crime at school, or if the State determines that the school is unsafe. Giving localities the ability to offer parents options other than relocation prompts schools to perform well to keep their students, and it gives students in failing schools additional options. At the same time, the financial consequences for failure engender market-like discipline.

Vouchers could also increase the power of school competition. Vouchers allow parents to use the money that would be spent in their public school district to purchase education at another existing public or private school. School vouchers of various forms are available to parents in 38 States and the

District of Columbia. In some cases, however, these voucher programs are thought to be too small to provide strong incentives for public schools to improve, shifting too few educational dollars away from failing public schools. Similarly, in some rural areas vouchers may be less effective if there are not enough students to support multiple schools. Preliminary academic evidence, however, suggests that vouchers can be effective. Evidence from randomized field trials in Dayton, Ohio, New York City, and Washington, D.C., found that African American students receiving vouchers achieved moderately large gains in test scores after 2 years. Evidence from voucher experiments in Milwaukee suggests that students realized gains in both reading and math. Tax credits are an alternative vehicle that can deliver the power of choice to families. What these initiatives have in common is that they exploit the ability of markets to give parents the power to choose the highest quality and most efficient education available for their children. The ability to make those decisions depends crucially on the availability of standardized and meaningful data, which testing can provide.

Providing for Vulnerable Populations: Government Partnerships

There is a compelling public interest in ensuring adequate educational opportunities for all children. Children who are well educated are likely to become more productive members of the work force, are less likely to need public assistance later in life, and tend to pass along their social and material well-being to their own children. To the extent that local school districts do not take these long-run effects into account, and given the difficulty of redistribution at the local level, subsidizing education for low-income children and children with special needs is a valuable State and Federal function. This Administration has made it a priority that no child be left behind.

Educational Resources for Low-Income Populations

The Federal and State governments have taken different approaches to ensuring adequate educational resources for low-income school districts. Most States have experimented with some form of school finance equalization (SFE) in the past 30 years to redistribute funds to low-income districts. SFE programs mainly seek to redistribute funds from districts with high property values per pupil to districts with lower property values per pupil. In practice, however, many SFE programs actually redistribute funds based on per-pupil education spending, not property values, and property values themselves may be affected by tax rates.

State SFEs, if not carefully crafted, not only may fail to increase the resources available to low-income students, but indeed may decrease the

resources available to all students. This can happen for any of several reasons. When redistribution of funds to poorer districts is based on district spending levels, it becomes, in effect, a tax on education spending by the high-spending districts, which may respond by reducing spending. Thus equalizations that rely on this approach may have the unintended consequence of “leveling down,” achieving greater equality only by lowering average spending per pupil; this could even result, perversely, in lower per-pupil spending in poor districts. SFEs that subsidize local education spending through matching may be able to “level up” through infusions of State funds.

The Federal Government, under Title I of the Elementary and Secondary Education Act, targets funds to low-income students through their school districts. Providing grants to high-poverty districts out of general revenue has the potential to be much more effective and less distortionary than State-level SFEs. Federal Title I aid may be particularly valuable to high-poverty districts in States with limited fiscal resources available to fund equalization programs. In fiscal 2001 the Federal Government allocated almost \$9 billion to Title I, to reach approximately 12.5 million students in both public and private schools. In fiscal 2002 the Federal Government will spend more than \$10 billion, and the President's 2003 budget requests an increase of roughly 10 percent. Federal education funds are more narrowly targeted to high-poverty school districts than State and local funds. The poorest quartile of school districts received 43 percent of Federal funds, but only 23 percent of State and local funds, in 1994-95. Title I, Part A, funds are generally targeted to students deemed most at risk of failure, but if half or more of a school's students are living in poverty, the funds may be used for school-wide programs. To discourage States and localities from shifting their funding responsibilities to the Federal Government, Title I conditions Federal funding on local and State resources being comparably allocated to Title I and non-Title I schools. Beyond these two conditions, schools have a great deal of flexibility in the use of Title I funds, and this flexibility should allow districts to use funds to meet their most pressing needs.

To promote quality in education, since 1994 the Federal Government has been using access to Title I funds to encourage districts to establish results-oriented infrastructures. States' Title I funding was made dependent upon their implementing final assessment systems and providing the Department of Education with evidence that such systems met Title I requirements by the 2000-01 school year. In addition, through Title VI the Federal Government provides grants to assist local education reform efforts that are in keeping with statewide reforms, and to support other promising local reforms. These programs are two examples of how the Federal Government can encourage the creation of desired institutional infrastructures while maintaining flexibility at the State level.

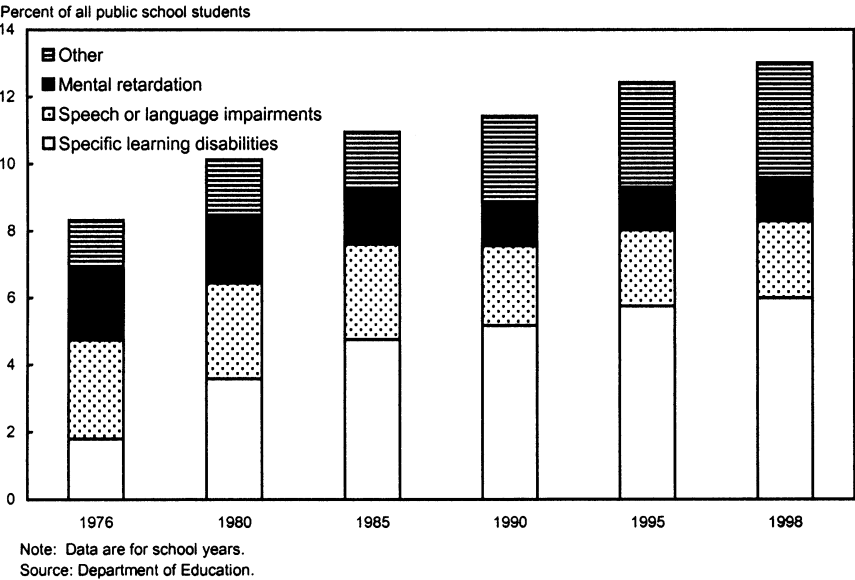
Special Education Funding

Although education of children with special needs is primarily a local responsibility, State and Federal resources also support this important work. The courts have determined that States and localities are constitutionally required to educate students with disabilities, and when Congress passed the Education for all Handicapped Children Act (now the Individuals with Disabilities Act, or IDEA) in 1975, States were given Federal dollars in exchange for providing free, appropriate education to all such students. One study estimates that Federal, State, and local governments bore, respectively, 8 percent, 47 percent, and 45 percent of the cost of public special education provision in 1998-99. The President's 2002 budget requests a 13 percent increase in IDEA grants to States. This spending can have significant payoffs for children with special needs: research shows that special education programs improve the math and reading test scores of special education students and do not undermine the achievement of other students.

The conflicting interests described in the earlier discussion of public-private partnerships can also be seen in intergovernmental partnerships. Special education is a prime example, demonstrating the issues that arise when those who provide services do not fully bear either the cost of those services or accountability for their results. In the past, the extra resources that categorical State and Federal funding made available for special education students may have created incentives for school systems and parents to expand the population identified as having special needs. Indeed, there has been a steady rise since the late 1970s in the percentage of students so classified, with the greatest increase in those categories, such as learning disabilities (as opposed to physical disabilities), where the identification of need is most subjective (Chart 5-1). African American and Native American students make up a disproportionate share of those referred into special education. Furthermore, school districts are often able to exclude special education students' test scores from State assessments; this may give them an incentive to refer students to special education inappropriately.

To address these undesirable incentives, the 1997 IDEA reauthorization changed the way in which Federal special education funds are allocated to States, but these funds account for less than 10 percent of all special education funds, and many undesirable incentives persist at the local and the individual levels. The subjectivity of such hard-to-observe classifications makes well-designed systems and incentives essential. On October 2, 2001, the President signed Executive Order 13227 to establish the President's Commission on Excellence in Special Education. This commission will examine these and other issues to prepare the Administration and Congress for the upcoming IDEA reauthorization.

Chart 5-1 Children in Federally Supported Programs for the Disabled
 Both the fraction of children in programs for the disabled and the proportion of children classified as having specific learning disabilities have dramatically increased.



Summing Up: Getting Incentives Right

Education is one area of public spending that has traditionally been subject to competition among localities, and between public and private providers. Research suggests that this competition has led to measurable gains in student achievement, but there is also an important role for the Federal and State governments to play in redistribution and social insurance. In designing systems that provide these valuable services while maximizing local flexibility, it is imperative to account for the influence of incentives on governments, schools, teachers, parents, and students alike. By rewarding good performance at all levels, programs can align individual incentives with public goals to promote efficiency and excellence. Indeed, these lessons pertain beyond the realm of education.

Welfare

Safety net programs such as welfare and Medicaid pose some of the greatest challenges—and the greatest opportunities—for more efficient provision of services in a Federal system. The ability of taxpayers to vote with their feet is more constrained in this setting than in education, because, as

discussed below, social insurance is harder to achieve at a local level. This does not mean that competitive forces cannot be harnessed to foster greater efficiency in providing support for low-income families. Rather, it is in these areas in particular that flexibility of method and careful accountability for results are likely to achieve the greatest gains, and where it is most important that the results to be evaluated be chosen and measured well.

The 1996 enactment of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) replaced Aid to Families with Dependent Children (AFDC), the primary Federal welfare program, with Temporary Assistance for Needy Families (TANF). PRWORA increased State discretion over the use of welfare funds by converting federally matched grants to block grants, thereby affording States greater flexibility. PRWORA also set time limits on benefit eligibility for recipients and created a framework for innovation in welfare reform. PRWORA was introduced in the wake of record highs in welfare participation and extensive program experimentation. Already before the passage of PRWORA, many States had been granted waivers, and 27 States had obtained major waivers exempting them from various aspects of AFDC's eligibility and process requirements, allowing them to experiment with alternative approaches. PRWORA widened this flexibility to all States. Welfare caseloads declined dramatically following PRWORA's enactment. Between August 1996 and June 2001, the number of TANF recipients was reduced by 56 percent nationwide. Although favorable economic conditions certainly played a role, research suggests that roughly a third of the decline was due to welfare reform (Box 5-1); estimates vary, however. PRWORA appropriated funds for TANF grants to States through fiscal 2002. Hence this year Congress will determine appropriations for fiscal 2003 and beyond. This provides an opportunity to review the program, the principles on which reforms were undertaken, and those that should guide the program in the future.

Focusing on Results

A prominent feature of PRWORA is its restrictions on benefits; these include 5-year lifetime eligibility limits and the condition that beneficiaries find work after receiving benefits for 2 years. Just as important, however, PRWORA also mandated the devolution of program design to the States (some States further devolved welfare provision to the counties) and increased flexibility and opportunity for innovation in welfare provision. When TANF replaced AFDC, the Nation moved from a welfare system in which the Federal Government prescribed the process of service provision to one in which it defines goals and creates incentives, leaving the process to be determined largely by each State. Under the former centralized, process-based approach, the Federal Government determined how funds were

Box 5-1. Why Have Welfare Caseloads Declined?

There is no question that strong economic performance and the resulting tight labor market of the late 1990s account for a portion of the recent decreases in welfare caseloads. However, the decline would not have been nearly as sharp were it not for the structural changes in the safety net programs that support working families.

In 1999 the Council of Economic Advisers found that only 8 to 10 percent of the decline in welfare caseloads between 1996 and 1998 could be attributed to changes in the unemployment rate; research also suggests that welfare reform was responsible for roughly one-third of the reduction. The lifetime time limits imposed under PRWORA create incentives for welfare recipients to find jobs (even before they reach the limit), and researchers have found that the imposition of time limits alone was responsible for over 10 percent of the decline in welfare caseloads between 1993 and 1999. In addition to encouraging self-sufficiency through time limits, PRWORA explicitly conditions benefits on welfare recipients engaging in work-related activities, and since its passage there has been a dramatic increase in the work participation rates of welfare recipients. This employment experience continues to help former recipients over their lifetimes by building their human capital and thus improving their future employment prospects.

Increases in other forms of support for working families also made work more appealing, by making it more lucrative relative to welfare receipt. After the passage of PRWORA, people could leave welfare without fear of losing valuable Medicaid coverage (as long as their income remained below eligibility limits, or for up to a year after it rose above those limits). They could also continue to receive child care subsidies, and many were eligible for an expanded Earned Income Tax Credit. These expansions were also likely responsible for part of the decline in caseloads. For example, one study found that in 1986 a single mother with two children, who left welfare to work full time at the minimum wage, would have increased her family income by only \$2,000 (and would still have been living on income of only 80 percent of the poverty line); she also would have lost her eligibility for Medicaid. The same woman in 1997 would have increased her family income, upon leaving welfare, by \$7,000 in constant dollars (almost doubling her income and raising her above the poverty line) and would have likely retained her family's Medicaid coverage for up to a year.

allocated as well as many other details of the program. Under PRWORA's outcomes-based approach, in contrast, funds are appropriated to decentralized providers for the pursuit of defined objectives, and these providers are then given discretion over how the funds are used. Although process and design are important features of any program, emphasizing ends rather than means can be a more effective way to reach goals.

Participation in some other assistance programs, for example, is conditioned on participation in job training. Although the goal of such requirements is noble—to enable recipients to become self-sufficient members of the work force—uniform training requirements may not be the answer for all workers. Some might benefit more from relocation assistance, or from income support to allow a longer job search. For some workers a greater obstacle to employment may be lack of child care or transportation. Thus, although training is one route to productive employment, it is neither the only route nor the best route for all. Assuming that the objective of these programs is to foster self-sufficiency, it is reasonable to judge the success of a program by the number of people it moves into lasting employment, rather than by the number of hours of training it provides.

The Importance of Measurement

When public policy objectives are broken down into measurable outcomes, providers can be paid and contracts awarded according to how well they achieve those outcomes. This encourages agencies and organizations to excel. By rewarding those programs that are succeeding, government can foster innovation, efficiency, and personalized solutions to the problems facing providers and their clients.

The first step toward reaching these goals is to turn public policy objectives into quantifiable measures and to set targets for those measures. When possible, such measures should accurately reflect broad policy objectives, not narrow intermediate steps. They should also strive to distinguish subpar performance due to labor market fluctuations and other anomalies from genuine program shortcomings. Providers can then have maximum flexibility and a minimum of restrictions, and be free from adverse incentives (such as the incentive to maximize training, when training is neither right for everyone nor the ultimate goal of the program).

Creating such measures is not always easy. Indeed, it is especially difficult when people and localities differ in their needs; such differences can affect both the appropriate goal of programs and the feasible outcomes. For example, getting welfare recipients into the work force is one measure of the success of welfare reform. Under PRWORA, Federal funding is conditioned on States meeting targets for the fraction of welfare recipients who are employed. Among the conditions are that 50 percent of recipient families

(and 90 percent of two-parent recipient families) be employed by fiscal 2002. States may reduce the target employment rate on which their funding is conditioned by 1 percentage point for each percentage point that welfare rolls are reduced from their fiscal 1995 levels. The dramatic decline in welfare caseloads actually observed since 1995 has meant that the overall participation requirement has been binding on very few States. Focusing solely on the size of welfare caseloads, however, could have created incentives to make recipients ineligible for welfare rather than make them self-sufficient. A broader goal of PRWORA is ending needy parents' dependence on government benefits, by promoting job preparation and work, while providing temporary income support for those who fall on hard times.

When measuring success by results, basing measures on the right outcomes is essential. These measures should ascertain the extent to which State programs are meeting the ultimate goals of PRWORA while still affording flexibility in program design. The Federal Government can help ensure that Federal, State, and local agencies have the tools they need to evaluate service providers. Although not all data may be collected currently, basing payments on progress toward those outcomes would encourage the collection of such data in the future.

The Value of Incentives

The second step in achieving the goal of innovative and effective provision is to create incentives for public and private service providers to succeed. Rewards for excellence can be paired with consequences for failure to meet minimum standards; this is especially useful when dealing with government agencies that cannot be replaced by more efficient entrants from the private sector if they fail. If a State fails to meet the work participation rate targets established in the TANF program, its block grant is reduced by an amount determined by the Department of Health and Human Services' evaluation of the duration and degree of its noncompliance. States can avoid these consequences if their performance improves in the following year under a corrective action plan. The Federal Government also has discretion in penalizing States and may choose to waive or substantially lower penalties in extenuating circumstances, such as regional recession, natural disaster, or a substantial increase in caseloads.

Flexibility is crucial to encouraging experimentation, because all experimentation entails risk. Despite a State's best intentions and efforts, reforms that appeared promising may not succeed. By giving the Federal Government discretion in penalizing failing States and using corrective action plans, TANF seeks to prevent such penalties from discouraging the very innovation it intends to foster. This furthers the ultimate goal of creating a system that encourages the development of effective and efficient programs.

The Benefits of Flexible Approaches

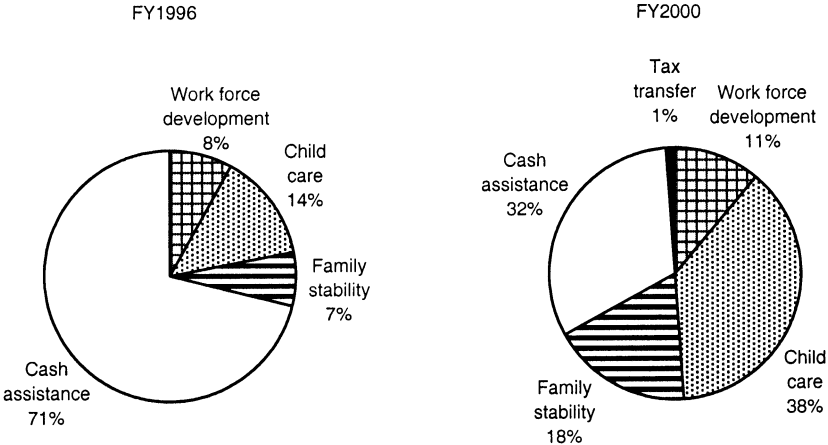
This Administration believes that welfare goals and targets should be flexible enough to accommodate local differences, encourage innovation, and foster excellence, and that such flexibility must be accompanied by accountability, careful monitoring, and rewards for progress toward meeting goals. Providing these rewards based on comprehensive outcome measures allows States, localities, and organizations facing different economic and demographic circumstances to design programs that work best for them. People on welfare face different obstacles to self-sufficiency and will therefore benefit from different services. Similarly, regional demographic and geographic differences shape the types of assistance that are appropriate, and State programs, capacities, and opportunities differ as well. The idiosyncrasies of local labor markets mean that the types of education and job training programs that are beneficial may vary widely across communities and over time. States have been using the flexibility granted under TANF to tailor programs to the needs of the populations they serve. As a consequence, between 1996 and 2000 the composition of welfare spending by type of assistance changed dramatically (Chart 5-2).

One example is subsidies for transportation. Lack of transportation can impede welfare recipients from getting training and holding a job. In an urban area with a well-developed transportation system, providing transportation subsidies to welfare recipients may make sense. Rural areas, however, may lack public transportation, and even some urban areas may have inadequate public transportation between the neighborhoods where many welfare recipients live and those where employment is available. States are using TANF funds to address these difficulties in a variety of ways. Governments in some States, such as Michigan and New York, are working with the providers of public transportation systems to expand access and service provision. Others are establishing programs to assist welfare recipients in purchasing or leasing their own automobiles, and some State agencies are providing transportation themselves.

Child care assistance is another area in which States are using their greater flexibility to increase funding, despite the disappearance of a mandate to provide this service. TANF released States from AFDC's conditions that they guarantee child care to all recipients who need it to work or go to school. Yet more stringent work participation requirements have likely increased recipients' need for child care services. In response, States have used the flexibility in TANF to increase child care funding: in fiscal 1999, Child Care Development Fund transfers and TANF funds directly spent on child care totaled \$4.4 billion, more than double the amount spent in fiscal 1998. Many States have experimented with child care vouchers, which have

Chart 5-2 **The Changing Allocation of Welfare Funds in Six States**

Welfare funds shifted away from cash assistance toward other needs after the passage of the Personal Responsibility and Work Opportunity Reconciliation Act.



Note: Spending data are for six States that participate in the Midwestern Welfare Peer Assistance Network: Illinois, Indiana, Iowa, Michigan, Minnesota, and Wisconsin.
Source: Smith, Courtney, Susan Golonka, and Fredrica D. Kramer, "The Evolving Nature of Welfare Reform: Where We Stand on the Eve of Reauthorization," National Governors Association, 2001.

reduced the paperwork required of them and made it easier for parents to take advantage of child care subsidies. States have clearly tapped into an important need among their populations and generated innovations in service provision.

These examples reflect broad shifts taking place in State welfare programs in the wake of PRWORA. Overall, between 1996 and 2000 State welfare spending shifted away from cash assistance toward providing social services. Beyond targeting services to communities, many States are using their newfound freedom to experiment with the structure of their welfare programs, recognizing that incentives matter for individuals, organizations, and governments alike. In 2000, 34 States offered “diversion payments” or services to families applying for TANF benefits. Most of these States provided lump-sum payments in lieu of monthly benefits. It is hoped that these payments, sometimes termed welfare avoidance grants, will enable families to weather a temporary emergency while avoiding attachment to the welfare system. Another structural innovation aimed at preventing welfare dependence is an intermittent time limit. Thirteen States are currently experimenting with such limits, which deny or reduce benefits for a period of time after a family has received assistance for a given number of months.

Some States are further devolving welfare to counties and local governments. California, Colorado, New York, North Carolina, and Ohio give counties block grants with which to provide welfare services. Like the Federal Government, these State governments are seeking to balance the desire to give local governments freedom to innovate, and to tailor programs to local needs, with the need to maintain standards. Most States that have ceded partial control of programs to localities, however, maintain some control over the criteria for eligibility, benefit levels, work requirements, and time limits.

One of the great advantages of flexibility in the laboratories of State programs is that each can learn from the experience of others. Even States that are succeeding in meeting specified outcome targets can benefit from information regarding other States' experiences. The Department of Health and Human Services, the National Governors Association, and other groups are already facilitating such information sharing. Because the Federal Government gathers and analyzes a great deal of State welfare program data in its monitoring of TANF compliance, it can play a vital role in helping States target their efforts, by disseminating information on the programs that have proved most successful.

Encouraging Broad Participation

In addition to affording States greater flexibility, PRWORA enlarged the pool of providers with whom States may contract. Under the Charitable Choice provision of PRWORA, States may administer and provide TANF services through contracts with charitable, religious, or other private organizations. Any State that chooses to involve nongovernment entities in social service delivery may not exclude providers because of their religious nature. This provision does not, however, amount to giving preference to religiously affiliated organizations. As the President stated in his executive order establishing the White House Office of Faith-Based and Community Initiatives, "This delivery of services must be results oriented and should value the bedrock principles of pluralism, nondiscrimination, evenhandedness, and neutrality."

Religious organizations have long been involved in poverty relief in the United States, and government partnerships with such groups have a long history. In 1999 Catholic Charities and Lutheran Social Services both received over half of their funding from the government. The Charitable Choice legislation prohibits agencies receiving government funds from discriminating against clients of different faiths, but it does not require religious organizations' beliefs to be strictly segregated from the services being provided. Federal funding is also conditioned on the government making an alternate service provider available if a client is uncomfortable receiving assistance from a religious provider.

The inclusion of nonsecular service providers in welfare programs is very much a work in progress. Changing agency policies and State laws that had made religiously oriented service providers ineligible for government funds is a time-consuming process. As of 2000, fewer than half the States had removed legal and policy barriers to religious organizations' participation in government-funded welfare provision, but at least 23 States had new cooperative relationships with newly eligible faith-based providers. The language of Charitable Choice extends beyond TANF to food stamps and Medicaid as well, but it has not been implemented in these programs because current law requires that a public official, not a private citizen, evaluate recipients' eligibility. Even in States and programs where legal barriers have been removed, small organizations often struggle to compete with agencies that have received government grants and contracts in the past and already have the necessary infrastructure to comply with government regulations. Federal grants and contracts typically require formal recordkeeping, monitoring, and substantial infrastructure, yet many religious congregations have outreach budgets of less than \$5,000, and few have more than one staff member assigned to such activities. Although smaller contracts might promote the incorporation of such agencies into the welfare provision network, they are not always cost-effective. Any gains from including small providers must be weighed against the costs of coordination and other increased costs associated with working with a greater number of providers.

Thus, in addition to affording States greater flexibility in the types of services they offer, PRWORA allows them to choose from a larger pool of service providers. Local organizations have a great deal to offer and can be a source of valuable innovation. They often have an established presence in the communities they serve, greater credibility than a government agency with local populations, and access to valuable volunteer labor.

Unfortunately, in the past, Charitable Choice language has not ensured that Federal administrators will require State and local governments to comply with new rules for involving faith-based providers. Faith- and community-based groups remain an underutilized resource, and this Administration will continue to work to eliminate barriers to their participation.

Medicaid and SCHIP

Maintaining a healthy citizenry is a compelling public interest, arising from the risk of the spread of disease, the loss of productivity from illness, and the altruistic motivation to provide for those who are ill but cannot afford health care. This can lead to inefficiencies in the health care system if only emergency room care is provided. For example, people without health

insurance are more likely to forgo cost-effective early or preventive care, to wait until very ill to seek health care, and when they do, to use the expensive option of emergency room care. The cost of this uncompensated emergency room care may then be passed on to the public in the form of higher medical fees or higher taxes. This compromises both the health of individuals and the public finances and suggests a role for government in subsidizing more efficient health care for low-income populations.

At present, the primary mechanism for such assistance is Medicaid, a Federal- and State-financed public health insurance program for low-income individuals who are aged, blind, disabled, or members of families with dependent children. In certain circumstances, Medicaid also provides medical care to those with high medical expenses but incomes modestly over the Medicaid threshold and to pregnant women. (States have discretion over the eligibility of both groups, and they are covered in 35 States and the District of Columbia.) The Federal Government matches each State's Medicaid spending at a rate inversely related to the State's income per capita; rates range from 50 to 76 percent in 2002. As discussed below, however, States are beginning to use their new flexibility to explore alternative ways to provide high-quality and high-value health care to their low-income populations.

States may seek waivers to use Medicaid funds to provide otherwise uncovered services and to experiment with Medicaid program design, and almost all States are now experimenting with different approaches, especially for populations whose Medicaid eligibility is not mandated. The State Children's Health Insurance Program (SCHIP; Box 5-2) provides health insurance for low-income children who do not qualify for Medicaid, under rules that provide more flexibility, and with a higher Federal match rate. These systems provide access to valuable health care for many low-income Americans and have improved the well-being of many.

Medicaid and SCHIP resources, however, could be allocated more efficiently than they are now, to provide greater benefits at lower cost, by using market mechanisms to promote access to private health insurance rather than relying on public production. Along with States' flexibility to experiment must come more consistent accountability for results. As in the education and welfare programs discussed above, this Administration believes that a Federal focus on ultimate goals and outcomes, rather than micromanagement of processes, is needed to promote innovation and efficiency.

Limitations and Shortcomings of the Current System

Medicaid enrollment grew by almost 60 percent between 1980 and 1993, from 19.6 million person-years to 31.2 million. Much of the enrollment growth since 1987 was driven by federally mandated eligibility expansions, which increased the pool of eligible individuals well beyond those eligible for

Box 5-2. The State Children's Health Insurance Program

The State Children's Health Insurance Program (SCHIP) is a joint Federal-State program, driven by Federal incentives to improve the health care of low-income children while still affording States a great degree of flexibility in reaching this goal. SCHIP was established in the 1997 Balanced Budget Act, under Title XXI of the Social Security Act, and provides health insurance coverage to Medicaid-ineligible low-income children. Every State currently has a federally approved SCHIP program, but the design and scope of programs vary widely. Fifteen States and the District of Columbia provide SCHIP insurance through existing Medicaid programs, 16 States have separate programs, and 19 States use a combined approach. States are experimenting with providing health insurance to entire families and with using sliding copayment scales.

Like Medicaid, SCHIP is funded through Federal matching of State expenditure, with poorer States eligible for higher match rates. In fiscal 2002 the Federal Government reimbursed individual States for between 65 and 84 percent of the cost of providing health insurance under the program. In addition to providing a substantial portion of the funding, in fiscal 2002 the Federal Government will use awards, based on the participation of former TANF recipients in Medicaid and SCHIP, as incentives for States to insure low-income children.

AFDC by raising income limits. Although those receiving TANF continue to be eligible for Medicaid, PRWORA severed the link between cash assistance and Medicaid enrollment. Since 1993, Medicaid enrollment has grown at a much slower rate, reaching 36.9 million in fiscal 2001, and is projected to grow by an average of only 1.9 percent a year over the next 5 years. Federal Medicaid expenditure, on the other hand, is projected to grow at an annual average rate of almost 9 percent, from \$159 billion in fiscal 2003 to \$206 billion in fiscal 2007.

These expansions to families with higher and higher incomes appear to have had diminishing effectiveness, both in improving health and in reducing the number of uninsured. One unfortunate side effect of the current system of publicly provided and publicly produced health insurance is the crowding out of private insurance: the existence of public insurance provides a disincentive for private employers to offer insurance to those eligible for the public program. Research shows that many of those to whom Medicaid eligibility was extended during the broad expansions of the late 1980s and early 1990s already had access to private insurance. Researchers

estimate that only 27 percent of the children made newly eligible for Medicaid between 1987 and 1992 were uninsured in 1987, and that almost half of those newly eligible may have lost private coverage. In fact, as the fraction of children eligible for the program rose from 15.2 percent in 1987 to 21.8 percent in 1996, the fraction of children who were uninsured not only failed to decline but rather increased, from 12.9 percent to 14.8 percent. This experience illustrates the potential pitfalls of expanding public programs without considering potentially offsetting responses in private markets.

There is other evidence that mandated expansions of this form are not the most efficient way to improve the health of low-income families. A more diverse population of patients is likely to have differing needs, making it more difficult for a one-size public insurance package to fit all. One symptom of the inefficiency of the current system is the failure to enroll all eligible children: nearly a quarter of uninsured children are eligible for Medicaid, and many more are eligible through SCHIP. Although Federal laws explicitly guarantee continued Medicaid coverage for many of those leaving welfare, researchers found that 49 percent of women and 29 percent of children lack health insurance 1 year or more after leaving welfare. Confusion about eligibility, the effort required to reapply for Medicaid after leaving welfare, and stigma may contribute to this lack of health insurance among former welfare recipients.

Fostering Market-Based Health Insurance

Greater flexibility is allowing States to address these shortcomings in varied and innovative, market-based ways. By increasing the access of low-income families to private insurance markets rather than trying to provide the same public health insurance to all, the Federal Government can promote the health of all citizens without a monolithic, slow-acting, and inefficient bureaucracy. States have requested waivers and demonstration projects to experiment with other means of provision and have highlighted the potential gains to such approaches, empowering patients and providers to choose the best health insurance options at the best price through unfettered markets. The process of applying for waivers used to be quite cumbersome for State agencies, as was the oversight of waiver programs for their Federal counterparts. The goal of the 2001 Health Insurance Flexibility and Accountability (HIFA) Demonstration Initiative is to increase State access to Section 1115 Medicaid and SCHIP waivers, simplify the waiver process, and create renewed interest in working with private insurance markets to provide health insurance to low-income individuals. The HIFA initiative encourages States to use available Medicaid and SCHIP funding to develop comprehensive health insurance coverage approaches. This offers States greater flexibility in designing benefit packages and cost sharing in exchange for increasing

coverage, particularly in support of private health insurance. Even without HIFA, the Administration has already approved over 1,400 waivers and State plan amendments through other programs. These waivers and amendments have already made an additional 1.4 million Americans eligible for health insurance and expanded coverage for over 4 million more, and the Department of Health and Human Services has cleared application backlogs for State plan amendments dating to the mid-1980s.

This use of Medicaid waivers parallels that of AFDC waivers before TANE. Since 1981 the Centers for Medicare and Medicaid Services (CMS, the agency formerly known as the Health Care Financing Administration) has granted over 250 home and community-based services waivers, which cover budget-neutral but previously uncovered services for Medicaid-eligible individuals who would otherwise be institutionalized. In 2001, 15 States were running statewide health care reform demonstrations under Section 1115 waivers. These waivers allow States to change provisions of their Medicaid and SCHIP programs in order to experiment with program improvements, provide coverage to groups not eligible under current law, or investigate an issue of interest to the CMS.

States are using these waivers to experiment with different methods of health care delivery. The waivers offer the most flexibility when used to extend coverage to “optional populations.” These are groups that States may use Federal Medicaid funds to insure, but whose coverage is not a condition of Federal funding. Because they often have higher incomes than other Medicaid recipients, these recipients are more likely to be employed and therefore to have access to employer-sponsored health insurance. Enabling them to purchase coverage through their employers is less likely to crowd out private provision than is public Medicaid insurance. States may choose to offer this insurance under their existing Medicaid plans, under group plans, or through other sources of the States’ choosing, as long as the coverage meets Federal cost and quality guidelines.

States have long had the option of using Medicaid and SCHIP funds to help eligible individuals purchase private health insurance through their employers. However, in part because of administrative and operational complexities, very few States were able to take advantage of this option. Massachusetts helps employees pay private insurance premiums through its own premium assistance program. Kansas provides small businesses with a \$35 health insurance tax credit for every employee to whom they provide coverage. The Administration’s HIFA model waiver initiative is designed to give States program flexibility to support approaches that increase private health insurance coverage options. HIFA quickly generated State interest in exploring other ways to use employer-sponsored insurance to provide coverage to Medicaid-eligible populations. The Department of Health and Human Services has already approved one such waiver for Arizona.

States are also using market mechanisms to expand access to health insurance through other Federal laws, such as the Health Insurance Portability and Accountability Act of 1996, and through high-risk health insurance pools. Both are discussed in Chapter 4. Each uses market mechanisms to set prices and expand access, while empowering individuals to choose the plans that suit them best.

State flexibility can also promote cost containment without sacrificing quality. Medicaid expenditure grew dramatically between 1988 and 1994, primarily because of cost increases and issues of program integrity, but partly from the eligibility expansions and enrollment increases discussed above. In an effort to control costs, States have enrolled an increasing fraction of Medicaid recipients in private health insurance programs. Fifty-four percent of Medicaid recipients were enrolled in some form of managed care in 1998. Other States are experimenting with directly providing care through public clinics and community health centers (Box 5-3). Although these measures have helped States (and the Federal Government) contain costs, continued innovation in cost containment is still greatly needed, as is flexibility to experiment.

Federal officials have expressed concerns about State financing practices that increase Federal Medicaid spending without increasing health insurance coverage. Recent studies by the Inspector General of the Department of Health and Human Services and by the Congressional Budget Office have identified provider payment policies that have allowed billions of dollars in Federal Medicaid funds to be used for purposes other than those intended, including nonhealth expenditure. The Administration has taken steps to increase State accountability while also increasing State flexibility.

Although the provision of health care poses challenges not seen in other safety net programs, the lessons drawn can inform a wide range of policies. By setting goals based on outcomes, promoting innovation, and rewarding achievement, the Federal Government can create a lasting institutional structure that adapts to the rapidly changing health care environment without saddling States and providers with cumbersome and quickly outdated conditions and regulations.

Box 5-3. Community Health Centers

The Community Health Center (CHC) program is a Federal grant program that offers funding to local communities for the provision of family-oriented primary and preventive health care services. In fiscal 2001 the program funded services to 10.5 million people living in medically underserved rural and urban areas throughout the country. In the last decade there has been a significant increase in the number of access points, primary care providers, and people served, as well as in appropriations; more than 3,300 CHC sites are now in operation, providing essential services that improve the health status of these underserved populations. To ensure that more communities benefit from the care provided by these centers, the Federal Government will expand the program to 1,200 more sites over the next 5 years, serving millions of additional patients. CHCs are discussed in more detail in Chapter 4.

Conclusion

Creating efficient, high-quality public programs requires balancing freedom against responsibility, and local needs against national interests. By tying Federal funds to meeting program goals, but not tying the hands of willing and able providers, Federal dollars can be stretched further and the quality of services provided can be higher. Rewarding innovation and requiring success can bring out the best in public and private providers alike.

The United States' federal system provides unique advantages for getting the most out of public spending. Competition among States and localities and public and private providers encourages the efficient use of public funds. Accountability for results can be achieved without rigid and burdensome Federal dictates. This Administration believes that it is the role of the Federal Government to create the infrastructure—including high-quality data, a level playing field, and incentives that promote the efficient use of taxpayers' money—that makes such competition and accountability possible.

Building Institutions for a Better Environment

The United States has achieved dramatic improvements in environmental quality over the past 30 years. Toxic releases have been reduced since they were first widely reported in 1987, waters safe for fishing and swimming have doubled, and air quality has improved markedly. This trend toward a cleaner, healthier environment, repeated in many of the world's other developed countries, is reflected in various indicators of environmental quality, including measures of sulfur dioxide, lead, and carbon monoxide emissions. Box 6-1 shows how emissions of these and other air pollutants have fallen significantly in the United States, with similar gains in a host of other countries.

These improvements are the result of policies that sought to address some of the most obvious risks to human health posed by air and water pollution, leakage from hazardous waste sites, and unnecessarily damaging mining and other extractive practices. In these early initiatives, the benefits often far outweighed the costs. Now that most of the largest and most glaring environmental problems have been tackled, however, the gains to be expected from further measures have become less obvious and more contentious. Meanwhile competition for resources and for the attention of policymakers and concerned citizens is as keen as ever. Medical research, national security, education, capital investment, and consumption all make valid claims on both government and private resources. As the environmental issues we address become ever more complex, research and careful analysis of both benefits and costs are required to formulate responsible policies that will improve Americans' well-being and are cost-effective.

Put another way, the task now before us is to build the right institutions to address these increasingly thorny environmental issues. For example, there is evidence that further improvements in air quality would improve health and reduce mortality, but these improvements might be extremely expensive. Similar tradeoffs are associated with reductions in certain toxic substances, such as arsenic in drinking water and mercury from the burning of coal. Although the health benefits from further reductions in these pollutants are surely desirable, the associated expense might be better directed toward alleviating other problems with the potential for even larger reductions in health risks. Ongoing efforts to protect endangered species, maintain biodiversity, and preserve ecosystems—all of which can influence long-term land use decisions and short-term economic activity—could pose tradeoffs between the welfare interests of current and future generations. Finally, concern over

Box 6-1 Trends in National and International Environmental Quality

Some of the most dramatic improvements in environmental quality have occurred in the air we breathe (Chart 6-1). The 1970 Clean Air Act Amendments identified six common, nationwide air pollutants for which emission limits were needed in order to achieve certain ambient concentration levels based on health criteria. Since the law was passed, emissions of most of these “criteria air pollutants” have declined significantly. Perhaps the most impressive achievement is the near elimination of lead emissions, which by 1998 were only 2 percent of their 1970 level.

One criteria air pollutant whose emissions have not fallen is nitrogen oxides, and one might be tempted to conclude that environmental quality with respect to this pollutant has gotten worse. But in fact the story of nitrogen oxides regulation highlights the importance of using the appropriate metrics in judging environmental quality: although emissions of a pollutant are often reported, it is ambient concentrations in the air we breathe that affect us directly and are the target of most environmental regulation. In the case of nitrogen oxides, and indeed for all criteria air pollutants, average national concentrations have fallen in the past 20 years (Chart 6-2).

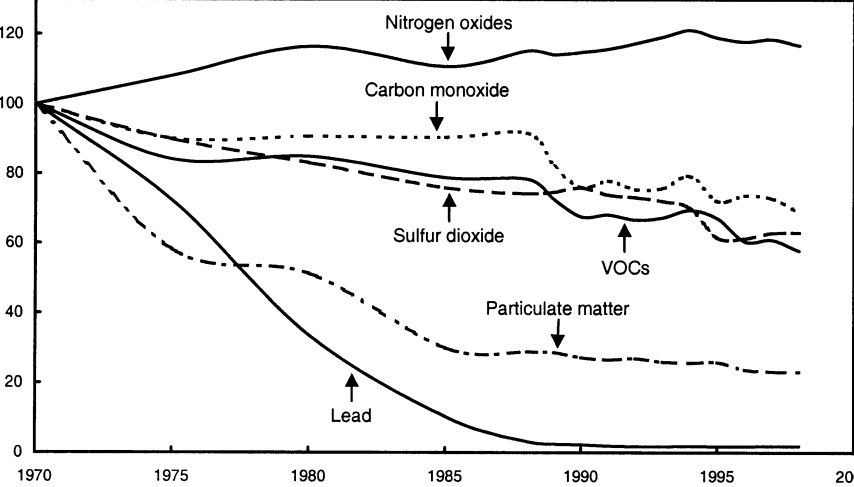
In addition to these reductions in criteria air pollutants, regulations and voluntary actions on the part of companies have resulted in substantial reductions in 188 toxic air pollutants that are either known or suspected to cause cancer or have other serious health effects. Nationwide emissions of these pollutants in 1996 were 23 percent below levels measured earlier in the decade. Concentrations of some of these toxic air pollutants have been reduced even more dramatically.

For many pollutants, such as sulfur dioxide, trends in the United States mirror those in other industrialized countries (Chart 6-3). The downward trend in such emissions is particularly impressive given the substantial growth in national income over the same period. Although it is sometimes assumed that economic growth leads to environmental degradation, studies show that environmental improvements usually accompany national income growth at higher levels of income, an observation that the chart supports.

Chart 6-1 **Emissions of Major Air Pollutants**

Emissions of most major air pollutants have fallen, some spectacularly, since the passage of the 1970 Clean Air Act Amendments.

Index, 1970 = 100

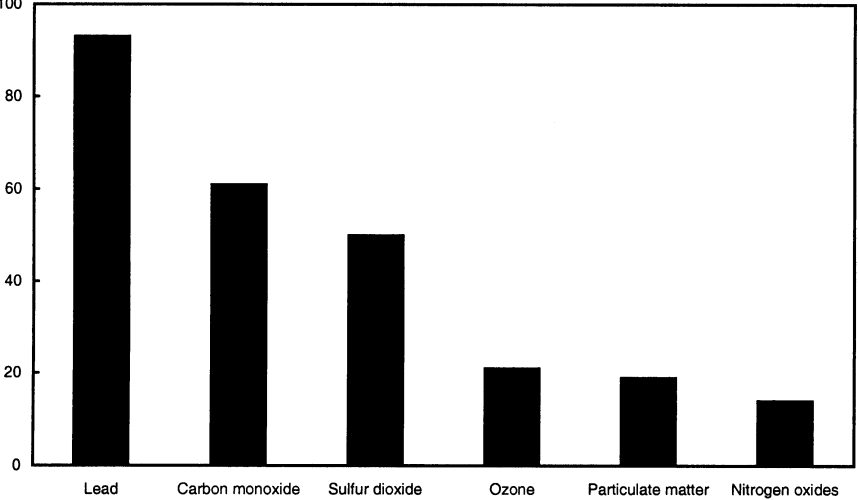


Note: VOCs are volatile organic compounds. Particulate matter refers to particles no greater than 10 micrometers in diameter and does not include miscellaneous or natural sources.
Source: Environmental Protection Agency.

Chart 6-2 **Reductions in Average Ambient Concentrations of Major Air Pollutants, 1981-2000**

Atmospheric concentrations of all six major pollutants have declined over the past 20 years.

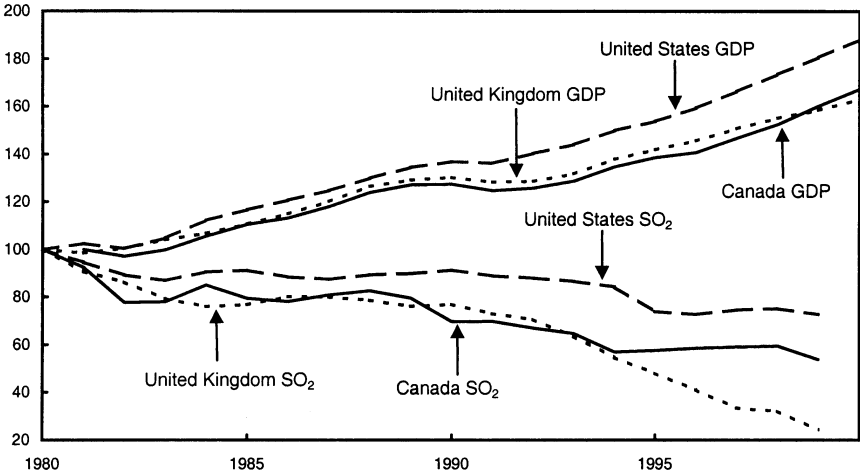
Percent reduction



Note: Data for particulate matter (as defined in Chart 6-1) begin in 1991. Reduction for ozone is that of one-hour levels.
Source: Environmental Protection Agency.

Chart 6-3 Sulfur Dioxide Emissions and GDP in Canada, United Kingdom, and United States
Sulfur dioxide emissions have declined in the United States and other countries alongside substantial growth in GDP.

Index, 1980 = 100



Note: GDP data for Canada begin in 1981.

Sources: Department of Commerce (Bureau of Economic Analysis), EMEP Program (Co-operative Program for Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants in Europe), Office for National Statistics (United Kingdom), and Statistics Canada.

potential climate change poses perhaps the greatest challenge. Sound climate change policy requires striking a balance not only between the well-being of current and future generations, but across countries as well. Choices must be made in the face of considerable scientific uncertainty and alongside competing concerns about energy security and diversity of fuels.

In many of these issues, the debate is frequently cast in terms of a tradeoff between environmental protection and economic growth. Yet the two are not necessarily mutually exclusive. As a society becomes more affluent, it is likely to demand a cleaner and safer environment. Prosperity also allows us to commit ever-increasing resources to environmental protection and to the development of science and technology that will lead to both future growth and a better environment. Indeed, empirical evidence suggests that growth eventually goes hand in hand with environmental improvements.

The design of appropriate institutions plays an important role in improving environmental quality; in particular, flexible approaches to environmental regulation can increase the benefits and lower the costs relative to alternative schemes. Such approaches often allow businesses to pursue established environmental performance goals or emission limits in the ways that

they find most effective, rather than following specific, detailed government mandates. This flexibility encourages innovation and the development of cleaner technologies. Over time, flexible approaches and other programs that promote technological innovation offer the promise of less pollution at even lower costs. The President's National Energy Plan, for example, builds on these ideas by encouraging both increased flexibility in regulation and the development of clean technologies.

Flexible programs also often involve a smaller, less costly regulatory and compliance apparatus. In place of lengthy wrangling and resorting to legal action between business and government over the interpretation and applicability of particular rules, requirements, and regulations, flexible approaches allow markets, financial incentives, and business-to-business transactions to efficiently allocate resources with minimal government supervision.

By institutions we mean not only the formal rules, regulations, markets, monitoring, and administrative features developed for environmental protection, but also the informal knowledge, experience, and norms that are essential for effective outcomes. Institutions of this kind that embody the flexible approaches described above do not appear overnight. Part of the challenge for environmental protection is designing and building the best institutions for the various problems we confront today, but another part is carefully constructing those institutions so that they can evolve to deal with emerging problems tomorrow. In exploring ways we can build institutions for a better environment, this chapter considers the pros and cons of alternative flexible mechanisms such as tradable permits, tradable performance standards, and emission charges. Several case studies of alternative schemes then illustrate these mechanisms in practice. Finally, we consider how this experience can be applied to the pressing environmental concern over the potential threat of climate change. We begin by briefly examining the motivation behind government involvement in environmental protection.

The Government's Role in Environmental Protection

At a basic level, environmental amenities have characteristics that frequently make them more of a public than a private responsibility. First, many environmental resources—notably the atmosphere, the oceans, and underground aquifers—are shared without becoming the exclusive property of anyone. Second, how one individual or business chooses to use air, water, and land resources influences the value of those common resources for many others. For example, marine fisheries are an important food source, but

excessive commercial fishing reduces the ability of a fish population to reproduce and provide more fish next season. Coal combustion provides an inexpensive and reliable source of energy, but the resulting emissions of sulfur dioxide (SO₂) increase the acid content of lakes and forest soils. Lead in gasoline is a convenient catalyst for boosting automobile performance, but it has adverse effects on children exposed to the consequent emissions from vehicles.

Economists refer to these environmental resources—healthy fisheries, healthy lakes and forests, and clean air—as public goods, and to the unintended, adverse effects resulting from the use of those resources as externalities. More broadly, externalities are the uncompensated effects of the activities of one individual or group on another: because these effects have no financial consequences for the individual or group undertaking the activity, they are external to the market. For example, until the government intervened, those who overfished a fishery did not bear the cost of that depletion to other fishermen and consumers; the power plants that emitted SO₂ had no financial incentive to reduce those emissions; and the refiners and users of gasoline faced no constraints on their use of lead as a catalyst. All these consequences were external to the market transactions that caused them and in some cases were not even appreciated at first. Even when they are identified and understood, however, such externalities by themselves are not necessarily a cause for government intervention. So long as the externality is identified, the individuals affected can, in theory, negotiate a solution. In our examples, some fishermen could have paid others not to overfish, the users of acidifying lakes and forests could have paid power plants to reduce SO₂ emissions, and communities could have negotiated with refineries to reduce the lead in gasoline.

The improbability of such solutions in the real world, however, highlights the fact that the corresponding problems, and environmental issues more generally, all involve public goods to some degree. This complicates arriving at a privately negotiated solution, because it is difficult to exclude those unwilling to pay to help solve these problems from enjoying the benefits of the improved resource. The productiveness of the fish stock, the recreational and commercial value of lakes and forests, and the health improvements from reduced lead emissions are all benefits that many if not all people can enjoy simultaneously and that are difficult to exclude people from enjoying. Under these circumstances, no single individual has the private incentive to negotiate a socially beneficial solution, because most of the benefits go to others. Nor is it easy for groups of individuals to band together informally to pursue a solution, because each has an incentive to “free ride,” allowing others to take care of—and pay for—the problem. Here the government can play an important role by representing the interests of a large group of individuals and compelling all those interested to share in the cost.

Measuring the Benefits and Costs of Environmental Protection

Rectifying an environmental problem—pollution in a river, for example, or depletion of a fishery—requires choosing both the appropriate level of control or use and the institution best suited to implement the controls. The level of control for many pollution problems has traditionally been set with an eye toward benefits. A prime example is air quality, where the Supreme Court recently upheld a decision that national air standards must be set to protect the public health without regard to costs, as set forth in the 1970 Clean Air Act Amendments. At the time this and other early statutes were passed, it may have appeared that the benefits were desirable at any cost, or that the costs were low, removing the need to consider them. However, as production technologies have become increasingly clean, the further reduction of pollution has become more difficult, and costs have risen. As a result, concern over costs has entered the regulatory process: levels of control on hazardous air pollutants are based not only on health concerns, but also on what control technologies are available. This means that consideration is given to whether the level chosen is feasible and cost-effective enough that someone has already developed technology for it. Costs also play a role in some fishery management policies, where the permitted annual harvest is set to maximize the sustainable catch.

Comparing the benefits and the costs of environmental policies is important because of the many competing needs for public and private expenditures. The optimal level of environmental protection is that where the benefit associated with one more unit of the resource equals the cost of providing it, with both benefits and costs appropriately added up across all individuals and over time. What should we include in our cost and benefit measures? On the cost side, most expenses associated with environmental protection arise from the use of marketed goods and services, making calculations relatively straightforward. For example, it is estimated that the recent decision by the Environmental Protection Agency to lower the acceptable level of arsenic in drinking water from 50 to 10 parts per billion will impose a total annual cost of more than \$200 million. This \$200 million will then be unavailable for other private and public activities—including other health and environmental programs. This therefore represents the cost of the program, which can then be compared with the benefits. Note that in the arsenic case—as well as in two of the case studies later in this chapter—concern over the distribution of costs and benefits was a particularly thorny issue, even though in theory it should be possible to make everyone better off when the overall benefits outweigh the costs.

The choice of policies and institutions to be used in achieving the environmental objective also plays an important role in determining costs. For example, cost estimates associated with implementing the Kyoto Protocol vary by orders of magnitude, depending on assumptions about the effectiveness of trading institutions. These trading institutions allow countries with higher abatement costs to seek out reductions in other countries with lower abatement costs. Because certain institutions—specifically, those that provide flexibility—offer the opportunity to achieve environmental goals at lower cost, it is important to understand the differences among the major types of environmental regulation, to which we return below.

On the benefits side, gains from environmental protection are often divided into two categories: use value and nonuse value. Use value refers to benefits that occur when individuals come into direct contact with the protected environment. These benefits may be associated with marketed goods and services, such as admission or transportation fees, or nonmarketed activities such as hiking, swimming, camping, or just looking at a beautiful natural landscape. They also include the health consequences of breathing cleaner air and drinking cleaner water. Nonuse value, which often involves nonmarketed goods and services, refers to the less tangible benefits that arise from individual preferences with respect to environmental amenities, as distinct from their direct use. This includes the value derived from knowing that a resource has been maintained and will be available to future generations, or to oneself if one should ever decide to use it.

Use values associated with marketed goods and services can often be estimated from observed behavior. For example, the willingness of people to pay to use a national park—as measured by the entrance fees they actually pay, or their travel expenditure to get there—can be used to estimate the value they associate with the park. Wage studies measuring the pay difference between low-risk and high-risk jobs can be used to infer the value associated with prolonged life, which can then be used to evaluate health-enhancing environmental proposals. Expenditures on water filters or bottled water can be used to value a reduction in water pollution. Nonuse values, as well as use values that are not associated with market activities, are more difficult to estimate accurately. Typically, individuals are surveyed and asked to place a dollar value on hypothetical levels of environmental quality, such as better visibility in scenic areas or enhanced protection of wilderness, ecosystems, and biodiversity. This approach is still a subject of scholarly research.

Types of Environmental Regulation

The policies and institutions used to achieve an environmental goal often have significant consequences for the associated cost. As environmental regulation has evolved, businesspeople and policymakers have worked together to find more flexible approaches that achieve the same goal at significant savings. These approaches range from standard tradable permit and fee programs, to more complex tradable performance standards and hybrid permit/fee programs, to more informal, flexible regulatory arrangements.

Command-and-Control Approaches

Traditional regulations for environmental protection, such as those legislated under the 1970 Clean Air Act Amendments, focused on developing specific technology and performance standards for pollution sources to meet. Technology standards mandate specific equipment that sources must use to control emissions, whereas performance standards mandate a limit on emissions allowed by each source. Because technology standards typically require the same technologies for all sources, and performance requirements require the same level of emission reductions or emission rates at all sources, both these approaches fail to take advantage of differences in the circumstances of each source. In particular, they fail to encourage more reductions where the cost of such reductions is low, and fewer reductions where the cost is high. Over the years, numerous studies have documented the added expense of limiting this kind of flexibility, with cost estimates of traditional regulation ranging from as little as 7 percent to as much as 2,200 percent (that is, 22 times) more expensive than an efficient, flexible program.

Standard Market-Based Approaches: Permit Trading and Fees

In the cases of marine fisheries, SO₂ emissions, and leaded gasoline noted earlier, market-based policies have been used to provide greater flexibility in meeting particular environmental goals. Fishermen, power plants, and gasoline refiners were required to hold a volume of permits (also referred to as allowances or quotas) equal, respectively, to the volume of fish caught, emissions created, or lead blended into gasoline. These permits were distributed on the basis of either past or current production. Unlike the earlier, command-and-control approaches, however, these permits could be freely traded, creating highly efficient markets in which firms holding more permits than needed could sell them to others or, in some cases, hold onto them for future use.

These permit markets have many advantages. They ensure that the most valuable uses of the affected resources are encouraged, they maximize economic activity and growth consistent with a given level of pollution reduction, and they encourage innovation in solving the environmental problem at hand. In addition, the market price of the permits provides a clear signal about the economic value of the environmental resource, which can then be used for both business planning and policy evaluation. Finally, although the permits in these programs were predominantly distributed freely to predetermined stakeholders, the government could choose in future programs to sell the permits, generating revenue that could be used to reduce taxes on capital and labor, thus improving the efficiency of the tax system.

Emission fees, where businesses pay a fee for each unit of emissions rather than buy and sell permits, share many of the advantages of tradable permits. They provide an incentive to engage in only the most valuable uses of the environmental resource, send a clear signal about its economic value, and generate revenue that can be used to reduce other taxes. Emission fees, however, provide greater certainty to businesses because the price associated with emissions (the charge rate) is fixed. In contrast, because tradable permits are in fixed supply, their price can fluctuate to reflect changes in demand—sometimes substantially. As an example, a market for nitrogen oxides (NO_x) emission permits was established in 1994 in the area around Los Angeles. At the end of 1999, permits for use in 2000 traded for around \$2 a pound, but by August 2000, during California's emerging electric power crisis, they sold for as much as \$50 a pound. Of course, the greater price certainty associated with emission fees comes at a cost: under an emission fee the actual level of emissions can fluctuate. Thus emission fees make it trickier for regulators to achieve a targeted level of emissions. Tradable permits also allow an administratively easier redistribution of the value associated with emission rights. Revenue from a permit fee can be rebated and redistributed, but this requires the government to distribute money after collecting fees, thus involving the government in myriad financial transactions. Under a tradable permit system, permits can be distributed in advance of the actual program, and financial transactions need occur only among private firms and individuals. Perhaps because of this, emission fees have received little attention in the United States, despite their considerable popularity in other countries (Box 6-2).

An intriguing possibility is the coupling of a tradable permit system with a fee-based “safety valve.” In this hybrid scheme, a regulatory agency operating an ordinary tradable permit program would create and sell extra permits on request at a fixed fee. If the fee were set above the typical trading price—for example, above the \$2 a pound price that prevailed before 2000 in the Los Angeles NO_x permit market—it would ordinarily not interfere with the permit market. However, in the event of an unusual demand spike like that

Box 6-2. Environmental Fees in Other Countries

Whereas the United States has tended to use tradable permits to encourage cost-effective reductions of pollutants, market-based environmental regulation in other developed countries has more commonly relied on fees, with particular focus on the transportation sector. For instance, in 1995 about 90 percent of the revenue from pollution control-related fees in 20 industrial countries came from fees on gasoline, diesel fuel, and motor vehicles. In the last decade, however, some European countries have developed fees specifically designed to reduce particular industrial pollutants.

In 1992 Sweden introduced a charge on NO_x emissions from large combustion power plants. This fee of 40 Swedish krona per kilogram of NO_x emissions, equivalent to about \$4 at the current exchange rate, was extended to smaller power plant boilers in 1996. Revenue from this fee is returned to the group of power plants that pay them in proportion to each plant's share of total energy production. This refund reduces the total financial burden on power plants from the fee. But the fee still provides an incentive to reduce NO_x emissions whenever the cost for each unit reduced is less than the fee. The Swedish government estimated that in 1995, as a result of the fee, NO_x emissions from power plants declined by 20 percent.

A Danish experiment with fees highlights one problem common to many existing environmental fees. In 1992 Denmark introduced a fee on carbon dioxide (CO_2) emissions by households, which was followed in 1993 by a similar fee on CO_2 emissions by industry. As a result of concern about the effect of these fees on Danish industrial competitiveness, the fees were altered in 1995 so that certain energy-intensive industries paid lower fees on CO_2 emissions than did less energy-intensive industries. Although this change had the desired effect of reducing the burden on the more energy-intensive industries, it also reduced the cost-effectiveness of the emission reduction scheme overall.

Firms facing CO_2 fees will reduce emissions up to the point where the cost of reducing another unit of emissions (that is, the marginal cost) equals the fee. Beyond that level it is cheaper to simply pay the fee than to further reduce emissions. Because different firms face different fees in Denmark, they should end up with differing marginal costs as well. This implies that the present arrangement is inefficient, because the total cost of the prevailing level of emission reduction could be reduced. Shifting some responsibility for emission reduction from firms facing high marginal costs to those facing lower marginal costs would lower the overall burden.

continued on next page...

Box 6-2.—*continued*

The Danish experience is not unique, however: throughout the industrialized world, environmental fees have frequently been accompanied by exemptions for particular products or industrial sectors. The goal of some of these exemptions, to reduce the burden of these fees on particular activities or sectors, can be achieved through other means that do not reduce the overall cost-effectiveness of the fee program: the revenue can be redistributed or rebated to program participants. The administrative and practical difficulties with such a redistribution point to an advantage associated with tradable permits: their initial allocations can be conducted in a way that alleviates burdens where desired.

resulting from the California energy crisis, the fee would provide additional flexibility and price stability, protecting both industry and the economy. In point of fact, California enacted something like this—whereby a reserve of NO_x permits would be available at \$7.50 a pound—after last year's permit shortage. Features like this have been used in the SO_2 trading program and in regulations for heavy-duty engines, both discussed below.

Other Flexible Approaches: Informal Markets and Tradable Performance Standards

In some cases it may be impractical to implement either an emission fee or a permit trading program. For example, monitoring actual emissions may be too expensive to make either viable. Emission fees also face opposition because they impose on regulated firms the burden of fee payments in addition to pollution control costs. At the same time, tradable permits may be impractical because the transactions costs associated with trading are too high, because there are too few potential buyers or sellers, or because different levels of sophistication among potential market participants are likely to lead to inefficiencies.

In these situations, alternative institutions can arise that approximate the efficiency of true market approaches by providing flexibility, but trade off some of the potential economic gains in the face of these practical constraints. One approach, discussed later in the Tar-Pamlico case study, is a less formal trading market. Another is a tradable performance standard.

The regulation of nitrogen oxides, particulate, and hydrocarbon emissions from various types of combustion engines provides multiple examples of

how a tradable performance standard can work. Since 1991, heavy-duty, on-highway engine manufacturers (who produce the engines used in trucks and buses) have been able to comply with some of these emission standards on new engines through a combination of averaging, banking, and trading—or ABT. This approach has been extended to emission standards for many other types of engines, including outboard boat engines, automobile and light truck engines, locomotives, and small nonroad engines such as lawn mowers.

A typical ABT program begins with a schedule of emission standards. For example, the NO_x standard for heavy-duty, on-highway diesel engines started at 6 grams per brake horsepower-hour for engines made in 1990, falling to 5 grams in 1991 and 4 grams in 1998. After 2004, even stricter standards will be applied. These are performance standards in the sense that they specify emissions (grams of NO_x) in relation to other outputs—in this case useful mechanical energy output measured in brake horsepower-hours. Engine manufacturers who lower their engines' emissions beyond the standard generate credits. The number of credits is related to how much lower the emissions are, over the life of the engine, than those for an engine that exactly meets the standard. With some restrictions, manufacturers that earn credits can use them to offset excess emissions from current-year engines that do not meet the emission standard (averaging), reserve them for similar use in future years (banking), or sell them to other manufacturers (trading).

Compared with a program that requires all engines to meet the same standard, these ABT programs make it possible to achieve the same (or lower) emissions at a lower cost. The banking element encourages manufacturers to overcomply in order to generate a stock of credits, providing flexibility in the future. This overcompliance reduces emissions below the standard in the current year. At the same time, the flexibility to produce some engines that do not meet the standard and others that surpass it—while achieving the standard on average—allows manufacturers to reduce emissions more among those engines where control costs are lower.

The program for heavy-duty, on-highway engines contains an additional flexibility mechanism called a nonconformance penalty. Manufacturers that fail to meet the standard, and fail to obtain credits from other sources, can choose to pay a penalty based on the degree to which their engines exceed the standard. As an example, in 1991 a manufacturer producing a heavy-duty diesel engine that was certified at 6 grams of NO_x per brake horsepower-hour (when the standard was 5) could have paid a penalty of about \$1,600 for each engine rather than seek out emission credits. The nonconformance penalty limits the maximum costs that can be incurred by manufacturers seeking to comply with the regulation, providing them an additional measure of financial certainty. True, unlike the ABT mechanisms, which can lead to

lower emissions than the required level, this kind of penalty (if used) allows emissions to rise relative to a program requiring strict adherence to the standard. However, this flexibility may actually allow the adoption of tighter standards, suggesting that such a straightforward comparison is not valid.

Myths About Flexible Approaches

Despite the demonstrated benefits of flexible programs, popular concern remains. Some of these concerns raise valid distributional and equity issues. The economic and environmental benefits of flexible programs are not always shared equally, and indeed, some stakeholders can end up worse off. But other concerns derive from misperceptions about how flexible approaches work. These misperceptions can be addressed by better information. Below we discuss some of the more common myths surrounding flexible approaches to environmental regulation, and counter them with rational economic explanations.

Myth #1: “It’s immoral to buy the right to pollute.”

A widely held belief is that it is somehow unethical or even immoral to allow firms to buy and sell the right to pollute. For example, it has been claimed that turning pollution into a commodity to be bought and sold removes the moral stigma properly associated with it, and makes pollution just another cost of doing business, like wages, benefits, and rent. Regarding climate change, it has also been asserted that an emission trading program may actually undermine the sense of shared responsibility that increased global cooperation requires.

Although it is difficult to refute arguments of a moral nature, claims such as these contain several flaws. Certainly it makes sense to maintain a moral stigma on pollution when polluters are making a discrete choice whether to pollute. However, in most cases the creation of some pollution is inevitable. Thus the question is not whether we will pollute, but rather how much. In this context it makes sense to evaluate pollution in terms of a tradeoff between the harm it causes and the cost of abating it—and tradable permits allow for this. Furthermore, arguments based on morality seem an inappropriate framework for the debate in light of the past achievements of tradable permits in reducing pollution. For example, it seems strange to debate the morality or immorality of the use of a tradable permit system to phase out leaded gasoline, given that such a system in the 1980s reduced atmospheric concentrations of lead more rapidly than anyone had anticipated, and at a savings of \$250 million a year. More generally, the premise

that environmental progress must be accompanied by sacrifice is not necessarily valid. Finally, the ability of a tradable permit program to make pollution an internal cost of business is actually very effective, because it forces polluters to incorporate the cost of their external environmental damages into their operating costs.

Myth #2: “Permit markets for pollution are unfair.”

It has also been claimed that a market-based system for environmental control is inherently unfair, allowing some participants (those for whom it is less costly to buy permits than to reduce their own emissions) to evade their obligations. For example, a proposal for an emission permit trading program for NO_x in the Netherlands met significant resistance in part because of policymakers’ concern that a free initial allocation of credits would benefit the most-polluting companies, while penalizing those that had been more proactive in limiting emissions. But those who oppose pollution permit markets on these grounds overlook the fact that trading usually makes all participants in a regulatory program better off, compared with the same program without trading. Consider the following hypothetical example: Suppose that company A would have to spend \$50 million annually to reduce its emissions as required by some new regulation, whereas company B could reduce its emissions by the same amount at a cost of \$5 million but is not required to do so. Trade in emissions would make both companies better off. If company A pays company B \$30 million in exchange for company B’s agreement to reduce its emissions in place of company A, company B would be better off by \$25 million, and company A would pay \$30 million rather than \$50 million to reduce emissions. Indeed, because trade is optional, its mere existence is evidence that trade is beneficial for both parties—if it were not, one party would opt out.

Along the same lines, it is often mistakenly assumed that emission trading somehow favors larger companies, allowing them to buy their way out of pollution reductions whereas smaller companies cannot. But in fact, smaller companies often benefit more from permit markets: because they may not have as many internal options for pollution reduction, the potential to buy emission permits gives them added flexibility. The mistaken assumption that emission trading favors large companies also ignores the distinction between the allocation of permits (and emission rights more generally) and their subsequent trading.

The allocation of permits provides an opportunity to assign responsibility for emission reductions in a way that addresses this concern. For example, one could issue proportionally more permits to smaller companies to reduce their burden. Or one could reward companies that have already reduced emissions by providing them with extra permits. The smaller companies, or

the ones receiving extra permits, would then be free either to use the permits themselves—forcing other companies to reduce more—or to sell them if they choose.

Moreover, almost no form of regulation (or, for that matter, of markets) is “fair” under all possible definitions. For example, consider a hypothetical industry in which some firms have invested in newer (more costly) equipment that is less polluting, whereas other firms still use older equipment that is more polluting. Suppose that the government now introduces a regulation requiring, explicitly or implicitly, that all firms in the industry use a third, new technology that is less polluting than either of the first two. Both companies will then have to spend money switching to the new technology. But not only will the firms that originally invested in the intermediate technology receive no benefit from having polluted less in previous years; they will in fact lose more money because they invested in this second-best technology that they now have to discard. Few would consider such a result fair—certainly these firms would not.

To take a real-world example, consider the United States’ upcoming ban on methyl bromide. Subsequent to the 1987 Montreal Protocol, participating developed and developing countries agreed to completely phase out the use of this ozone-depleting chemical by 2005 and 2015, respectively. Currently, California strawberry and Florida tomato production relies on methyl bromide to control for pests and weeds. Substitutes for methyl bromide are expected to be less effective and produce lower crop yields. Meanwhile, competing strawberry and tomato growers in Mexico can continue to use methyl bromide for an additional 10 years, thus allowing them to increase their imports to the United States, at the expense of U.S. production. Surely the U.S. farmers would not consider this form of traditional regulation fair.

Finally, those who believe it is unfair for some firms to purchase permits rather than reduce emissions or limit resource use sometimes overlook a feature of a fully tradable permit system that they themselves can take advantage of, to remove permits from the system. If they are unhappy that firms are buying permits in order to comply, they can simply purchase existing permits themselves and retire them, thereby reducing the number of permits available to those firms. This method has been used, for example, by people concerned about wetland preservation to buy water rights from agricultural users in Nevada.

In thinking about fairness generally, society first needs to determine what it believes is fair. Second, groups in society need to remember that those adversely affected by a policy change can in principle be compensated if it is felt that such compensation would make the policy more fair. Compensation can occur under any form of regulatory tool, whether traditional or market based.

Myth #3: “Tradable permits and other flexible mechanisms will never work in the real world.”

Flexible mechanisms do work, and we know this from real-world experience: the successful results of many different pollution abatement and resource management programs that have used them. These mechanisms have been shown to be a highly effective (but certainly not the only) means of controlling pollution and managing resources. The case studies below document this experience for a variety of environmental concerns. Although the setup and structure of these programs vary considerably, each has allowed for flexible methods of compliance. As a result, many have achieved their reduction and conservation goals at substantially lower cost than traditional command-and-control approaches. For these programs to work well, however, certain conditions must prevail; these are discussed in greater depth in the section on lessons learned, at the end of the chapter.

Myth #4: “Traditional regulation encourages technological innovation and adoption of new technologies more than do market-based mechanisms.”

As discussed above, the circumstances of some environmental issues may favor traditional regulatory approaches, including technological standards mandating the use of a specific technology, and performance standards, which require each firm to demonstrate a certain performance level, expressed as an emission rate per unit of input or output. However, the requirement to use a particular technology prevents firms from seeking out cheaper alternatives. And because individual firms are usually in the best position to find those cheap alternatives, it is likely that technological mandates retard innovation. By specifying compliance in terms of a fixed technology or performance level, both kinds of standards provide little incentive for ongoing improvements in pollution control techniques. That is, firms may get no benefit from improvements they might discover that would allow more emission reductions for the same price. Lacking this incentive, firms may not invest continuously in research and development to enhance environmental quality. Barriers such as these have contributed to declining private sector funding for environmental technology development once firms have met the established standards.

Flexible mechanisms, in contrast, encourage firms to constantly seek out the most cost-effective technology to reduce their pollution. Moreover, the wider technological choice that results from such research creates greater opportunities for still further innovation, which cannot be predicted or captured in a government-controlled technological mandate. One example demonstrating that flexible permit trading programs promote innovation is

the success of the Title IV SO₂ program established under the 1990 Clean Air Act Amendments. This program is discussed in greater detail in the case study below. Because the program did not impose a technological requirement, and consequently rewarded all emission reductions, firms began to experiment with blending the high-sulfur coal that many of them had been using with low-sulfur western coal. Blending worked far better than had been thought possible, resulting in low-cost emission reductions.

Because the SO₂ program also included a flexible banking mechanism, firms had an incentive to use these low-cost opportunities to reduce emissions substantially below the required levels. Excess emission reductions such as these are unlikely in programs that limit compliance to a fixed technology or performance level, because they provide no incentive for overcompliance.

As a second example, research shows that stricter building codes have had little effect on homebuilders' choice of insulation technology. On the other hand, higher energy prices and adoption subsidies (which pay homebuilders directly to use more energy-efficient insulation) would have had a much greater effect. In this case, flexible incentives would have led to the more rapid adoption of new technologies, where traditional regulation failed to do so.

Finally, fisheries have long been subject to command-and-control regulation, which, for example, set limits on the time spent fishing. There is strong evidence that, under this type of regulation, fishing operations built up excess capital: too many ships were acquired, and too much equipment was installed, in order to catch as many fish as possible in the short time allowed. In the case of the Federal surf-clam fishery, in contrast, tradable permits succeeded in reducing the number of ships and the amount of capital used, and thus led to a more efficient use of existing technology than the various size limits and time restrictions that they replaced. One of the case studies below discusses fisheries in more detail.

Case Studies in Flexible Environmental Protection

Recognizing that flexible approaches to environmental protection can work solves only part of the puzzle. The other part is identifying the right institutional arrangement for the environmental problem in question, and the right development path along which to build those institutions. Perhaps the best way to understand how flexible programs are put into place is to consider several examples. Below we review three such programs that use varying approaches to address different environmental problems.

The Sulfur Dioxide Permit Trading Program

History of Sulfur Dioxide Regulation

Sulfur dioxide, when released into the atmosphere, reacts with water, oxygen, and other chemicals there to form an acidic deposition known as acid rain. Acid rain has the potential to raise the acidity of lakes, resulting in fish kills; to reduce the alkalinity of forest soils, harming various tree species; and to degrade various other ecosystem functions. Studies have also linked SO₂ with degradation of visibility and with increases in fine particulate matter in the atmosphere, which can cause respiratory problems in humans. In North America, acid rain is a concern mainly in the northeastern United States, particularly in the Adirondacks and New England, and in southeastern Canada. The majority of SO₂ emissions come from industrial activities, although natural sources—volcanoes and sea spray—also contribute.

Historically, SO₂ pollution control has focused on fossil fuel-burning electric power generators, which are responsible for approximately two-thirds of all SO₂ emissions in the United States. The 1970 Clean Air Act Amendments, the first significant Federal air pollution legislation, led to the establishment of national air quality standards for permissible concentrations of SO₂ in the air. States were largely held responsible for meeting these standards in each local area through the development of a State Implementation Plan (SIP), specifying actions to be taken to bring the State into compliance. As part of their SIPs, States required some existing power plants and others not yet built to have high smokestacks, so as to disperse emissions over a wider area. However, because acid rain can sometimes fall hundreds of miles downwind from its source, tall stacks may actually have increased SO₂ concentrations at distant locations. The 1970 amendments also imposed New Source Performance Standards (NSPS), which applied only to new power plants. These standards set new coal-fired plants' maximum allowed emission rates significantly below the emission rates of existing plants.

In projecting States' future air quality, it was assumed that existing plants not meeting the NSPS would gradually be retired, following historical patterns. However, this assumption failed to account for the strong incentives that the rules themselves created to extend the lives of older plants, which were expensive to replace with plants meeting the NSPS. By 1975 it had become clear that, because older plants were continuing to operate longer than expected, many States would not be able to comply with the air quality standards within the mandated time period. As a result, the 1977 Clean Air Act Amendments extended the deadline until 1982 and tightened the NSPS in those areas unable to meet the original deadlines. These new NSPS rules

required coal-fired plants built after 1978 to remove a specified percentage of potential emissions. This, however, reduced the advantages of using low-sulfur coal as a means of compliance, because percentage reductions were still required regardless of the type of coal used. Thus regulations may actually have dirtied the air on balance, by encouraging utilities to burn high-sulfur coal and by strengthening the incentives to extend the lives of old, dirty plants. The NSPS requirements also raised fairness issues, as some industries (such as high-sulfur coal producers) benefited while others (such as low-sulfur coal producers) suffered losses. Also among the losers were those States, mostly in the West, that were already using low-sulfur coal to generate electricity and were growing rapidly.

The 1990 Clean Air Act Amendments

Because current controls were not successful at achieving the SO₂ emission reduction goals, a new acid rain program was launched under the 1990 Clean Air Act Amendments. Title IV of the amendments set a goal of reducing annual SO₂ emissions by 10 million tons from the 1980 level. To achieve these ambitious reductions, the law required a two-phase tightening of the restrictions placed on fossil fuel-fired power plants. Phase I, which began in 1995, affected 263 units at 110 mostly coal-burning electric utility plants located throughout 21 eastern and midwestern States. An additional 182 units opted into the program during the course of Phase I. Phase II, which began in 2000, further tightened annual emission limits on the larger, higher emitting Phase I plants and set emission restrictions on smaller, cleaner plants, some of which were fired by oil or natural gas.

To achieve these goals, the 1990 amendments directed the Environmental Protection Agency to design a trading program in SO₂ emission allowances. The program provides incentives for energy conservation and technology innovation that both lower the cost of compliance and increase pollution prevention. Under the program, units are allocated allowances based on their historical fuel consumption and a specific emission rate. The large size and relatively small number of plants made it easier for emissions to be monitored continuously, increasing the credibility of emissions accounting and simplifying verification of the achievement of emission reduction goals. The majority of allowances are allocated by the agency without cost to the recipient. However, every year a small fraction (about 3 percent) of allowances are held back and sold in an auction administered by the Chicago Board of Trade. The SO₂ program also has a reserve of allowances that provides firms with the opportunity to purchase additional allowances at a fixed price of \$1,500 (in 1990 dollars; this figure is adjusted each year for inflation). Each allowance permits a unit to emit 1 ton of SO₂ during or after a specified year. Allowances may be bought, sold, or banked for future

use. If a plant's annual emissions exceed the number of allowances held, the owners must pay a penalty of \$2,000 (in 1990 dollars, also adjusted for inflation) per excess ton of emissions. Violating units are also required to make additional future emission reductions. Trading is not restricted to utility plants; anyone may buy or sell allowances. For example, speculators have acquired some allowances in hopes of future price increases, and environmental groups and some individuals have acquired allowances in order to reduce emissions more than the law requires.

Results

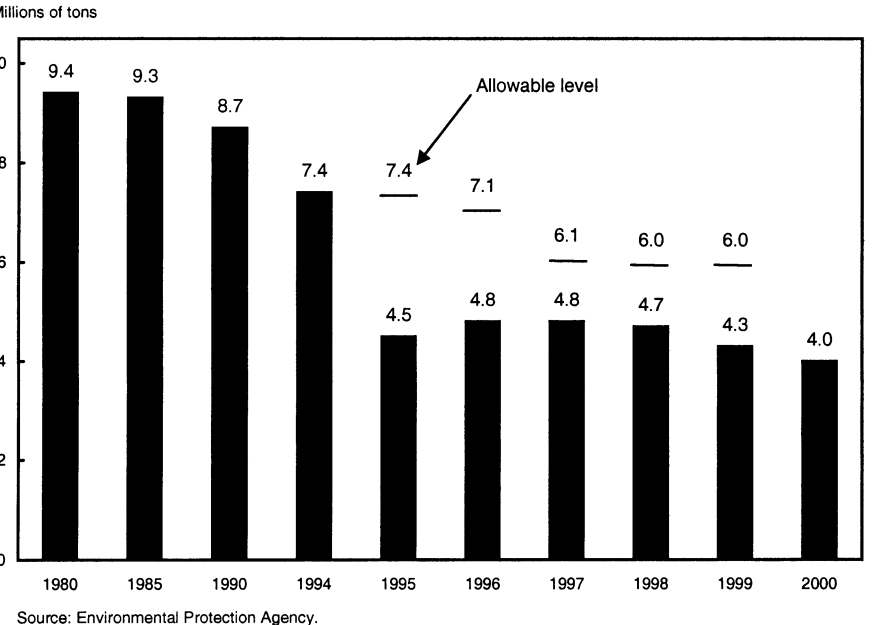
Participation in the trading program has been strong. Through the end of 2000, over 11,600 transfers had taken place, involving 111 million allowances. Approximately 59 percent of these (66 million) were transferred within organizations, and the remainder between economically distinct organizations. Both the number of transfers and the associated number of allowances have increased greatly since the program's inception (Chart 6-4). In the first year of trading (1994), 66 transactions took place, exchanging 0.9 million allowances between economically distinct organizations. In 2000, 2,889 transactions resulted in the transfer of 12.7 million allowances.

The trading program has lowered emissions substantially while yielding considerable cost savings, especially compared with the previous, command-and-control regime. Emissions data indicate that in the program's first target year (1995), nationwide emissions by the units required to participate in Phase I were reduced by almost 40 percent below their required level (Chart 6-5). This overachievement was encouraged by the provision allowing firms to bank credits for future use when they reduce emissions in excess of current requirements. The General Accounting Office projects that, compared with the command-and-control approach, the allowance trading system could save as much as \$3 billion a year, or more than half the total cost of meeting the standards. Some economists, however, believe this estimate overstates the program's cost reduction. As low-cost options for emission reduction emerged that had not been foreseen in 1989, there has been over time a clear downward trend in the predicted cost of the program. This primarily results from the fact that, as it turned out, low-sulfur coal could be substituted for high-sulfur coal much more easily than had been anticipated at the program's inception. On the other hand, this less costly method was likely adopted, in part, precisely because of the flexibility allowed for in the SO₂ trading program. A command-and-control program, whether based on performance standards or on technological requirements, might have afforded much less opportunity to take advantage of this low-cost alternative. In this case, flexibility allowed adoption of the optimal, most efficient solution available.

Chart 6-4 Sulfur Dioxide Allowances Traded Between Economically Distinct Organizations
Trading activity in the sulfur dioxide emissions permit trading program has risen almost without interruption.



Chart 6-5 Emissions from Phase I Facilities in the Sulfur Dioxide Trading Program
Sulfur dioxide emissions from the original 263 units were far below allowable levels during Phase I.



Tradable Quotas in the Alaskan Halibut and Sablefish Fisheries

The preceding example focused on a national pollution problem, which required a national solution. But flexible approaches have also been successfully applied to local and regional environmental problems, as the next two case studies demonstrate.

Fish in the coastal waters and open seas are the private property of no one; they are there to be caught by anyone with a boat, a fishing permit, and the necessary equipment. This public access nature of saltwater fisheries results in economic inefficiencies. If fish could be fenced in and counted like cattle, property rights could be allocated for each fish, or for a school, or for an entire fishery. Owners of such rights would have an incentive to limit their catch, so that enough fish are left each year to ensure the sustainability of the fish population, and thus of the owners' profits, in future years. However, because rights to individual fish or to fisheries cannot be established, and no one private fishing operation can control the actions of others, it is often in each fisherman's best interest to catch as many fish as possible, as quickly as possible, before the others do. As a result, many fisheries have suffered from an excess of capital, participation, or effort given the amount of fish available. This, in turn, has led not only to overfishing and depletion of the resource, but also to increased conflict and hostility, undesirable price and market effects, and increased physical danger to fishermen.

Regulation of U.S. fisheries was established in 1976 with the passage of the Fishery Conservation and Management Act (later renamed the Magnuson-Stevens Fishery Conservation and Management Act). Since then the act has been amended more than a dozen times, marking significant changes in its course and emphasis. The 1996 amendments emphasized the goal of biological conservation of fish stocks and protection of habitats, along with other resource management objectives. For the first time, the amendments made the prevention of overfishing an enforceable obligation on the part of the Federal Government.

In some fisheries, authorities have sought to achieve these goals through the use of a market-based output control mechanism called individual fishing quotas (IFQs, sometimes also called individual transferable quotas). An IFQ is defined as "a Federal permit under a limited access system to harvest a quantity of fish, expressed by a unit or units representing a percentage of the total allowable catch (TAC) of a fishery that may be received or held for exclusive use by a person." Ideally, regulators should set the TAC equal to the socially optimal catch (that is, the maximum sustainable catch). To date, IFQs have been adopted in a number of U.S. fisheries, such as those for surf clams and ocean quahogs, South Atlantic wreckfish, and Alaskan halibut and sablefish. Such mechanisms have also been used in other countries, including Iceland and New Zealand.

The experiences of the Alaskan halibut and sablefish fisheries are particularly illustrative. When the IFQ program was launched in 1995, the estimated coastwide biomass of halibut was above the 25-year average, but was declining and expected to continue to drop in the future. As of 1999, sablefish biomass had been declining since 1986 and was 30 percent below the recent average. Before the IFQ program, efforts to maintain fish stocks took the form of traditional management: regulators set an annual TAC on commercial fishing of halibut and then attempted to achieve the TAC through a combination of area, season, and gear restrictions. These regulations resulted in a host of problems, such as gear conflicts, fish kills due to gear lost at sea, discarded fish mortality, excess harvesting capacity, declines in product quality, safety concerns, unmonitored catch of regulated species in other fisheries, and economic instability within both the fishing industry and fishing communities. Evidence of some of these problems can be seen in the extremely short annual season for halibut fishing: from 1980 to 1994 the season averaged only 2 to 3 days in the management areas responsible for the majority of catches.

IFQ Design

Consideration of limits to entry began in 1977, but because of implementation delays, IFQs for halibut and sablefish were not approved until the end of 1991 and were implemented only in 1995. A primary objective of the program was to eliminate the fishing derby associated with the shortened season and the limit on the catch. This frantic race for fish was not only unsafe but inefficient as well. To increase their individual catch, some fishermen brought in additional vessels, and this imposed higher costs both on themselves and on others. These higher costs included increased harvesting and processing costs and decreased product prices, as well as the potential for higher debt service, additional unmonitored fish mortality, and increased accidents.

The design and management of the IFQ programs for Alaskan halibut and sablefish are largely the same. Landing data for halibut are collected by individual State governments and then forwarded to the International Pacific Halibut Commission (IPHC). Catch data for sablefish are collected by the individual States and the National Marine Fisheries Service (NMFS). Both programs require IFQ owners to be on board the vessel when the IFQ is being fished. They also set limits on the accumulation and transfer of quota shares. No person may own more than 0.5 percent of the total quota share for halibut, or 1 percent of the share for sablefish, in particular areas. Transferability is restricted across vessel size and across vessel categories.

IFQs were allocated to vessel owners and leaseholders who had verifiable commercial landings of halibut or sablefish during any of the eligibility years

1988, 1989, and 1990. Specific allocations were based on the best 5 years of landings during the qualifying years of 1984-90 for halibut and 1985-90 for sablefish.

The catch is monitored through a combination of real-time and post-transaction auditing. Deliveries may be made only to registered buyers, and notice must be given to the NMFS. Real-time auditing is through IFQ landing cards and transaction terminals. Post-transaction auditing compares the records submitted by registered buyers with the fishermen's landing records. Provisions also exist for over- and underharvests: limited amounts of annual quota shares can be either deducted or credited to the next year's allocation. In part because of this extensive monitoring system, administration of IFQ programs is somewhat costly. Nevertheless, it is believed that the program's economic benefits will far outweigh the increase in management costs. In addition, as mandated by the new Magnuson-Stevens Act requirements, a cost recovery program to help defray monitoring and enforcement costs was established in March 2000.

Results

Measured against the program's stated goals, IFQs for halibut and sablefish have been highly successful. Most notably, the race for fish was eliminated. The season has increased from less than 5 days to 245 days a year for both species, and landings are now broadly distributed throughout the season. As a result, safety has improved. The program also reduced the frequency with which the TAC was exceeded, in both fisheries. In addition, the IPHC estimates that discarding of halibut bycatch fell by about 80 percent between 1994 and 1995, as did halibut mortality from lost or abandoned gear (although significant uncertainty surrounds both these estimates). There does not, however, appear to be any difference in sablefish bycatch before and after IFQ implementation. There is anecdotal evidence of highgrading (discarding all but the most profitable fish), but comparisons of halibut size-composition data suggest that any highgrading that does occur is insignificant. Underreporting of either halibut or sablefish catches does not appear to be a problem.

Meanwhile the quota share markets have been active, with more than 3,800 permanent transfers of halibut quota shares to date and more than 1,100 transfers of sablefish quota shares. Trading under the IFQ program has also led to some consolidation: the number of quota holders declined by 24 percent for halibut and 18 percent for sablefish between January 1995 and August 1997. In both fisheries the bulk of this consolidation has taken place among those with smaller IFQ holdings. Although it seems likely that the overall efficiency of the fisheries has increased, it remains uncertain how costs and revenues have been affected.

Despite these successes, some concerns remain. Most complaints center on the allocation of IFQ permits, while the rest tend to reflect problems common to any fishing restriction. The primary complaint concerning the initial allocation relates to the delay between the qualifying years and the implementation of the program. Some fishermen who have become active since the qualifying years received no initial free allocations and had to purchase all their quota rights. Conversely, some quota shares were awarded to individuals who had been active during the qualifying years but inactive in the years immediately preceding implementation. Crewmembers and processors also allege that the initial allocation rewarded vessel owners and redistributed market power in favor of quota shareholders. In addition, there is ongoing concern about community effects, adequacy of enforcement, the potential for localized depletion, and the preemption of productive sport-fishing grounds (which are not regulated) by commercial fishermen. Many of these issues could plague any fishing regulation scheme.

Informal Permit Trading in the Tar-Pamlico River Basin

In 1983 local fishermen and citizens in the basin of the Tar and Pamlico Rivers of eastern North Carolina noticed sores on fish, algal blooms (aquatic algae consuming the water's available oxygen), and fish kills in their local rivers and estuaries. Because studies link many of these problems to increased concentrations of phosphorus and nitrogen in water systems, the North Carolina Environmental Management Commission (EMC) designated the region a Nutrient Sensitive Water in 1989.

Laying the groundwork for future regulation was somewhat complicated by the fact that these nutrients came from different types of sources: 83 percent of nitrogen and 66 percent of phosphorus loads originated from non-point sources, such as agricultural runoff and natural phenomena. The remainder came from point sources such as water sewage treatment facilities and local industry. Given the political and technological constraints on detecting, monitoring, and enforcing non-point source nutrient reduction, the proposed EMC regulation targeted point source discharges, setting strict limits on new dischargers and the expansion of existing ones. The ultimate goal of this command-and-control regulation was to reduce phosphorus and nitrogen loading into the region's waters by 200,000 kilograms a year by 1995.

Some of the publicly owned treatment works (POTWs) affected by the regulation estimated that together they would have to spend between \$50 million and \$100 million to achieve compliance with the State's plan.

Concerned about these high capital costs, the POTWs, in conjunction with a private firm, asked the North Carolina State government if a better solution could be found. Working with the Environmental Defense Fund (a private nonprofit group, now called Environmental Defense) and the Pamlico-Tar River Foundation, a coalition of dischargers called the Tar-Pamlico Basin Association proposed an alternative solution involving collective nutrient trading.

Under the arrangement, which was approved in 1989, two types of trades are allowed: collective trading among point sources and collective trading between point sources and non-point sources. In the first case, members of the association operate within a “bubble,” offsetting one another’s discharges to achieve a specified overall limit. In the second case, the members collectively have the option to achieve all or part of the total nutrient reduction goals by funding agricultural best management practices (BMPs) through the State’s Agricultural Cost Share Program, which pays farmers to reduce nutrients and runoff. These offset funds are used to pay willing farmers 75 percent of the cost of adopting nutrient-reducing BMPs on farms within the basin. In this manner the Tar-Pamlico program establishes responsibility at the group rather than the individual level, as no transactions occur between individual point source and non-point source polluters.

So long as the association succeeded in reducing phosphorus and nitrogen emissions by the originally targeted 200,000 kilograms a year, no specific emission reduction requirements would be imposed. Given this flexibility, the association estimated that it could meet this reduction for about \$11.5 million, far less than the estimated cost of the proposed command-and-control regulation. The agreement between the association and the State also required the association to fund a computer model simulating nutrients’ flow and effects; to hire a consultant to evaluate existing wastewater treatment plants, to determine the changes needed to ensure that they are operating at maximum efficiency; to monitor each member’s weekly phosphorus and nitrogen discharge; and to provide upfront funding for the Agricultural Cost Share Program.

In all, 15 dischargers, contributing about 90 percent of all point source flows to the basin, eventually joined the association. Some of those that decided not to join cited the risk involved: there was no guarantee that the association would achieve the required nutrient reduction by 1995. If it failed, the investment and membership costs would be forfeited, and the State’s original command-and-control plan would be implemented. Other point source dischargers that had already planned or begun upgrades in plant facilities could meet the State’s stricter limits without the need to trade.

A tricky feature of this program is the arrangement for trading between point and non-point sources. Whereas the amount of nutrient load entering the water from a point source is easily measurable, that from a non-point source is not. This is in part because the amount of nutrient loading resulting from a given amount of fertilizer can vary considerably, depending on the weather and other conditions outside anyone's control. Because of this added uncertainty, expected non-point source emissions are imperfect substitutes for point source emissions: more than one unit of non-point source reductions is necessary to equal, in quality-adjusted terms, a unit of point source reductions. It was recognized that, because of this, trades between these two types should not occur at a one-to-one ratio. But it was also recognized that the choice of the trading ratio between point and non-point sources would be key to the program's success: too high a ratio would discourage trading, but too low a ratio might fail to achieve abatement goals. In the end, the trading ratio was set at two to one for effluents from non-point sources involving livestock (such as pastureland and poultry operations), and three to one for cropland. That is, to acquire a one-unit credit, the association must pay the State's Agricultural Cost Share Program for the reduction of two (or three) units of a non-point source's nutrient emissions.

To date, compliance has been achieved entirely through trade among point sources. It is uncertain whether this indicates that the trading ratio was set too high, or that abatement costs at point sources are in fact the lowest-cost alternative. But an important outcome is that, thus far, internal "trades" have taken place rather informally. Instead of paying one another to undertake pollution control measures, association members reportedly have each agreed to incorporate nutrient removal systems whenever they expand their facilities. The association maintains that this approach is less costly: economies of scope make it less expensive to expand a facility and upgrade the control technology simultaneously, rather than on separate occasions as trading might require.

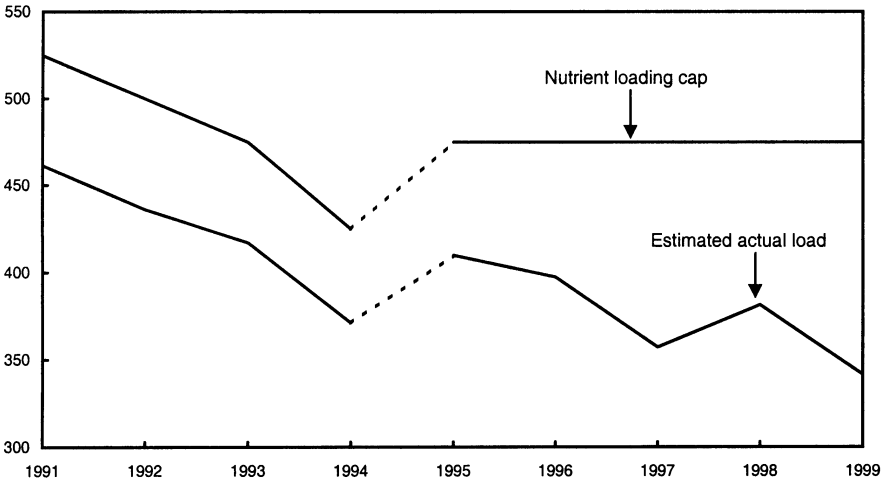
The two largest emitters in the group, both POTWs, were among the first to implement nutrient removal systems. Smaller members have since followed suit. The association expects to achieve the reduction requirements through internal trading for the next 4 or 5 years, after which members may begin to take advantage of trading with non-point sources, or shift to a more formal trading system within the organization, or both.

The results of this market-based program have been impressive (Chart 6-6). Because of growth in nearby communities, dischargers have had to become even more efficient with respect to their nutrient emissions. Even though the association's combined discharge flow increased approximately 20 to 35 percent from 1991 to 1997, total nitrogen concentrations fell by 10 to 20 percent, and total phosphorus concentrations by 20 to 40 percent, in the same period.

Chart 6-6 Nutrient Loading by the Tar-Pamlico Basin Association

The association's nutrient loads under the market-based nutrient reduction program in this North Carolina region have remained well below loading caps.

Thousands of kilograms



Note: Nutrients measured are nitrogen and phosphorus. The periods 1991-94 and 1995-99 represent two distinct phases of the program.

Source: North Carolina Department of Environment and Natural Resources (Division of Water Quality).

When Markets Don't Work

The preceding case studies highlighted three examples where flexible, market-based approaches have been used to achieve environmental goals at substantial savings over less flexible alternatives. In each case the institutions and their historical development differed substantially. An important lesson is that these different settings required different approaches in order to succeed.

In other words, flexible approaches do not succeed simply by virtue of their flexibility. Other elements are necessary as well. First, tradable permit markets typically require a large number of participants to work well. As the Tar-Pamlico case study suggests, one way around this dilemma of a small number of participants may be to create a more informal trading association. Second, it is important that trading not be inhibited by overly cumbersome restrictions. For example, in 1981 the Wisconsin Tradable Discharge Permit system was organized on the Fox River, allowing rights to biochemical oxygen demand discharges (which decrease the oxygen available for fish and other aquatic species) to be traded among point sources. By 1996, however, only one trade had taken place. It is likely that trading was infrequent because administrative impediments discouraged the transfer of permits. Dischargers are not allowed to trade unless they can demonstrate need, and

therefore they cannot trade solely for the purpose of reducing treatment costs. Moreover, the traded rights are guaranteed for a maximum of 5 years, with no assurance that rights will be renewed.

In addition to liquidity among participants at a given moment, liquidity across time is necessary to smooth out temporary fluctuations in aggregate permit demand. For example, the SO₂ trading program allows firms to bank unused permits for future use. By 1996, after just 2 years of operation, the total volume of banked permits actually exceeded annual emission levels. This bank provides an effective cushion against demand fluctuations, as the banked permits can be increased or drawn down as needed. In contrast, the Los Angeles area NO_x program initially lacked a permit bank or other source of aggregate flexibility. As a consequence, the permit price skyrocketed from its historical level of around \$2 a pound to nearly \$50 a pound in the summer of 2000, because of increased demand from fossil-fuel electricity producers. Similarly, an innovative internal greenhouse gas emission trading program at a major energy company has seen fluctuations in demand cause the price to jump to \$99 per ton of carbon dioxide in less than 1 year from almost zero the year before, in the absence of a substantial bank. These aggregate liquidity problems could be solved either by developing a bank or, as suggested above, by empowering the regulatory agency to provide a safety valve, selling additional permits when the price reaches a specified threshold.

Finally, flexible programs work best when monitoring costs are low and when financial incentives—fees or permit requirements—are easily associated with actual emissions or resource use. Automobile emissions, for example, are poor candidates for a trading program: it is impractical to require the drivers of the Nation's more than 100 million registered automobiles to both monitor their individual emissions and acquire tradable permits accordingly. Still, we see flexible approaches—in the form of tradable performance standards described earlier—applied to these sources.

Lessons for Future Policy: Climate Change

One of the most controversial and complex environmental policy challenges facing the United States—and the world—is the long-term issue of climate change. This potential problem spans both generations and countries, implicating simultaneously the environment, on the one hand, and the world's fundamental economic reliance on fossil fuels—a key source of climate change risk—on the other. What do the lessons learned in this chapter suggest about a reasonable approach to climate change?

Base Policy Action on Sound Science

In each of the case studies presented in this chapter, government policy responded to an environmental problem in a manner designed to protect not only the environment but also economic well-being. Sound science guided those responses and must do so in our response to climate change, as articulated by the President in his speech in the Rose Garden on June 11, 2001. Yet the risks arising from climate change are less clear than the risks identified in the case studies, as is the appropriate response. We are uncertain about the effect of natural fluctuations on global warming. We do not know how much the climate could or will change in the future. We do not know how fast climate change will occur, or even how some of our actions could affect it. Finally, it is difficult to say with any certainty what constitutes a dangerous level of warming that must be avoided.

Therefore an important element of a reasonable climate change approach must be more research into both the science of climate change and mitigation technologies, in order to learn more about the risks and the appropriate response. The President has committed the United States to do just that, with research initiatives in both the science of understanding climate change and the means of mitigating its effects. This includes the President's Climate Change Research Initiative and his National Climate Change Technology Initiative, which will add to the more than \$18 billion spent on climate research since 1990.

Choose a Flexible, Gradual Approach

The President has also directed an effort to consider approaches to reducing greenhouse gas emissions. All of the case studies in this chapter demonstrate that flexible approaches consistently provide environmental benefits at a lower economic cost than traditional methods. Flexibility is even more important when balancing climate change and fossil energy use. An effective program must include all greenhouse gases, all emission sources and sinks, and, given the global nature of the problem, all countries. It should provide for flexibility to shift emission reductions over time in response to both short- and long-term opportunities. Flexibility is needed in the face of changing economic conditions, scientific uncertainty, and the development of affordable, advanced energy and sequestration technologies. Finally, an effective program needs to consider non-greenhouse gas emissions that contribute to climate change, such as tropospheric ozone and black soot. Because all of these dimensions offer promising opportunities to address climate change, each must be used in a way that maximizes the mitigation benefit for every dollar spent.

Ideally, this could be accomplished by creating the same incentives for equivalent emission reductions in all these different dimensions: across gases, across sources, across countries, and over time. These incentives would necessarily adjust in response to changing economic conditions and additional knowledge concerning benefits and costs. Yet concepts such as a worldwide tax on greenhouse gas emissions or a worldwide tradable permit system, sometimes advertised as solutions, are at best useful theoretical benchmarks against which to measure alternative, practical approaches. At worst, they can be a distraction from meaningful, realistic steps forward.

Why are such proposals impractical? Because they fail to recognize the enormous institutional and logistical obstacles to implementing any sweeping international program. Institutionally, it is important to learn to walk before trying to run. The United States implemented its successful SO₂ trading program only after gaining experience in the 1970s and 1980s with netting and banking programs, experimenting with control technologies for more than 20 years, and recognizing the limitations of alternative command-and-control approaches. Most other countries have significantly less experience with flexible approaches. A flexible international program would be unprecedented.

As the case studies have also shown, flexible programs have been remarkably successful—but sometimes they run into glitches. For that reason, it would be dangerous to make any serious U.S. policy or commitment dependent on newly designed and untried international institutions—a point highlighted by the President's Cabinet-level climate change working group in its initial findings. Moreover, the current uncertainty surrounding climate change implies that a realistic policy should involve a gradual, measured response, not a risky, precipitous one.

What would constitute a practical policy? In addition to the science and technology initiatives noted above, we could begin investigating reasonable ways to set emission goals and to facilitate efforts by businesses and individuals to think about their own emissions and opportunities for reductions. Internationally, we should continue to expand our cooperation with both developed and developing countries. This will build experience along the various dimensions required for a flexible response and will set the institutional foundation for any further policies that might be necessary in the future.

Set Reasonable, Gradual Goals

A reasonable national goal for greenhouse gas emissions could serve as a benchmark for our progress in terms of mitigation, and thus as an investment in one aspect of a climate change policy that encompasses science, technology, cooperation, and mitigation. One of the problems with climate

policy over the past decade has been a focus on unreasonable, infeasible targets. For example, reducing U.S. emissions to 7 percent less than their 1990 level (the Kyoto target) over the next 10 years could cost up to 4 percent of GDP in 2010—a staggering sum when there is no scientific basis for believing this target is preferable to one less costly. Worse yet, by imposing such high economic costs and diverting limited resources, the Kyoto targets could have reduced our capacity to find innovative ways out of the environmental consequences of global warming. But what defines a reasonable emission goal in the absence of better science?

The uncertainty surrounding the science of climate change suggests that some modesty is in order. We need to recognize that it makes sense to discuss slowing emission growth before trying to stop and eventually reverse it. There is an unfortunate tendency to treat greenhouse gases—especially carbon dioxide (CO_2)—in a manner analogous to SO_2 and NO_x , for which strict quantitative limits have been imposed. SO_2 and NO_x can be controlled by adding equipment to existing facilities. CO_2 , however, can only be reduced by either reducing energy use or replacing fossil fuel facilities, equipment, and transportation fleets with ones that use fuels with lower or zero emissions (that is, unless and until capture and sequestration of CO_2 become feasible). This is vastly more expensive than the end-of-pipe treatment appropriate for SO_2 and NO_x , and it raises concerns about fuel diversity, national security, and the ability to sustain our economic strength and quality of life.

A modest, near-term goal to mitigate greenhouse gas emissions could be described in many ways. A greenhouse gas emission target could be indexed to economic output or other measures of economic activity. Or one could express the goal in terms of greenhouse gas emission intensity, that is, the amount of emissions per unit of economic activity. Both these ideas describe targets that are flexible in the face of economic growth, encouraging reductions without threatening the economy.

A reasonable, gradual goal specified in this way offers advantages over the reductions set out in the Kyoto Protocol. The Kyoto Protocol focused on rather dramatic short-term reductions with unclear environmental benefits. Those reductions risked damaging economic consequences and, in turn, jeopardized the ability to invest in long-run scientific and technological solutions. A reasonable goal offers insurance consistent with existing climate science without putting the economy at risk. A gradual approach balances the need for mitigation with the need for economic growth to power future innovation. A gradual approach also allows us to adjust as we learn more from the science and are able to take advantage of technologies as they develop. Finally, a gradual goal provides time to develop stronger institutions for a long-term, global solution.

Provide Information and Encourage Reductions

In addition to setting a reasonable goal, we need to facilitate efforts by firms and individuals to track their own behavior and to recognize cost-effective mitigation opportunities. The government has a useful role here, both in providing information and in acknowledging progress. No matter how sensible the near-term national goal, firms and individuals need to understand their role—and opportunities—in order to succeed.

One portion of an information program could be the development of procedures and pilot programs to measure both project-based reductions and carbon sequestration. Project-based measurement is important domestically to the extent that offsets are eventually used in certain sectors or for certain gases. It is important internationally if the United States wants to encourage domestic firms to seek out meaningful reductions in developing countries where fully market-based programs are unlikely to be implemented.

Sequestration of greenhouse gases in agricultural and forestry sinks offers considerable opportunity, both domestically and internationally, to achieve inexpensive near- and medium-term reductions—if an environmentally sound accounting method can be devised. We can continue work aimed at reducing the concerns and uncertainty associated with sink usage. In all cases, research, rules, and pilot programs should be developed in consultation with other countries pursuing alternative climate change programs, to ensure both consistency and fair competition.

In addition to educating businesses and individuals about their own greenhouse gas emissions and the opportunities to reduce them, we can encourage them to reduce emissions in innovative ways. This might involve incentives, voluntary challenges, or public recognition, again focusing on flexible, gradual efforts.

Give Technology—and Institutions—Time

These first steps concerning reasonable goals, information, and accounting, along with continued international cooperation, can serve as building blocks toward long-term institutions. To get the institutions right and to protect the economy, however, this movement must be gradual. Initial steps should signal our intent and thereby encourage the development of new technologies—technologies designed to eventually stabilize atmospheric concentrations of greenhouse gases at a level that does not dangerously interfere with the climate system. Such stabilization, in contrast to an arbitrary short-term emission limit, remains the long-term goal recommended both by the United Nations Framework Convention on Climate Change and by the President.

These efforts and goals will require time in order to accomplish them effectively. Science, markets, technology, and global participation must be wound together in an effective policy response. To do so requires building sound institutions for a better environment.

Supporting Global Economic Integration

The world economy has become increasingly integrated. Goods, services, capital, and people flow across borders with greater frequency and in ever-greater volumes. For some, cross-national interaction has become even more a part of day-to-day activity than interactions within their own country.

Americans benefit tremendously from their interactions with other countries, just as they do from their interactions with each other in different States. Such interactions allow Louisianans to drink California wine, Chicagoans to eat bananas and pineapples from Hawaii, and savers in Ohio to provide financing to business startups in Florida. In the same way, international trade allows Americans to enjoy French wine and Colombian coffee and to take advantage of investment opportunities in the United Kingdom.

Despite these benefits, many geographic, institutional, and historical factors impede the free flow of goods, capital, and people across national borders. Realizing the full benefits of international interactions requires building into our economic system mechanisms that facilitate the removal of such impediments. National compacts such as the interstate commerce clause of the Constitution help to link the activities of different States. In the same way, international institutions have developed to promote linkages around the world. Such institutions seek to provide a stable framework for international transactions, while respecting the sovereignty of each country that chooses to participate, as well as serving a valuable coordinating role. International financial institutions such as the International Monetary Fund (IMF) help to promote international monetary and financial cooperation. All of these institutions also evolve in response to changes in the global economy, just as the transactions themselves are likely to change in response to institutional initiatives.

This chapter begins by describing the increasing integration of the world economy and of the United States with the world economy. It then sets out some of the benefits of this globalization and addresses some of the concerns it has engendered. Finally, it discusses the role of institutions within the international economy, covering both recent activities and some likely areas for change.

The United States in the International Economy

Trends and Patterns in U.S. and World Trade

Several factors have contributed to the increased integration of the U.S. economy with the rest of the world. For one, the costs of communicating between a producer in one country and a buyer in another have fallen dramatically, thus reducing the total costs of dealing with a foreign trade or financial partner. One measure of these falling costs is the cost of international telephone service: the average amount billed to end users for a minute of international telephone service fell from \$2.23 in 1975 to \$0.45 in 2000 (in dollars unadjusted for inflation).

In 2000, of the 10 largest international telecommunications carriers in the world as measured by minutes of outgoing traffic, three were U.S. companies, and they held first, second, and sixth place. International telephone traffic worldwide continued to grow rapidly, by more than 20 percent in that year. The flow of international telephone traffic to and from the United States continues to exceed that for any other country in the world. Worldwide satellite industry revenue also grew by 17 percent in 2000. These numbers suggest the continuing significance of international and global communications to U.S. and foreign business firms, who sell and purchase products and services in all parts of the world, and to U.S. and foreign consumers.

The costs of transporting goods between countries have also fallen, and this, too, stimulates international trade. Average nominal freight and insurance costs for U.S. imports fell by about 50 percent between 1975 and 2000, and air cargo rates on long-distance routes declined substantially. Over the same period, the share of U.S. imports that arrives by air increased from 9.2 percent to 25.4 percent. With this widespread use of speedier delivery times, trade in perishable goods as well as in inputs used in just-in-time production processes has grown. The United States now imports eggs from New Zealand and electronic components from Malaysia. Exports from the United States, such as the telecommunications equipment we send to Japan, are also available more quickly to consumers and producers in other countries.

In tandem with these falling communications and transport costs, international efforts to reduce policy barriers to trade have helped to further link the economies of different countries. Average tariffs on industrial goods in developed countries have fallen from 40 percent 50 years ago to around 4 percent today. Nontariff barriers to trade, such as quotas and some regulatory barriers, have also been dramatically reduced.

All these changes in transactions costs have profoundly affected international flows of goods, services, and capital. On a pure volume basis, global merchandise trade has increased substantially in the last two and a half decades, growing by 277 percent between 1975 and 2000 (Chart 7-1). During this same period, U.S. exports grew by around 393 percent, from \$230 billion to \$1.1 trillion (in 1996 dollars). The importance of international transactions in relation to overall U.S. economic activity has also risen. In 1975 total trade (measured as exports plus imports) was equal to less than 16 percent of GDP, but by 2000 that figure was over 26 percent (Chart 7-2). About 8 percent of the labor force is now engaged in producing goods and services that will be sold in foreign markets.

The United States trades with many countries around the world. Canada is our top-ranking trading partner, accounting for 20.3 percent of trade in 2000 (again measured as exports and imports combined). Mexico (12.4 percent) and Japan (10.6 percent) rank second and third, respectively. The countries of the European Union together account for 19.3 percent of U.S. trade. This concentration of U.S. trade in transactions with other high-income countries follows a historical pattern. But trade with a broader

Chart 7-1 **World Merchandise Trade Volume**
Worldwide trade in goods has nearly quadrupled since the mid-1970s.

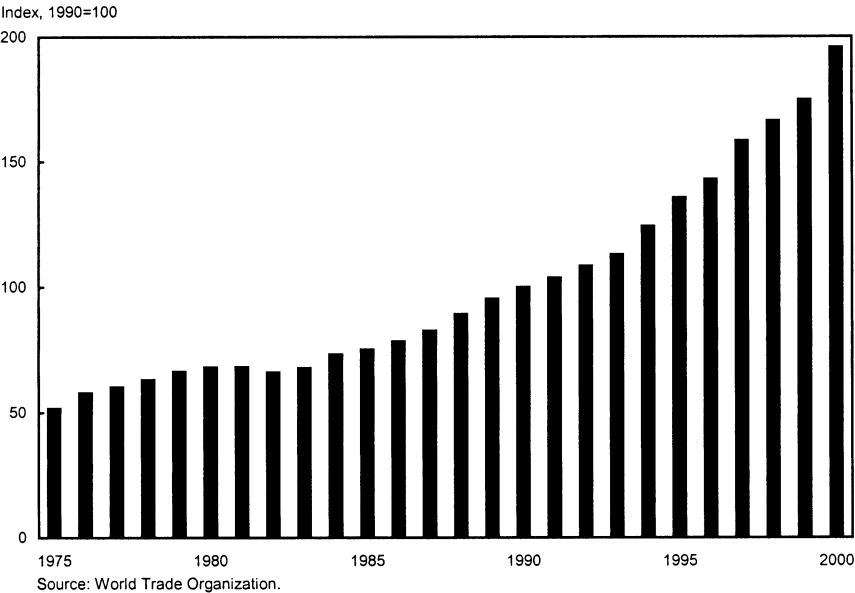
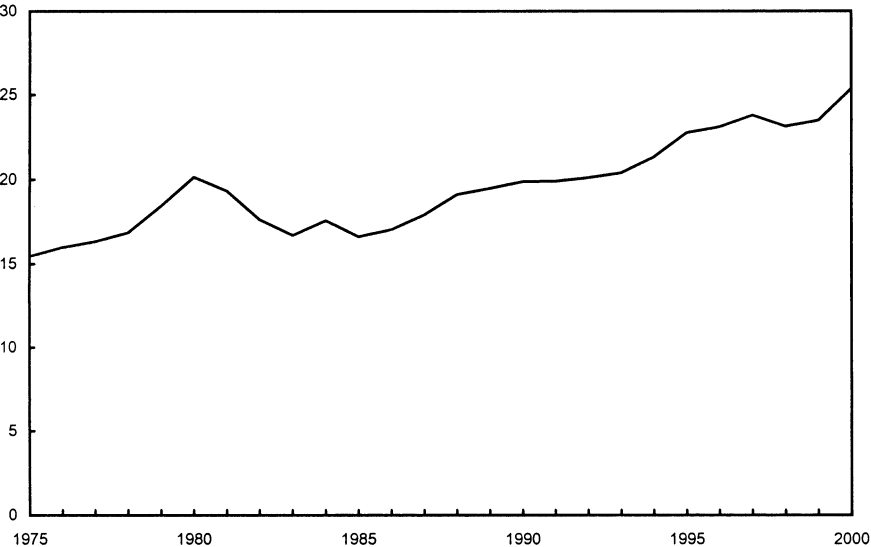


Chart 7-2 **U.S. Trade Relative to National Output**

Trade is substantially more important to the economy than it was 25 years ago.

Percent of GDP



Note: Trade is measured as exports plus imports.

Source: Department of Commerce (Bureau of Economic Analysis).

range of countries already constitutes an important share of our international transactions, as Mexico's high ranking demonstrates. And this trade is growing: trade with low- and middle-income economies grew from \$78.5 billion in 1975 to \$750.2 billion in 2000.

The reduction in impediments to international transactions has also been accompanied by changes in the types of goods being traded. Manufactures have become an increasingly important element of world trade in goods: their share of world merchandise exports rose from 69.8 percent in 1975 to 74.8 percent in 2000. About 80 percent of both U.S. merchandise exports and imports in 2000 were manufactured goods; as recently as 1980 only 55 percent of imports and 70 percent of exports consisted of manufactures. Within manufacturing, certain industries are particularly trade-oriented. Ranked on the basis of exports as a share of shipments, nonelectrical machinery and computer and electronic equipment were the leaders. In each of these industries, exports accounted for 30 percent or more of U.S. firms' total shipments (Table 7-1).

This increasing importance of manufactures reflects in part another important change in the nature of U.S. trade: more and more trade now involves the exchange of intermediate inputs across borders. For example, a

TABLE 7-1.—*U.S. Manufacturing Trade as Share of Shipments
and Consumption, 2000*
[Percent]

Product category description	Exports as percent of shipments	Imports as percent of consumption
Total manufacturing.....	19.8	26.3
Food.....	7.1	5.3
Beverages and tobacco products.....	6.0	9.0
Textiles and fabrics	26.0	25.4
Textile mill products	5.2	14.7
Apparel and accessories	15.5	57.5
Leather and allied products	33.5	80.1
Wood products	6.6	17.8
Paper.....	11.2	13.1
Printing, publishing, and similar products.....	5.8	4.9
Petroleum and coal products.....	4.7	12.2
Chemicals	21.7	19.9
Plastics and rubber products	11.5	11.3
Nonmetallic mineral products.....	10.0	16.7
Primary metals.....	15.4	27.1
Fabricated metal products, not elsewhere specified.....	10.5	12.6
Machinery, except electrical.....	36.0	33.4
Computer and electronic products	44.6	50.8
Electrical equipment, appliances, and components.....	24.8	32.4
Transportation equipment	22.9	33.0
Furniture and fixtures.....	4.6	20.1
Miscellaneous.....	26.3	45.2

Note.—Product category descriptions based on the North American Industry Classification System (NAICS).
Consumption is defined as shipments minus exports plus imports.

Sources: Department of Commerce (Bureau of the Census) and U.S. International Trade Commission.

firm may purchase one input to its production from one country, and another from another country, and assemble the final good at home or even in a third country. One way to measure such interactions is to look at the amount of imported inputs used in goods that are in turn reexported. One study found that, in 1990, such vertical specialization accounted for about 20 percent of all exports in a sample of 14 major trading economies, including the Group of Seven (G-7) large industrial economies (Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States). Increases in such vertical trade have been found to account for more than 30 percent of the growth in the ratio of world exports to world GDP. Such trade may help to enhance the efficiency of producers, since they now have access to a wider range of input sources than are available domestically. (Box 7-1 discusses the importance of vertical trade in overall U.S. trade.)

Box 7-1. Vertical Trade and Production Sharing

A large portion of U.S. trade, both imports and exports, is trade in partially finished products, also called intermediate inputs. Examples include the steel used in automobile manufacture, and the cloth and other textiles from which finished apparel is made. This type of trade goes by many names, such as vertical trade, vertical specialization, and production sharing, although these terms have somewhat different meanings. Vertical trade, the broadest category, includes any production process that is not confined to one country. Vertical specialization is slightly narrower. It is defined as the use of imported inputs to produce goods that are subsequently exported. Production sharing is narrower still: imported inputs are used to produce goods that are then exported to the country from which the inputs came.

Some of these production processes are organized by a single (vertically integrated) firm, but in a growing number of cases separate companies in different countries manage different stages of production. In the past, many companies felt that the only way to guarantee the timely arrival, exact adherence to specifications, or quality of an intermediate good was to own all the steps on the supply ladder (hence the name “vertical integration”). For similar reasons, it may sometimes have been difficult to locate plants overseas. However, the past decade or so has seen large improvements in the technology available to coordinate and monitor manufacturing in different parts of the world. This includes everything from cheaper and better international telephone service to fax machines to Internet-linked computer-aided design packages. These advances have allowed companies and countries to specialize in those steps of the production process that they are best at performing, leading to an increase in vertical trade.

The extent of vertical trade can be gauged in a number of different ways. One way is simply to measure the amounts of intermediate goods that are imported or exported. However, it is sometimes difficult to decide whether a good should be classified as intermediate, because this depends on its intended use, which may not be known. Auto tires are a good example of this. They can be used as an intermediate good and put on cars to be sold as part of a final product, or they can be sold in retail stores as a product themselves. The ideal would be to look at how much of a traded good's value is added in each of the countries involved in its production. One measure of this is the imported input share, that is, the share of the value of production that is attributable to imported inputs. Another such measure would be the amount of

continued on next page...

Box 7-1.—*continued*

production sharing, which is defined as U.S. materials shipped abroad for processing and then sent back to the United States. Note that production sharing is a special case of vertical trade, since vertical trade also covers inputs shipped to Mexico or Canada, finished there, and exported to any country, not just the United States.

The U.S. Government has kept statistics on production sharing since about 1963. These numbers are collected because products assembled abroad from U.S. manufactured components qualify for different tariff treatment: only the portion of the product's value not accounted for by U.S. inputs is subject to duties. The tariff provision that governs such production sharing is number 9802. Two main categories of goods covered under this provision are goods assembled of U.S.-made components, and metals. Of course, the data collected do not capture the entire extent of production sharing, as certain products are exempt from duties under various agreements such as the North American Free Trade Agreement (NAFTA). In fact, in the first table below, which traces U.S. imports from selected economies in the Asia-Pacific Economic Cooperation (APEC) forum, the total recorded in 2000 fell from the previous year, possibly because of increased exemption of goods. In the table, "customs value" is the total value of the goods imported into the United States, and "U.S. content" is the percentage of value that comes from U.S. inputs. Therefore, under provision 9802, duties would only have to be paid on the difference between the customs value and the value of U.S. components: the value added abroad. For example, in 2000, the United States imported \$1.38 billion worth of goods from Korea for which a 9802 exemption was claimed. The U.S. content of those goods totaled 54.6 percent, or \$750 million, and therefore the value added abroad was 45.4 percent, or about \$630 million.

In addition to collecting statistics, the U.S. Government occasionally publishes surveys of developments in production sharing. According to a recent survey, major industries involved in vertical trade include the automotive industry and various electronics industries. For example, the United States imports motor vehicles from Canada (\$45.7 billion, or 35 percent of the total), Japan (\$34.5 billion, or 27 percent), and Mexico (\$21 billion, or 16 percent). Exports of motor vehicles from Japan, which is not covered by NAFTA, contained U.S. components comprising 2.4 percent of the value of these imports. Exports of motor vehicles from Canada and Mexico, however, have historically contained U.S. components equal to one-quarter and

continued on next page...

Box 7-1—continued

U.S. Imports from Selected APEC Economies under Tariff Provision 9802

Economy	1998		1999		2000	
	Customs value (millions of U.S. dollars)	U.S. content (percent)	Customs value (millions of U.S. dollars)	U.S. content (percent)	Customs value (millions of U.S. dollars)	U.S. content (percent)
Australia	25.0	16.6	18.7	22.3	18.8	26.4
Canada	427.8	45.4	358.9	49.0	483.1	48.0
China	1,477.2	15.7	1,612.0	16.9	1,242.4	20.3
Hong Kong, China	558.9	41.2	451.2	38.1	253.2	38.8
Indonesia	298.0	18.3	296.8	18.0	190.1	26.2
Japan	12,363.1	4.1	15,058.2	3.8	17,851.3	3.0
Korea	1,601.2	49.1	2,002.3	52.0	1,378.0	54.6
Malaysia	1,830.7	50.0	2,109.1	47.3	1,639.3	54.0
Mexico	27,162.2	53.3	25,875.0	53.8	19,429.9	52.9
New Zealand	2.0	36.9	.9	51.6	3.2	18.3
Peru	.9	34.2	4.0	6.4	1.6	1.8
Philippines	2,253.7	50.1	2,331.3	48.8	2,098.7	44.5
Russia	2.7	26.6	1.8	18.0	5.8	39.9
Singapore	556.4	27.1	200.6	40.7	235.5	40.3
Chinese Taipei	1,511.2	35.9	1,716.7	34.1	881.8	44.8
Thailand	663.6	55.3	592.0	56.8	396.3	56.4
Vietnam	78.5	11.2	114.2	13.8	47.9	20.9
Total	50,813.3	38.6	52,744.2	36.7	46,157.1	32.0

Source: U.S. International Trade Commission.

two-fifths of their value, respectively. (The last years for which such data are available are 1988 for Canada and 1993 for Mexico. After that, those countries were covered by free-trade agreements and no longer recorded values for provision 9802.) And indeed, the United States exported \$17 billion worth of automotive parts to Canada in 2000, and \$7.3 billion to Mexico.

Another sector in which production sharing is prevalent is electronic products. U.S. content in machinery and electronic products imported from Mexico under the production sharing provision was \$4.9 billion in 2000. As mentioned previously, however, not all production sharing is captured by provision 9802, as there may be other programs under which the goods in question get more favorable treatment. Luckily, we can get a rough idea of the discrepancy through the following calculations. Mexico also collects statistics on U.S. products imported as

continued on next page...

Box 7-1—continued

inputs to planned exports under its maquiladora and PITEX programs. The measured value of imports of machinery and electronics intermediate goods from the United States was \$37.2 billion in 2000 (a much larger number than \$4.9 billion). Overall, Mexico exports 92 percent of its maquiladora products to the United States, and so one can estimate that the U.S. content of machinery and electronic products under all production sharing arrangements was at least \$34.2 billion in 2000. This implies that the 9802 statistics capture only a small portion of all production sharing between the United States and Mexico. As an illustration, the second table in this box lists the top 20 production sharing commodities from Mexico. The U.S. content, measured as a percentage of the final value, is typically quite high.

*Top 20 Product Categories in Production Sharing in
U.S.-Mexico Trade, by U.S. Content, 2000*

Product category description	Customs value (millions of dollars)	U.S. content (percent)
Cotton sweaters, pullovers, and similar articles	232.0	80.4
Parts and accessories of motor vehicles	355.3	78.0
Manmade fiber sweaters, pullovers, and similar articles	273.2	76.8
Cotton T-shirts, singlets, tank tops, and similar garments	588.8	75.5
Safety seat belts for use in motor vehicles	491.6	74.5
Insulated electric conductors	236.7	66.3
Motor vehicles for transport of goods, 5-20 metric tons	297.5	60.6
Switches for electrical connections	246.6	60.2
Connectors such as coaxial, cylindrical multicontact	417.4	59.0
AC motors	264.8	56.1
Other electrical telephonic apparatus	266.6	55.2
Insulated ignition wiring sets and other wiring sets for vehicles	699.7	48.0
Motor vehicles for transport of goods, not over 5 metric tons	247.9	46.9
Boards, panels, consoles, etc., for electrical control consoles	252.4	43.9
Non-high-definition color television reception apparatus	759.7	38.3
Cotton women's or girls' trousers, breeches, and shorts	934.1	35.5
Cotton men's or boys' trousers and shorts	825.4	35.3
Parts of motor vehicle seats	283.5	16.0
Display units for ADP machines	273.8	2.5
Digital processing units	249.8	2.4

Note.—Product category descriptions based on the Harmonized Tariff Schedule (HTS).

Sources: Department of Commerce (Bureau of the Census) and U.S. International Trade Commission.

Interestingly, the often back-and-forth nature of vertical trade means that a significant portion of the value of U.S. imports simply represents the value of previous U.S. exports. Many domestically produced goods are shipped abroad for further processing or assembly and then returned to the United States, in another illustration of how international trade becomes part of the overall production process. This is a particularly striking feature of U.S. trade with Mexico. In 1998, for example, the United States imported \$93 billion worth of goods from Mexico, \$27.2 billion of which entered the country under a special “production sharing” provision of U.S. law that gives duty-free treatment to the reimportation of goods produced with U.S. components. Of this \$27.2 billion, \$14.5 billion (53 percent) represented the U.S.-made content of these imports. That \$14.5 billion also represents at least 15 percent of all U.S. imports from Mexico.

Lower international transactions costs have facilitated trade in services as well as in goods. Between 1986 and 2000, total U.S. trade in services grew by over 200 percent. One reason is that falling communications costs have allowed many products that were not traded in the past, such as financial services, to become more readily available on the international market. U.S. trade in financial services quadrupled between 1986 and 2000, from \$5.1 billion to \$21.5 billion. Other categories of U.S. services trade, such as travel, education, and royalties and license fees, have also greatly increased.

Trends and Composition of Capital Flows

Like trade and services flows, global capital flows have increased enormously over the past 30 years. These flows represent funds channeled from savers in one country to borrowers in another. From the end of World War II through the early 1970s, capital controls in most countries heavily regulated or even prohibited the international flow of capital. Only when these controls were liberalized, especially in the late 1970s and early 1980s, did cross-border financial transactions begin to surge.

Global capital movements can be analyzed in terms of both gross and net flows. For example, suppose that early in December German residents purchase \$200 worth of U.S. securities from U.S. residents, and that later that month they sell \$50 worth to U.S. residents. Considering only these transactions, capital flows into the United States from Germany amount to \$150 (\$200 in purchases minus \$50 in sales). Suppose further that, over the same month, U.S. residents first purchase \$100 worth of German securities from German residents and then sell them \$30 worth. Considering the latter two transactions, capital flows into Germany from the United States amount to \$70 (\$100 in purchases minus \$30 in sales). From the perspective of the United States, net capital inflows amount to \$80 (\$150 of inflows minus \$70 of outflows). One measure of gross capital flows, used in the tables in this

chapter, would sum the capital flows into and out of the United States to arrive at a total of \$220. A broader measure, usually not available from official data sources, would sum all cross-border purchases and sales to arrive at a total of \$380. Regardless of which concept is used, gross capital flows will be larger than net flows by definition.

Although it may appear that the gross basis overstates the importance of capital flows, gross flows do measure the amount of international funds flowing in and out of a country's financial system. Especially for developing economies, it is important to know if these flows are so large that they might overwhelm the capacity of the domestic financial system to process them.

Unfortunately, data on gross capital flows come from different sources and are often fragmentary. Since cross-border financial transactions are usually not subject to tariffs or quotas, national authorities have lacked a strong incentive to document their size. Nonetheless, the IMF estimates that, in the 30 years since 1970, gross capital flows as a percentage of GDP have risen almost tenfold for the advanced economies and more than fivefold for developing economies. Table 7-2 presents more recent measures of capital flows. From 1990 through 2000, estimated capital flows on a gross basis in advanced economies more than quadrupled.

TABLE 7-2.—*Estimated Gross Private Sector Capital Flows*¹
[Billions of U.S. dollars]

Item	1990	1995	1996	1997	1998	1999	2000
Advanced economies: gross flows.....	1,536.8	2,285.6	2,975.4	4,163.8	4,053.4	5,885.2	6,432.1
Direct investment	404.7	515.5	567.6	674.7	1,104.3	1,774.8	2,070.7
Portfolio investment	377.5	818.3	1,182.8	1,348.8	1,871.4	2,731.1	2,628.7
Other ²	754.6	951.8	1,225.0	2,140.3	1,077.7	1,379.3	1,732.6
<u>Memoranda</u>							
Gross financing to other markets ³	38.1	151.2	209.8	274.9	148.9	163.7	216.5
Equities	1.2	10.0	17.8	26.2	9.4	23.2	41.8
Bonds	8.7	59.2	103.0	126.2	79.5	82.4	80.5
Loans	28.2	82.0	89.0	122.5	60.0	58.1	94.2
United States: gross flows	189.1	697.5	878.9	1,226.9	876.8	1,218.8	1,566.3
Direct investment	85.7	156.5	178.4	210.4	320.7	456.4	440.1
Portfolio investment	30.4	218.9	280.1	316.9	354.2	475.2	610.6
Other ²	73.0	322.1	420.5	699.6	201.9	287.2	515.6

¹ Gross flows are the sums of the absolute values for inflows and outflows of each country.

² Generally, bank loans.

³ Data include new formal international offerings or syndicates, but exclude bank lending that is not syndicated and investments that do not occur through public offerings. Thus, substantial amounts of financing are excluded.

Note.—Advanced economies comprise Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States.

Detail may not add to totals because of rounding.

Capital flows can also be categorized by the nature of the investment being undertaken. Capital used by a firm in one country to establish a plant in another is labeled foreign direct investment, as are large purchases of equities that imply a lasting interest in an enterprise. Purchases of long-term bonds, money market instruments, and small amounts of equities are labeled portfolio investment. Residual transactions such as loans fall into the category labeled “other” in Table 7-2. Gross capital flows have shifted toward direct and portfolio investment in the past decade.

The explosion in gross capital flows obscures the fact that, on a net basis, capital flows have grown much less rapidly (Table 7-3). This difference in the two measures means that larger amounts of funds are crossing borders, but that the balance of inflows and outflows is remaining roughly constant. These net flows also reflect the balance of domestic saving and investment in a country. If a country saves more than it invests, the excess savings must go abroad. Similarly, if a country invests more than what is available from domestic saving, the extra funds must come from abroad.

These net capital flows are also just the mirror image of the country's current account balance, which, roughly speaking, consists of the balance in its combined goods and services trade and the net flow of income generated from cross-border investments. A country that sends savings abroad, on net, is enabling the rest of the world to spend more on that country's goods and services than that country is spending on goods and services produced by the rest of the world; such a country has a current account surplus. A country that is attracting savings from abroad, on net, is able to spend more on goods and services produced by the rest of the world than the rest of the world is spending on goods and services that the country itself produces; that country has a current account deficit.

Although net capital flows on a global basis have increased relatively little in recent years, this is not the case for the United States, as Table 7-3 also shows. The United States recorded large current account deficits over the past decade, reflecting an increased desire on the part of foreigners to invest in the United States. The United States also ran large current account deficits in the 1980s. An important source of financing for these deficits was foreign official purchases of U.S. government debt securities. In the 1990s, however, the bulk of foreign investment entering the United States consisted of purchases of private assets. In particular, direct investments in the United States have shown a very rapid rate of increase over the past several years. In short, rapid rates of productivity growth and increases in economic activity over the past decade have made private assets in the United States more attractive for foreign investors.

Because the world's developing economies have relatively little capital compared with the developed economies, there is a presumption that capital

TABLE 7-3.—*Estimated Net Private Sector Capital Flows*

[Billions of U.S. dollars; inflow (+), outflow(-)]

Item	1990	1995	1996	1997	1998	1999	2000
World.....	162.7	100.7	280.4	213.0	127.0	343.9	423.8
Direct investment	-45.5	-12.1	1.5	10.3	-13.7	46.8	133.3
Portfolio investment	46.1	81.8	22.0	50.0	-140.4	254.2	287.5
Other	162.2	31.0	256.9	152.7	281.1	42.8	3.1
Memoranda							
Emerging markets	39.2	205.7	233.3	116.8	69.6	59.6	8.9
Direct investment	19.3	96.5	119.6	145.2	155.4	153.4	146.2
Portfolio investment	.5	41.2	86.9	48.6	-4.2	31.0	-4.3
Other	19.4	68.0	26.8	-77.0	-81.6	-124.8	-133.0
United States.....	26.3	14.2	39.7	253.6	172.0	321.6	406.9
Direct investment	11.3	-41.0	-5.4	.8	35.7	145.6	135.2
Portfolio investment	-27.2	-26.1	-19.6	78.9	82.0	212.7	360.7
Other	42.1	81.3	64.7	173.9	54.3	-36.7	-89.0

Note.—World is defined here as advanced economies (Australia, Austria, Belgium, Canada, Cyprus, Denmark, Finland, France, Germany, Hong Kong (China), Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States) plus emerging markets (the developing countries, countries in transition, and Israel, Singapore, South Korea, and Taiwan (China)—the IMF definition in “World Economic Outlook,” December 2001).

Detail may not add to totals because of rounding.

Sources: Department of Commerce (Bureau of Economic Analysis) and International Monetary Fund.

should flow from the latter to the former. Hence capital flows to and from these developing economies receive much attention. Table 7-3 shows that these flows have varied enormously over the past decade. In the early 1990s some developing economies made enormous strides in structural economic reform and removed restrictions on capital flows, leading to a renewed interest on the part of international investors. Net flows skyrocketed, reaching \$233 billion in 1996. However, the financial crises that began in East Asia in 1997 and then occurred in Russia and Brazil in 1998 and 1999 dampened investors’ appetites. Net flows fell to close to zero in 2000 but are believed to have increased moderately in 2001. A swing in net banking flows accounts for most of the decline since 1996. This was due to both a decrease in international bank lending to developing economies and an increase in deposit outflows from developing economies to international banks. (The lower international bank lending reflects in part a move from cross-border lending to more lending by subsidiaries within the countries.) However, direct investment flows have remained fairly stable over the past 3 years, a sign that investors are still willing to undertake long-term investments in the developing economies.

Cumulating net capital flows for a given country and accounting for changes in the prices of assets held across borders yields the net international investment position for that country with the rest of the world. For example,

suppose that a country begins international transactions with the rest of the world and for 10 years enjoys net capital inflows of \$1 billion a year (possibly including reinvested earnings). At the end of these 10 years that country's net international investment position would show that the rest of the world has accumulated a total of \$10 billion in claims on that country, assuming that the prices of these claims did not change over the 10-year period. These claims could be in the form of portfolio investments (if, for example, investors in the rest of the world bought bonds issued by the country's corporations) or direct investments (if the rest of the world bought controlling interests in the country's corporations).

Table 7-4 indicates that, worldwide, these cross-border claims are quite large in the aggregate, at over \$21 trillion, equal to almost 70 percent of world GDP. The claims are largely divided among bank loans, equities, and bonds. Central bank reserves make up a fourth, relatively small category. These holdings are now much smaller than those of private investors, having grown at about half the rate of gross capital flows over the last 30 years.

TABLE 7-4.—*Estimated World Cross-Border Claims and U.S. International Investment Position, Year-End 2000*

Item	Billions of U.S. dollars
World cross-border claims	21,261.0
Bank loans and deposits	8,317.6
Equities	4,516.5
Debt securities.....	6,377.2
Central bank reserves ¹	2,049.6
U.S. claims on rest of world ²	7,189.8
Bank assets	1,276.7
Corporate stocks.....	1,828.8
Bonds	577.7
Central bank reserves ³	128.4
Other	3,378.2
Rest-of-world claims on United States ²	9,377.2
Bank liabilities.....	1,139.8
Corporate stocks.....	1,589.7
U.S. Treasury securities, corporate and other bonds	2,013.9
Central bank reserves ³	922.4
Other	3,711.4

¹ Gold valued at SDR 35 per ounce..

² Direct investment at market value.

³ Gold valued at market price.

Note.—Detail may not add to totals because of rounding.

Sources: Department of Commerce (Bureau of Economic Analysis), Bank for International Settlements, and International Monetary Fund.

Table 7-4 also indicates that the United States is a party (either a lender or a borrower) in roughly 80 percent of global cross-border claims. As noted above, foreign investors have found the U.S. economy very attractive and have built up their holdings of U.S. assets. At the same time, U.S. citizens have substantial holdings of foreign assets. Foreign-owned assets in the United States total \$9.4 trillion, and U.S. claims on the rest of the world total \$7.2 trillion, so that the United States is today in the position of a net debtor.

In most cases, transferring capital across borders requires a foreign exchange transaction, in which the currency of one country is exchanged for that of another. As capital flows have increased, so has turnover (the total value of transactions) in the foreign exchange market. Data for foreign exchange turnover correspond to the broadest measure of capital flows discussed earlier. There is no attempt to net purchases and sales against each other, either across trading days or across transactions that finance one country's purchases versus those that finance its sales. Since 1989 daily nominal foreign exchange turnover has more than doubled; it now averages \$1.2 trillion. But turnover has actually fallen since 1998, for two reasons. One is that the introduction of the euro as the common currency of the European economic and monetary union means that many cross-border transactions within Europe no longer require an exchange of currencies, and the other is that consolidation has occurred in the international banking sector.

Given the annual capital flow data summarized in Table 7-2, the turnover data suggest that gross flows for the year as a whole are the product of extraordinarily large flows on a daily basis within the year. This provides yet another explanation for policymakers' concern that in some cases the sheer size of these flows could overwhelm the resources of a poorly supervised financial system in the event of a sharp reversal. This issue is discussed further later in the chapter.

The Benefits of Globalization

The various trends, described in the previous section, toward increased interaction between people and firms in different countries—increases in trade as well as increases in capital flows—are often collectively referred to as globalization. Each of these forms of globalization, and others such as international migration, benefit the United States in a variety of ways, as this section will show.

The Benefits of Trade

International trade, both exports and imports, benefits the economy in a number of different ways. In a general sense, exports benefit the economy

because American workers have another market—the global market—in which they can sell the goods and services they produce. Over 12 million American jobs are supported by exports. Opening foreign markets for U.S. producers allows them to expand their output and hire more American workers. Before the North American Free Trade Agreement (NAFTA) went into effect in 1994, for example, U.S. shipments of assembled motor vehicles to Mexico were severely hampered by Mexico's high tariffs and other regulations designed to protect the local automotive industry. Under NAFTA, Mexico was required to reduce these barriers: in 1998 Mexico eliminated its tariffs on light trucks produced in the United States, and all remaining Mexican tariffs on medium and heavy trucks and buses were eliminated on January 1, 2002. Subsequently, U.S. exports of motor vehicles to Mexico rose from \$975 million in the 5 years preceding NAFTA to \$6.6 billion in the 5 years after NAFTA. And this happened despite a major recession in Mexico following that country's financial crisis of 1994-95.

The health of many sectors of the American economy depends upon trade. America's farmers, for example, rely on sales to foreign markets. Exports of U.S. agricultural products amounted to \$53 billion in 2000, and roughly 25 percent of cash sales by farmers and ranchers come from sales to foreign consumers. U.S. agricultural exports support 740,000 American jobs.

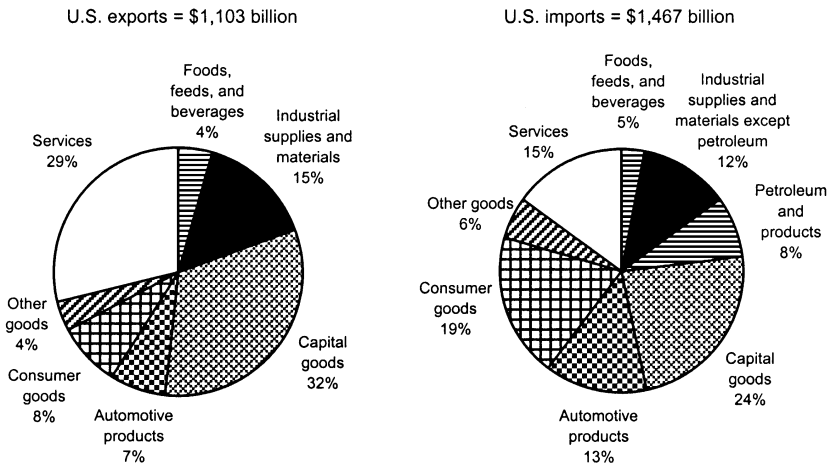
Trade also benefits the economy in a number of more specific ways. First, trade may reduce the prices of some of the goods that we consume. When a country is closed to trade, domestic consumers are forced to buy only those goods produced in their home market. Often, however, a producer in another country is able to produce the same goods more efficiently, that is, at a lower cost. When trade is open, consumers have the choice of buying the imported good at the lower price. In addition, now that domestic producers are competing with imports, they will have greater incentive to produce using the lowest-cost methods possible. Thus international trade tends to reduce the prices of some goods traded. Of course, if the United States is already the lowest-cost producer of a good, domestic consumers will continue to purchase it from domestic suppliers.

A second specific benefit of trade is that it gives a country's consumers access to the many different goods and services produced around the world. For example, without trade, we would not be able to purchase coffee from Costa Rica, or enjoy certain fresh tropical fruits year-round. We would not have access to some products at all, or would be able to consume only the domestic variety. Similarly, when a firm needs a specialized input for a production process, trade often allows it to choose from many options available around the world, rather than only those produced at home. This option allows the firm to produce more efficiently, and be more competitive internationally, than without this choice.

As a third benefit of international integration, trade helps boost productivity in the United States. Increased competition from trade provides incentives for domestic firms to produce using the most efficient, lowest cost methods possible. Firms that are successful in international competition are likely to be more productive than those that sell only at home. In fact, recent evidence shows that exporters tend to be relatively more efficient and to pay higher wages than nonexporters. One study found that, in 1992, a worker at an exporting plant earned wages that were 10 percent higher, and nonwage benefits that were 11 percent higher, than a worker at a nonexporting plant.

Trade also allows the U.S. economy as a whole to specialize in the products that it is comparatively best at producing. This is because trade between nations is the international extension of the division of labor. The United States exports some of the goods and services that it is relatively better at producing, and receives in exchange goods and services that other countries are relatively better at producing. For example, the United States exports manufactured goods that require high levels of technical skill, such as telecommunications equipment and professional scientific instruments. Some of these industries, such as electronics and computer equipment, sell at least a quarter of their merchandise overseas (Table 7-1). This reflects the relative abundance of highly skilled labor in the United States. U.S. imports, on the other hand, tend to be in areas such as consumer goods (Chart 7-3). This specialization of economic activity based on comparative advantage

Chart 7-3 U.S. Trade by Sector in 2000
 The largest category in both exports and imports is capital goods.



Note: Data are on a national income and product accounts basis.
 Source: Department of Commerce (Bureau of Economic Analysis).

allows the United States as a whole to use its resources most effectively, and it allows Americans to purchase goods from the world's best sources of those goods. Thus both exports and imports are beneficial and help make the United States a richer and more efficient economy.

Trade also increases productivity because it gives exporters access to a larger total market. Because some goods, such as automobiles, are produced most cheaply in large quantities, a larger market may allow exporters to reduce their production costs through economies of scale. Finally, trade benefits the economy through the access it provides to foreign technology and ideas. We can import innovative products from abroad and use them to increase our own efficiency, or to create even newer technologies, raising the rate of economic growth.

The Benefits of Capital Flows

Just as trade flows result from individuals and countries seeking to maximize their well-being by exploiting their own comparative advantage, so, too, are capital flows the result of individuals and countries seeking to make themselves better off, in this case by moving accumulated assets to wherever they are likely to be most productive. Increased capital flows benefit both the lender and the borrower. From the lender's perspective, cross-border capital flows provide an opportunity to diversify an investment portfolio. To the extent that returns on international assets do not move in lockstep with returns on domestic assets, diversification through cross-border investments both increases expected returns and lowers risk. These benefits lie behind the large increases in capital flows documented earlier in the chapter. The "home bias" to investment portfolios is falling; whereas in the late 1980s only 6 percent of U.S. residents' equity holdings were in foreign assets, more recent estimates put that share at more than 10 percent. Even that, however, is below the percentage that most models of optimal portfolio selection would predict.

For the borrower, cross-border capital flows allow for an expansion of production possibilities. Lending from abroad allows more capital to be combined with other inputs to increase the production of valuable goods and services. Some of the increase in output will be used to pay back the lender, but a substantial fraction should contribute to a rise in domestic standards of living. This is particularly important for developing economies, where overseas capital effectively substitutes for or augments often-scarce domestic sources of investment. Capital inflows can help keep domestic interest rates low, making sure that government borrowing to finance programs for education and health care does not crowd out private domestic investment.

Capital flows also boost efficiency in the borrowing country. New ideas and techniques accompany capital flows across borders, allowing for a more

efficient allocation of resources within the country. Such knowledge transfers boost productivity in the receiving country, allowing for more rapid technological economic progress there. This is most evident in the case of foreign direct investment, where new plants and new management methods can lead to sharp increases in output. Capital inflows also help expand and diversify the financial system in the recipient country, and this, too, leads to a more efficient allocation of capital and faster growth.

The increases in economic well-being associated with increased capital flows require a supportive domestic environment. Without this support, capital flows can reverse themselves sharply, imposing large adjustment costs on the borrowing economy. The risks of a reversal are heightened if the borrowing economy is pursuing unsound macroeconomic policies, or if supervision of the financial system is inadequate.

Quantifying the positive relationship between increased capital flows and faster growth is difficult, for several reasons. First, poor macroeconomic or regulatory policies may render some countries unable to harness investment capital in ways that promote sustainable growth. Second, causation between capital flows and economic growth is likely to run both ways. An increase in capital available to an economy will boost growth, but as an economy grows, it is more likely to attract foreign capital. This confronts economists with a chicken-and-egg question: which came first, the capital flows or the growth? Recent empirical research has struggled with these problems but, on balance, concludes that the increased capital flows brought about by capital liberalization spur economic growth. All else being equal, a country that opens up to capital flows can expect to enjoy an increase in its growth rate per capita of half a percentage point or more per year. For example, if an economy is growing at an annual rate of 2 percent, opening up to capital flows would allow its economy to double in size 7 years sooner than otherwise.

There is every reason to expect that in the long run international capital flows will continue to increase in importance, as economies around the world become more interlinked. Continued increases in trade volumes, discussed earlier in the chapter, will require capital flows to finance them. Investors will continue to obtain the benefits of diversification from increasing their international exposures. And, as we have seen, the average investor is still a long way from holding an optimally diversified international portfolio. Finally, although world living standards are improving on average, both the relative and the absolute gap in incomes per capita between rich and poor countries continue to increase. This gap indicates that the rate of return on capital in the world's poor economies is likely to be several times that in the rich economies, providing an enormous incentive for continued—and indeed, augmented—flows. Of course, this will only be true to the extent that productivity gains achieved in the developed economies can be transferred across borders. And most important, it requires that the least developed

economies have sound policies and educated work forces in place, to make effective use of the capital coming in.

The Role of Migration

Migration is another important aspect of the internationalization of the economy. Just as trade in goods, services, and capital allows resources to be used most efficiently, so, too, the movement of people from country to country around the world can enable them to make the best use of their skills and abilities. Thus removal of barriers to immigration allows for more efficient worldwide distribution of workers.

The United States has a long history of accepting people from other countries, as witnessed by the numbers collected by the Bureau of the Census on the foreign-born population. In 2000 foreign-born residents made up 10.4 percent of the U.S. population (although in 1900 they represented an even greater 13.6 percent). Immigrants have been a key building block for the U.S. economy. Our openness to immigration has allowed us to reap the benefits of the presence of newcomers from many countries.

Immigrants benefit the economy in several ways. First, people are a resource, similar to the other resources of our economy such as land or minerals. Immigrants who come to the United States to work allow the country to produce more. It has been estimated that if immigrants make up 10 percent of the population, the net overall gain from their presence is somewhere between 0.01 and 0.14 percent of GDP per year. Given that, in 2000, U.S. GDP was \$9.9 trillion, the overall gain is between \$1 billion and \$14 billion.

The increase in the labor force from immigration also affects prices. The goods and services that immigrants produce tend to become cheaper as more immigrants enter, and all consumers benefit from this reduction in prices. This price drop is an average price drop across all goods and services. Some goods and services—in particular, those that use a lot of unskilled labor—will see sharper drops in prices than others. Household services and services to dwellings are examples. On the other hand, the prices of goods and services that use less unskilled labor are likely to fall by less or stay the same, and may even increase.

Legal immigrants who work may also contribute to government finances by paying taxes on the wages they earn. Because they tend to be younger workers, immigration also improves the current balance sheet of Social Security. Of course, legal immigrants may receive welfare benefits, which impose a cost on the government and taxpayers. Recent research provides some estimates on the balance between taxes that immigrants pay and the benefits they receive. These calculations indicate the ultimate effect on taxpayers of a given legal immigrant now and into the future, taking account of the effects of that

specific immigrant on taxes and benefits, as well as the effects of his or her children into the future. Overall, according to this research, the average immigrant makes a net positive fiscal contribution of about \$80,000.

Some Myths About Trade and Globalization

Although globalization, by increasing the movement of goods and services, capital, and people across the Nation's borders, has provided a variety of benefits to the United States, many have expressed concerns about globalization's effects, both in the United States and abroad. This section reviews some of those concerns and explains why globalization is, in fact, unlikely to have the adverse effects often feared.

Trade and the Environment

A variety of concerns have been raised about the impact of globalization on the environment. One is that government action to implement domestic environmental regulations may be interpreted in other countries as protectionism and, consequently, in violation of trade agreements that the United States has entered into. Domestic environmental regulations may then be challenged, and the case adjudicated by international dispute settlement mechanisms. The concern is that the United States might be forced to change or eliminate its own environmental standards.

In fact, environmental regulations do not normally raise issues of consistency with international trade agreements, which are aimed at preventing discrimination against foreign products, not at lowering environmental standards. There is generally no reason for environmental regulations to lead to discrimination against or among foreign products. If a product is judged to inflict environmental harm, its production and use are normally regulated, or prohibited, without regard to its origin; if this is the case, such regulations are unlikely to breach international trade obligations. Even if they did, international trade agreements contain exceptions that allow a country to take environmental measures against imported products that might otherwise violate obligations under the agreement.

For example, Article XX of the 1994 General Agreement on Tariffs and Trade—one of the agreements among members of the World Trade Organization (WTO)—lists a number of general exceptions to members' obligations. One of these confirms that a WTO member may adopt and enforce measures “necessary to protect human, animal or plant life or health” or “relating to the conservation of exhaustible natural resources.” These exceptions are subject to a number of conditions, among them that the measures not arbitrarily or unjustifiably discriminate among countries and

that they not constitute a disguised restriction on international trade. (NAFTA incorporates similar exceptions and conditions.) Thus, nothing in these international agreements prevents the United States from establishing and maintaining legitimate environmental measures, so long as it does so in a way that does not unjustifiably discriminate against its trading partners or create unnecessary barriers to trade. In fact, the General Accounting Office concluded in 2000 that, “The WTO rulings to date against U.S. environmental measures have not weakened U.S. environmental protections.”

Other concerns about globalization may stem from the fear that growth in developing countries resulting from increased trade may lead to environmental degradation. But in fact, there is no clear relationship between development and pollution levels. Indeed, some evidence shows that organic water pollution intensity falls substantially as a country’s income per capita rises from \$500 to \$20,000, with the decline beginning before the country reaches high-income status (about \$10,000 in annual income per capita). Trade may also give countries access to cleaner technologies, allowing them to build their industries in a more environmentally sound fashion.

Trade and Employment

Some argue that globalization leads to the loss of jobs for American workers. It is true that some domestic firms will not be able to compete effectively with imports, and these firms may be forced to reduce their work force or even cease operations. At the same time, however, the opportunity for increased trade will lead other firms to expand their operations and increase hiring, in order to serve the international market as exporters. These firms tend to be the more productive ones in the economy. Exporters also tend to pay higher wages than firms that do not export—in 1992, up to 18 percent higher on a simple average basis, according to one study.

It is also true that the firms forced by import competition to eliminate jobs may be in different sectors from the exporters who are increasing hiring. This can make it difficult for those who lose their jobs to import competition to find new jobs with exporting firms that use the skills they have acquired. But such shifts in employment also reflect one of the benefits of trade for the aggregate economy, namely, that it allows the economy to produce the goods and services that it is comparatively best at producing, and to buy from other countries those goods and services that it is relatively ill equipped to produce. The expansion of trade that may precipitate such a shift of workers may, as a result, lead to an increase in the average income of the American worker, because wages in import-competing industries tend to be below the average, whereas wages in exporting industries tend to be above the average. Workers in export-competing industries such as aircraft and pharmaceuticals earned about 22 to 60 percent more than the average wage in 2000. The reverse is

true for import-competing industries: wages in the apparel industry, for example, were 36 percent lower than the average in manufacturing, those in the leather industry were 29 percent lower, and those in the textile industry 35 percent lower.

The shifting of jobs across sectors may take time, and some workers may face dislocation. However, the displacement of some workers by imports should not be an excuse for discouraging trade, any more than the costs to some workers of technological change should stop the development of innovations. It would have made little sense to discourage the diffusion of personal computers just because it jeopardized the workers of typewriter manufacturers. Imposing trade restrictions in an effort to save those jobs will only destroy, or prevent the creation of, jobs in other sectors. If, for example, government-imposed trade barriers were to hinder access to imported capital goods, the domestic firms that purchase those inputs would be forced to operate at higher costs of production. This would adversely affect their competitive position relative to foreign rivals who have free access to such capital goods. Domestic producers might lose sales, and this might force them to downsize their work forces, or even to shift production to locations abroad where the inputs are freely available.

Of course, finding a new job in another firm or another industry, after losing one's job to import competition, may be difficult. The Federal Government recognizes this possibility and has put programs in place to assist those who lose their jobs because of trade in finding new ones, and to provide them with financial assistance while they make the transition. For example, the Trade Adjustment Assistance (TAA) program provides training, job search aid, and relocation allowances; these benefits are on top of unemployment insurance and other programs. In 1999 close to 130,000 workers were estimated to be in groups certified as eligible for TAA. This Administration is committed to reauthorizing and improving existing TAA programs that are due to expire. The Administration worked during 2001 to strengthen the performance of these programs, so that they are more effective at easing the transition into new employment. In addition, for certain sensitive sectors such as textiles and agriculture, trade liberalization is designed to proceed in gradual stages so that workers have more time to adjust.

Trade and Relative Wages

Over the last three decades, the returns to education, in the form of higher wages, have increased dramatically, although the rise has flattened out in more recent years. In 1979 a male with a college degree could command a 30 percent wage premium over a male with only a high school diploma. This premium had risen to 60 percent by 1995 but has remained relatively constant since then. Because workers with less education often work in

industries that compete most closely with imports, particularly those from developing countries, some have blamed increased trade for these changes in wages. However, although the United States did increase its imports from developing countries over this period, it also experienced a great deal of technological change, which increased demand for workers with higher skill levels. This tends to increase the relative wages of those with higher skill levels. In fact, it appears that this increased demand for more educated workers, and not increased trade with developing countries, has led to the recent change in relative wages.

The Effects of Trade on Developing Nations

Some have suggested that international trade may harm workers in developing countries, because countries like the United States import goods produced under poor working conditions or at very low wages. Those who hold this position argue that the United States should use trade measures, such as withholding access to our markets, as a weapon to force developing countries to improve working standards or to increase wages.

The use of trade policy to force such changes, however, would have perverse effects, actually hurting those it aims to help. For example, if the United States and other countries refused to import from countries where wages are below a certain standard, workers in those countries would be denied the opportunity to work in an export-producing industry. Unfortunately, jobs in other industries may not be readily available in that country, or if they are, may pay even lower wages and impose even worse working conditions.

In addition, to cut off imports from such countries may be to deny them one of their best opportunities for economic growth. A number of recent studies show that participation in an open trading system has a positive effect on a country's income per capita. One study finds that increasing the ratio of trade to GDP by 1 percentage point raises income per capita by 1.5 to 2 percent, and an increase in average incomes is generally associated with higher incomes for the poor. Several studies by the World Bank also point to a linkage between trade liberalization and faster economic growth, as liberalization encourages higher rates of investment and more rapid technological innovation. Thus, limiting trade with developing countries may only serve to keep the poor in their poverty. Perhaps because of the negative effects of linking trade and labor outcomes, many developing countries are strongly opposed to including discussions on labor standards in international trade negotiations.

Many countries, including the United States, do adhere to certain core labor standards, such as the prohibition of exploitative child labor. Trade in and of itself does not cause poor working conditions. Rather, they are more

likely to be the result of domestic policies and economic circumstances. In fact, trade may help to improve working conditions, just as it may facilitate an increase in incomes. Benjamin Franklin summarized it well: “No nation was ever ruined by trade.”

International Policy Issues and the Role of International Institutions

An important factor in the continued worldwide growth in trade and capital flows has been the creation and development of international institutions dedicated to promoting that growth. The United States is a participant in these institutions and has benefited from their important work. The United States has also participated in recent efforts to reform some of these institutions. The present section discusses some of the most important of these organizations and recent proposals for their reform.

International Trade Institutions and the Benefits of Trade

International trade institutions and agreements are designed to ensure that all parties are able to enjoy the benefits of free and open trade. These institutions allow many countries to negotiate together to reduce barriers to trade in ways that are acceptable to all. They also create a stable framework for international transactions. If progress is to continue toward the goal of increased trade, it is crucial that the United States encourage its trading partners to maintain the focus of trade negotiations on this main purpose, rather than stray into areas, often very controversial, that could stall greater progress toward free trade.

The international trade agreements in which the United States has participated can be classified into several broad types. Those of the first type are called multilateral agreements, in which a large number of countries around the world agree to reduce barriers to trade among themselves. As a rule, agreements of this type, such as the General Agreement on Tariffs and Trade (GATT), are structured such that each participating country agrees to reduce trade impediments to all other participants. One of the foundations of the GATT/WTO system is the most-favored-nation (MFN) principle, which mandates that if a WTO member extends any benefit (such as a reduction in tariffs) to a product of another WTO member, it must extend the same benefit to like products of all other members.

A second type of trade agreement is the regional trade agreement, examples of which include NAFTA and the trade agreements of the

European Union. In such agreements, each participant agrees to reduce trade barriers only with respect to the other participating countries in the region. So, for example, in NAFTA, the United States reduced its barriers to Mexican and Canadian exports but made no such changes for exports of European or Asian countries. (Such favorable treatment of regional trade might seem to violate the MFN principle for countries that are WTO members; however, Article XXIV of the 1994 GATT explicitly allows for such regional agreements under certain conditions.)

Although regional agreements generally make good progress toward free trade among the participants, they may introduce some distortions in trade patterns. A country may end up importing goods from a country in the region that has high costs of production but is subject to a low tariff, rather than from one outside the region (or a nonparticipant within the region) that has a low cost of production but faces a high tariff. Such trade patterns (called trade diversion) may hinder the most efficient use of global resources. However, an advantage of regional trade agreements over multilateral agreements is that a smaller group of countries may find it easier to come to a consensus on trade liberalization. Also, if the agreement is among countries that would naturally engage in a great deal of trade with each other in the absence of artificial barriers to trade (for example, countries in close geographical proximity to each other), the amount of trade diversion may be very small.

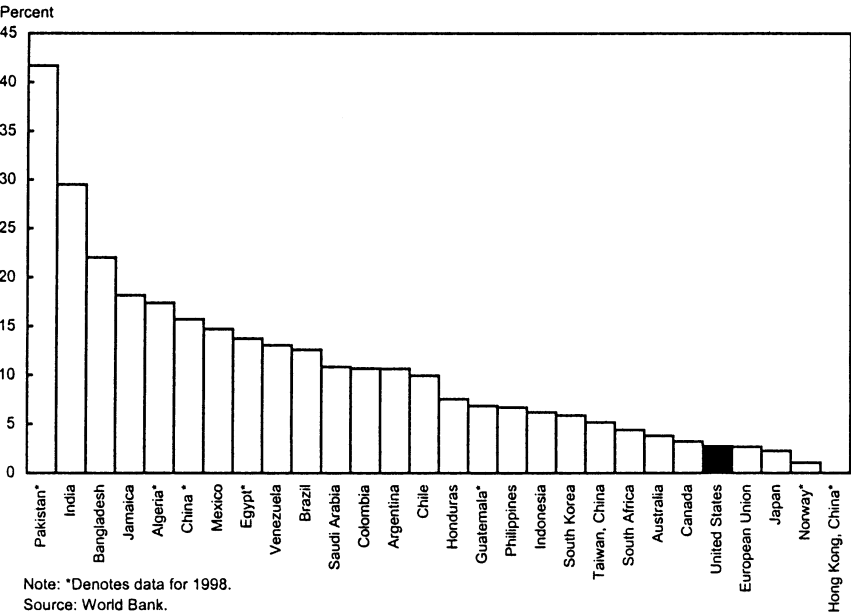
The WTO has reported a massive proliferation of regional trade agreements in recent years, with an average of one per month being notified to the organization. A recent study by the WTO Secretariat identified a total of 172 regional trade agreements currently in force (including some that have not, or not yet, been notified to the WTO), and this number could well grow to about 250 by 2005. On the basis of the 113 regional trade agreements notified to the WTO and deemed to be in force as of July 2000, it is estimated that some 43 percent of world trade occurs within such agreements. This share would rise to 51 percent if all 68 or so of the regional trade agreements currently under discussion and scheduled to be in force by 2005 were already in place.

Economists are divided as to whether regional agreements help or hinder progress toward broader, multilateral agreements. On the one hand, negotiation over regional proposals may divert negotiating resources from multilateral talks, or a proliferation of different regulations under various regional agreements may raise transactions costs for trade. On the other hand, if all countries engage in regional agreements, there will be competition to get the best trade deals, and this competition can lead to bidding down barriers to free trade. It may also be easier for a small country to get larger countries to recognize and understand its needs in a regional than in a multilateral setting.

Finally, a third type of trade agreement is the bilateral trade agreement, such as the recent agreement between the United States and Jordan. Others include the agreement between the United States and Israel and that between Canada and Chile. Such agreements have pros and cons similar to those of regional agreements.

The United States benefits significantly from its participation in international trade institutions, for a number of reasons. For one, because U.S. tariffs on imports are already among the lowest in the world, any agreements to further liberalize trade will likely lower other countries' tariffs more than they lower U.S. tariffs. U.S. tariffs average about 2.5 percent on comparable, trade-weighted terms (Chart 7-4), but U.S. producers face extremely high tariffs in many developing countries. For example, average tariffs on U.S.-produced goods are 13.7 percent in Brazil, roughly 17 percent in Thailand, and up to 35 percent in India. (The numbers for Brazil and Thailand are average applied rates; that is, they are averaged over all imports from the United States. The rate for India is a ceiling rate, which means that no tariff is supposed to be higher than 35 percent. However, because of exceptions put in by the Indian government, the applied rate could be higher.) Many of the United States' trading partners, including the European Union and Japan, maintain high barriers on a range of agricultural goods.

Chart 7-4 Import-Weighted Average Tariffs, 1999
 U.S. tariffs are among the world's lowest; many developing countries impose tariffs at far higher rates.



Thus, multilateral agreements on tariff reduction often disproportionately benefit U.S. exporters.

However, tariffs are not the only artificial barriers to trade. Other barriers include quotas (quantitative limits on import volumes), technical regulations and standards (such as for telecommunications equipment), rules for the valuation of goods subject to tariffs (which affect how the tariffs are calculated), and rules regarding investment (for example, limiting the percentage of foreign ownership of a domestic company). Unfortunately, whatever their stated purpose, such rules are often in fact designed to protect domestic industries from foreign competition. The United States faces discriminatory regulations in many countries. Discriminatory foreign health and safety regulations cost the United States over \$5 billion in agricultural exports in 1996, according to the Department of Agriculture.

To circumvent this problem, most trade agreements establish the principle of nondiscrimination, or national treatment. This means that all countries that are parties to the agreement must treat the exports of other parties as if they were domestically produced. Since many international agreements now include provisions on regulatory barriers and government procurement policy, this requirement allows U.S. exporters to avoid such impediments in other countries. As tariffs fall, these kinds of negotiations become increasingly important to the opening of markets.

The United States has participated in a number of different trade institutions and agreements over the years. For example, the United States was a member of the GATT from its inception in 1948 until 1995, when the WTO was formed. Until the WTO came into being, the GATT was both the agreement (which is still in effect) and the international organization formed on an ad hoc basis to support it. The United States benefited significantly from the outcome of the Uruguay Round, a recent major round of multilateral negotiations under the auspices of the GATT. The reduction in U.S. tariffs that emerged from that agreement had an effect on an average American household of four similar to a tax cut of \$310 a year, or the equivalent of a per-year income gain of more than \$600.

The WTO is an international institution in which the United States negotiates agreements with 143 other members to reduce barriers to trade. In addition, the WTO maintains a forum for dispute settlement that enables its members to resolve trade disputes arising under the WTO agreements. At the fourth WTO Ministerial Conference in Doha, Qatar, in 2001, the members of the WTO agreed to launch a work program that includes further negotiations on trade liberalization. Negotiations will commence in a number of areas, including agriculture, services, industrial market access, a limited set of environmental issues, antidumping and subsidies, and WTO dispute settlement rules; it will also include important work on trade-related capacity building for developing countries. Members also committed

themselves to maintain their current practice of not imposing customs duties on electronic transmissions at least until the Fifth Session of the Ministerial Conference, which is likely to occur in 2003. Negotiations on certain issues, such as investment and competition policy, are delayed until that conference.

Some of the issues slated for negotiation have proved particularly difficult to deal with in the past, suggesting that gains from the new WTO agenda could be large. The new work program will address market access barriers to trade in agricultural products as well as government subsidies in this sector. Some countries, such as those of the European Union, rely heavily on export subsidies. The potential gains to the United States from these discussions are indeed sizable, in part because the multilateral negotiations promise to reduce barriers to U.S. trade around the entire world. One study finds that if a new trade round reduced world barriers on agricultural and industrial products and on trade in services by one-third, the gains to the United States could amount to \$177 billion, or about \$2,500 for the average American family of four.

The United States is also a founding member of the Participants to the Arrangement on Guidelines for Officially Supported Export Credits, an independent body within the Organization for Economic Cooperation and Development (OECD). The arrangement was established in 1978 to limit the terms and conditions under which governments can finance their exports, with the goal of opening export markets by eliminating official financing subsidies. Financing subsidies close markets by eliminating competition on the basis of price, quality, and service and directing business to those countries willing to spend budget resources to provide below-market export financing. The arrangement is currently operated by 24 OECD member governments and governs official export credits totaling \$45 billion in 2000, as well as aid financing of about \$9 billion to \$10 billion a year. The WTO leaves much of the discipline for such indirect subsidization to the OECD Arrangement, and therefore the U.S. antisubsidy efforts in the OECD are complementary to its broader WTO work to eliminate subsidies. The Treasury Department estimates that OECD disciplines over aid financing subsidies alone have opened export markets worth \$5 billion to \$6 billion annually, leading to increased U.S. exports of about \$1 billion each year. The overall U.S. budget savings from all OECD disciplines on financing subsidies amount to around \$300 million a year.

NAFTA has been another important example of U.S. participation in international trade institutions. From 1994, when NAFTA went into effect, until 2000, total trade among the United States, Mexico, and Canada increased from \$297 billion to \$676 billion, or 128 percent. The share of worldwide U.S. goods exports that has gone to NAFTA partners more than doubled over the same period, from 14 percent to 37 percent. Trade restrictions imposed on U.S. exports by our NAFTA partners have fallen

significantly. For example, in 1993 Mexico's average tariffs on U.S. goods were more than twice as high as U.S. tariffs on Mexican goods. Under NAFTA, Mexico's average tariff on U.S. exports has fallen below 2 percent, and two-thirds of U.S. exports now enter Mexico duty-free. Nearly all of the \$406 billion in goods traded between the United States and Canada enters duty-free.

The United States has benefited from this agreement, which when fully implemented will, according to some estimates, yield an increase in U.S. GDP of between 0.1 percent and 0.5 percent, or between \$10 billion and \$50 billion relative to the size of the economy in 2000. For an average household of four, this translates into a per-year income gain of \$140 to \$720. The NAFTA liberalization is also roughly equivalent to a tax cut of \$210 for the same family. U.S. producers of various commodities also benefit from NAFTA. Exports of beef and processed tomatoes to Canada, as well as of cattle, dairy products, apples, and pears to Mexico, are 15 percent higher than they would have been had the Canada-U.S. Free Trade Agreement, and later NAFTA, not reduced barriers to U.S. goods in those markets, according to the Department of Agriculture.

The United States is currently involved in efforts to liberalize trade with a larger number of our hemispheric neighbors. Discussions toward a Free Trade Area of the Americas (FTAA) began at the Summit of the Americas in Miami in December 1994. Thirty-four countries agreed to construct a free-trade area in which barriers to trade and investment would be progressively eliminated, and to complete negotiations toward the agreement by 2005. The FTAA thus aims to establish free trade across the Western Hemisphere, from Hudson Bay to Tierra del Fuego. The nine FTAA negotiating groups cover a range of areas, including market access, agriculture, services, investment, intellectual property, government procurement, competition policy, dispute settlement, and antidumping, countervailing duties, and subsidies.

The potential market that an FTAA would create is enormous: the combined GDPs of Central and South America amount to \$1.57 trillion. (This figure leaves out Mexico, as it is already covered under NAFTA.) And the obstacles currently faced by American exporters in Latin America are formidable, particularly since other countries in the region already have negotiated reductions in barriers with each other. For example, when Chile and Canada recently concluded their bilateral free-trade agreement, Chile's across-the-board 8 percent tariff was eliminated on Canada's exports, but it remains in effect on U.S. exports. Under the MERCOSUR trade arrangement—a customs area agreement signed in 1991 among Argentina, Brazil, Paraguay, and Uruguay—imports and exports among these four countries and Chile are largely duty-free; U.S. exporters to those countries face average tariffs of almost 15 percent. The FTAA promises to eliminate the discrimination against U.S. products in these markets.

The importance of breaking down barriers throughout the hemisphere is epitomized by the experience of Caterpillar Inc. Caterpillar's motor graders made in the United States for export to Chile face nearly \$15,000 in tariffs. Yet when Caterpillar manufactures motor graders in Brazil for export to Chile, the tariff is just \$3,700. And if Caterpillar's competitors were to produce a similar product in Canada, it could be exported to Chile duty-free under the Canada-Chile free-trade agreement. One result of these high trade barriers against the United States may be to create incentives for U.S. firms to locate factories abroad.

If an FTAA were to eliminate barriers to trade in agricultural and industrial goods and in services among the countries in the hemisphere, the United States could reap a gain of \$53 billion, according to one study. An FTAA would also promote greater economic integration and regional cooperation, bringing greater economic opportunity and political stability to the region. Negotiations toward this agreement continue.

As this review has shown, past U.S. participation in international trade institutions and agreements has benefited the United States significantly. Our continued ability to exercise effective leadership in trade negotiations, however, depends on restoration of the President's Trade Promotion Authority (TPA). TPA allows the President to submit a negotiated trade agreement to Congress subject to an up-or-down vote, without amendments. Congress retains the final decision on whether or not the United States signs any trade agreement, but TPA provides the President with more negotiating leverage and gives the United States enhanced credibility in negotiations with its trading partners.

TPA has a long history. In the 1934 Reciprocal Trade Agreements Act, Congress for the first time agreed to give its prior approval to any trade agreement reached by the executive, although it did require that the negotiating authority be renewed every 3 years. Although the Trade Act of 1974 required that Congress approve trade agreements after their negotiation, it also provided a "fast-track" procedure in which Congress would vote in a timely fashion and without amending the agreement. This fast-track procedure has been used to pass legislation implementing the United States' most recent important international trade agreements, including NAFTA in 1993 and the Uruguay Round of the GATT in 1994. These procedures, however, lapsed in 1994 and have not been renewed.

Role and Reform of International Financial Institutions

International financial institutions (IFIs) exist to help countries cope with short-term balance of payments problems and address longer term development challenges. Capital flows have played an increasingly important role in

both these areas, calling for policy responses from countries and from the IFIs themselves.

As already noted, capital flows represent a transfer of resources across time, as savers lend to borrowers today in exchange for repayment plus interest or dividends tomorrow. Increased uncertainty about those repayments can render unattractive an investment that was once attractive. In particular, changes in economic policies or political developments can cause investors to sharply reevaluate the prospects for future payments. Thus their very forward-looking nature can make capital flows subject to abrupt reversals.

Sharp reversals of international capital flows have occurred many times in history. The United States in the 1800s was a developing economy that benefited from European capital inflows. Financial disruptions in the 1850s, 1870s, and 1890s were associated with sharp reversals in these flows. The same situation played out in Latin America in the 1930s. As capital markets collapsed with the onset of the worldwide depression, governments in the region were hit particularly hard. By 1935 almost 70 percent of Latin American national government bonds were in default.

More recently, the emerging market debt crisis in the 1980s was another example of a sharp reversal in capital flows. Rising real interest rates associated with the effort to contain global inflationary pressures made investment projects in developing economies look less attractive. This reversal of capital flows led to a “lost decade” for the Latin American economies until expectations improved when new policies involving structural reform were put in place. Most recently, the crises of the 1990s—in Mexico in 1994-95, East Asia in 1997-98, and Russia and Brazil in 1998-99—again demonstrated how investments based on forward-looking calculations of risk and expected return can quickly reverse, especially when weaknesses in the recipient country’s policy framework are exposed.

These abrupt reversals in capital flows are extremely costly. The withdrawal of foreign investment drives up interest rates in the borrowing country, retards domestic investment, and often leads to a sharp contraction in economic activity and a shrinking of future production possibilities. The balance sheets of domestic firms that depended on these flows are considerably weakened, and there is often a wrenching reallocation of domestic resources away from the nontradable goods sector to the tradable sector, to accomplish the current account adjustment necessitated by the drop in capital flows.

Finally, many of the world’s poorest economies, plagued by years of economic mismanagement, have had little access to private capital flows of any kind. Investors are unwilling to extend loans without some prospect of repayment. But the possibility of repayment is bleak given an unstable system of governance that cannot guarantee property rights, or establish the necessary legal, financial, and physical infrastructure that would foster the

productivity of their citizens. Often, the result is a cruel paradox: the countries most in need of capital—and that might offer the highest potential rates of return on that capital, were the proper policies in place—are precisely the ones with the least access to international capital flows.

The Evolution of Today's International Financial Institutions

Two of today's principal IFIs were created as part of the post-World War II international financial arrangements that came to be known as the Bretton Woods system. Chief among the IFIs is the International Monetary Fund, established in 1945. One of the original goals of the IMF was to provide short-term loans to countries to help with balance of payments adjustment. Under the system of pegged (but adjustable) exchange rates in place from the late 1940s until 1971, it was expected that countries on occasion would require help to manage a set of macroeconomic policies that was inconsistent with the country's fixed exchange rate. The usual manifestation of this inconsistency was a current account deficit that could not be offset by private capital flows at the prevailing exchange rate. One alternative in such a situation would be to devalue the domestic currency in an effort to close the current account deficit. However, following a series of such devaluations in the 1930s in which countries essentially competed for trade advantage, the IMF was created to provide short-term funding to countries in such distress. This funding was meant to provide countries with the breathing room necessary to implement a more rational set of macroeconomic policies that would allow them to avoid the devaluation option.

With the abandonment of the Bretton Woods system of fixed exchange rates in the early 1970s, the IMF essentially lost its original role. Over the past 25 years, the IMF's mandate has broadened to include promoting international monetary cooperation and orderly exchange arrangements with the aim of fostering economic growth. To carry out this mandate, the IMF undertakes surveillance of the macroeconomic policies of its 183 member economies and provides them financial and technical assistance. In this sense, the IMF no longer functions merely as a crisis lender to economies facing balance of payments adjustments. The IMF has also become involved in supporting development programs, aiding the world's most impoverished countries through loans, help in devising a macroeconomic policy framework, and technical assistance.

The IFIs also include what are known as the multilateral development banks (MDBs), of which the World Bank Group is the largest. The World Bank was established in 1945 and had its initial focus on the reconstruction efforts following World War II. As Europe and Japan rebuilt, that focus shifted toward development, targeting the poorest countries, which were unable to obtain access to private international capital flows. The late 1950s saw the creation of the Inter-American Development Bank, the first of four

regional MDBs. Together the MDBs worked toward the goal of financing the development of the world's poorest economies. However, during the crises of the 1980s and 1990s the scope of the MDBs' mission was broadened, and, often encouraged by governments in the developed economies, they participated in the financial crisis lending packages organized primarily by the IMF. Thus the missions of the IMF and the MDBs have sometimes overlapped, with the IMF providing some nonemergency financing for developing economies and the MDBs contributing to crisis financing packages.

Performance of the International Financial Institutions in the 1990s

The turmoil in the international financial system in the second half of the 1990s indicated a shift in the nature of financial crises. The increase in the size of capital flows during the 1990s, documented earlier in this chapter, led to larger, more sudden crises when those flows reversed. These crises also appeared harder to contain, and the result often was large-scale IMF lending. The nature of these new crises focused attention on the role of the IFIs and raised key questions for policymakers. First and foremost, were the resources of the IFIs adequate to deal with these crises? Second, was the provision of assistance itself encouraging further crises? And finally, were countries becoming overly dependent on crisis financing provided by the IFIs?

From the mid-1980s through the mid-1990s, the IMF's resources available for crisis lending (also called its available liquidity) were adequate. However, over the 6-year period beginning in 1995, the average size of IMF stand-by arrangements (traditional lending programs), relative to the recipient country's IMF quota, more than tripled compared with the 6 years beginning in 1989. This is not surprising given the increase in gross capital flows over the 1990s. The new type of crisis was met with a larger official sector response. As a result, it became clear that, in the second half of the 1990s, IMF resources were shrinking relative to private financial flows. This was especially apparent during the Asian financial crisis, when IMF available liquidity fell to \$56 billion in December 1997 from \$83 billion the year before. By December 1998, available liquidity had dwindled to \$54 billion.

Over the mid- to late 1990s, as crises developed and the size of IMF assistance programs increased, policymakers began to revisit the concern that the provision of official assistance was contributing to the development of new crises. The logic in support of such a proposition emphasizes the expectations of private investors. If investors come to expect that countries will automatically receive assistance in the event of a financial crisis, they are likely to exercise less prudence when making loans. Countries that are pursuing unsound policies may still get loans from private investors, since the investors believe that any future problems are likely to be resolved by the provision of

funds by the IFIs. This is an example of moral hazard: an increase in risky behavior (in this case on the part of the borrowing countries and their lenders) when insurance or a guarantee is provided (in this case by the IMF). Thus the concern is that IFI support can encourage risky activity on the part of private lenders and borrowing countries, which often ends badly in further rounds of crises.

The resolution of the crises of the late 1990s was also complicated by a shift in the composition of capital flows away from syndicated bank loans toward bond issuance. Such a shift protected the banking and payments systems of the industrial countries from the worst consequences of international financial crises. However, it also complicated the task of crisis resolution, because restructuring a country's debt now required dealing with a large number of bondholders spread around the world, rather than a small group of bank creditors. When a country's creditors are few in number, it may prove possible to coordinate an orderly restructuring that does little to interrupt economic activity (although this proved surprisingly difficult with bank loans to Latin American governments in the 1980s). But when the lenders are a large, diffuse group of bondholders, an orderly restructuring may be next to impossible. In fact, the switch from bank finance in the 1980s to bond finance in the 1990s in part may have reflected efforts by creditors to safeguard their positions by making such a restructuring more difficult for borrowers. In addition, the shift from bank to bond finance is part of a larger trend, seen not just internationally but in domestic capital markets as well, away from financial intermediaries to direct finance.

Efforts to Reform the International Financial System

As early as 1995, following the Mexican crisis, it became clear to international policymakers that the set of policies and institutions collectively known as the international financial system might be in need of overhaul, especially the IFIs themselves. Various official bodies commissioned reports that examined ways in which the system could be improved. These reports tended to focus on four key areas: transparency and accountability, strengthening national financial systems, management of crises, and debt relief. The following sections deal with each in turn.

Transparency and Accountability. Market-based transactions work best when parties are fully informed. Absence of important information on the part of the lender or the borrower in a transaction can lead to less than efficient outcomes (a finding recognized in the work of the most recent Nobel laureates in economics). Thus reform proposals have called for additional transparency and accountability both on the part of countries receiving capital flows and on the part of the IFIs themselves. In response, the IMF has established the Special Data Dissemination Standard to facilitate the flow of information from countries. In addition, the IMF has encouraged the publi-

cation of documents related to its surveillance (the annual Article IV consultations on each member's economic policies) and of the supporting documents submitted by the country and the IMF when a financial assistance program is put in place and reviewed. Over the last year, 45 percent of the full Article IV consultation reports were made publicly available.

Strengthening National Financial Systems. Several of the crises of the 1990s involved lax practices in the financial and corporate sectors of borrowing economies (see the 1999 *Economic Report of the President*). As a result, calls for the reform of the international financial system have included measures to strengthen national financial systems through the implementation of best practices in financial regulation. To meet these needs, the G-7 authorized the creation of the Financial Stability Forum (FSF) as a way to coordinate the activities of finance ministries, central banks, financial regulators from key economies, the IFIs, and international standard-setting bodies such as the Basel Committee on Banking Supervision and the International Organization of Securities Commissions. The FSF identified key standards and codes for countries' financial systems and has worked toward fostering their implementation. Beginning in May 1999, the IMF and the World Bank introduced the Financial Sector Assessment Program (FSAP) and a key byproduct, the Reports on the Observance of Standards and Codes (ROSCs), in order to assess countries' implementation of these standards. As of September 30, 2001, 57 countries had undergone review of their standards and codes, and reports for 36 had been published. As of the same date, 22 FSAPs had been completed, with 4 assessments published. The IMF has identified 11 main standards and codes that will be addressed in the ROSCs, including the Basel Committee's Core Principles for Effective Banking Supervision.

Management of Crises. As noted earlier, resolving the capital account crises of the second half of the 1990s required much larger IMF programs and caused a dwindling in available liquidity. One aspect of reform efforts was therefore the decision to increase IMF resources in 1998. The IMF resolution required that new commitments by member countries to the IMF be \$89 billion. In February 1999 the United States increased its share by \$15 billion. For crises affecting the global financial system as a whole rather than that of an individual country, additional funds are available to the IMF through borrowing agreements with a number of IMF members and other institutions. Provisions for a New Arrangement to Borrow (NAB) were agreed to in 1998, to supplement the existing General Arrangement to Borrow (GAB). At the end of 2001, total resources available to the IMF stood at \$125 billion, of which \$43 billion was available under the GAB and NAB facilities.

Steps were also taken to shorten the response time of IMF programs and to restructure programs to ensure that countries do not become overly

dependent on IFI resources. In 1997 the Supplemental Reserve Facility (SRF) was created, providing another type of loan arrangement for IMF programs. Explicitly short-term in nature (loans are expected to be paid back in 12 to 18 months and required to be paid back in 24 to 30 months) and carrying a higher interest rate than the more traditional stand-by arrangement, the SRF was designed to create incentives that would favor its use only by truly illiquid borrowers. Essentially solvent countries that have temporarily lost liquidity could afford the higher interest rates and would be able to repay any loan in a shorter period. Countries that have more fundamental problems would have recourse to programs with loans that would be paid back over a longer period.

To shorten response times, the IMF in 1999 created the Contingent Credit Line (CCL), a facility that allows countries with sound policies to prequalify for a line of credit that would protect against contagion in a systemic crisis. (Contagion refers to a sudden cutoff of private capital inflows to one country in response to a crisis in another.) Despite subsequent modifications to the terms of the facility, to date no countries have chosen to participate. This lack of interest appears to relate to the stigma that might be associated with seeking a CCL. Countries may worry that their pursuit of a CCL might be taken by market participants as a signal of problems in the country.

The extent to which the private sector should be involved in any solution to financial crises has been the most contentious issue in discussions of international financial system reform. Private sector involvement is generally taken to mean some sort of burden sharing or participation on the part of private creditors in the provision of financing to a country in crisis. Such burden sharing could be a formal part of the official program to aid the country. For example, IFI financing for the second program for the Republic of Korea in 1997 included an agreement by commercial bank creditors to extend the maturity of their loans to Korea. Burden sharing could also come about through a reduction in the value of private sector claims against the distressed country; a reduction in principal was part of Ecuador's restructuring of its debt, for example (Box 7-2). Absent such commitments by private creditors, policymakers worry that crisis financing provided to a country by the official sector may only serve to reduce the losses that private sector creditors would otherwise bear. This might encourage lenders to behave less prudently in the future, raising the moral hazard concerns discussed above.

In September 2000 the IMF released a framework for advancing the discussion on private sector involvement. The framework encourages countries and private lenders to make every effort to forestall crises through a variety of measures. Borrowers and lenders are to use information provided under the transparency and accountability initiatives discussed above, as well

Box 7-2. Crisis and Restructuring in Ecuador

Ecuador's experience in 1999 and 2000 presents an interesting case, in that during this time it became the first country to default on Brady bond obligations. (Brady bonds were issued by 18 governments between 1990 and 1997, under a plan proposed by the then-Secretary of the Treasury. The Brady Plan offered a means for sovereign countries to restructure past-due loans extended to them by commercial banks, by converting the loans to bonds.) Ecuador's decision to default was not taken lightly and was explained by dire economic circumstances. Output had stagnated in 1997 and had fallen sharply in 1998 because of declining oil revenue and agricultural and coastal infrastructure damage due to the El Niño effect. Many firms came under financial pressure, compounding difficulties in the banking sector. Over the first half of 1999, real GDP fell at an annual rate of 15.4 percent.

The decline in economic activity made it difficult for Ecuador to service its external debt. Ecuador's poor prospects, and financial markets that were destabilized by the Russian default in 1998, precluded new private lending. In late August 1999 Ecuador announced it would defer a coupon payment on PDI (past-due interest) Brady bonds, but in September Ecuador made payment on its discount Brady bonds. Creditors disliked the idea that Ecuador had tried to limit default to one type of Brady bond, and shortly thereafter bondholders accelerated their claim for full payment of outstanding interest and principal on all Brady bonds. As a result, Ecuador defaulted on its other Brady bonds and its Eurobonds as well.

At the same time, the IMF announced it would approve a stand-by arrangement if Ecuador would make certain recommended changes to its economic policies and pursue good-faith efforts to reach a collaborative agreement with its creditors. However, no agreement was reached. To facilitate restructuring of the debt, Ecuador established a consultative group consisting of representative institutional bondholders. The group was given economic and financial information, which was simultaneously made public. No confidential economic information was shared with the group, nor was any information about the terms of the planned restructuring. Although there were many one-on-one meetings between the Ecuadorian authorities and major bondholders, in general there were no large-scale negotiations with the bondholders. Unfortunately, this process failed to provide a meaningful forum. With the rapid turnover of finance ministers and a lack of political consensus, it was hard for Ecuador to sustain a dialogue until political stability was restored.

continued on next page...

Box 7-2.—*continued*

Consultations continued over the next several months with no progress. Private investors expressed concern that Ecuador had shown little willingness to engage in open dialogue or negotiations, and about the slow pace of progress. In January 2000 President Jamil Mahuad announced that Ecuador would convert its monetary base from the local currency, the sucre, to the U.S. dollar and adopt the dollar as the country's official currency (the sucre had depreciated more than 65 percent in 1999). Shortly thereafter, Vice President Gustavo Noboa assumed the presidency after President Mahuad was deposed in a popular uprising. President Noboa continued with dollarization, with the support of the IMF. The new political regime made progress in restructuring negotiations, and in March a \$2 billion aid package was announced, which was funded by the IMF, the World Bank, the Inter-American Development Bank, and Corporación Andina de Fomento. The loans were designed to assist the implementation of dollarization, to resolve the banking crisis, and to strengthen the public finances.

In mid-May 2000 the Ecuadorian authorities held an open meeting with bondholders to discuss the country's economic prospects. IMF staff also attended and presented key features of the new economic program. Bondholders received the details with interest, and in August, 98 percent of them accepted a debt exchange offer. A combination of exit consents and cash incentives provided the motivation to accept the package. (Exit consents allow the majority of bondholders to exercise their power to amend old debt just before these creditors leave the old debt and accept the new debt. This provides an incentive for all other holders to come along with them.) With the exchange, Ecuador reduced the face value of its debt by roughly 40 percent, realizing a projected cash flow savings of \$1.5 billion over the succeeding 5 years.

Since the restructuring of its debt and the implementation of the IMF program, Ecuador's economy has recovered strongly. Real GDP growth for the year ending in the third quarter of 2001 was 5.0 percent. Dollarization pushed inflation down from 91 percent in 2000 to 22 percent at the end of 2001. Interest rates on 10-year bonds were roughly 12 percentage points above those on U.S. Treasuries at the end of 2001, down from 46 percentage points at the height of the crisis in September 1999. Although the banking system has improved, there is room for further reform, such as implementation of key Basel principles. Analysts point to restructuring nonperforming loans and additional structural economic reforms as keys to further boosting economic activity in Ecuador.

as to maintain continuing dialogues, perhaps through the establishment by borrowing countries of investor relations offices. The IMF itself, in July 2000, formed the Capital Markets Consultative Group to enhance communication with the private sector. Lenders are also encouraged to promote the inclusion of collective action clauses in future bond issues (discussed further below), to allow for easier coordination of creditors in the event of a crisis.

The framework stresses that, should a crisis develop, voluntary solutions between debtors and creditors are to be preferred over involuntary solutions that involve unilateral actions. In most cases, it is hoped that policy adjustments and temporary official financing will suffice to restore an economy to sustainability. In a minority of cases, however, the official sector is envisioned as encouraging creditors to reach voluntary agreements to help overcome their coordination problems.

In some such cases, the country may have no choice but to suspend payments on its debt. The IMF has reaffirmed its policy of “lending into arrears” in such cases, that is, providing lending to countries that are experiencing debt-service difficulties before those difficulties are fully resolved. Lending into arrears is to be decided on a case-by-case basis and is to occur only where prompt IMF support is considered essential for a successful adjustment program, and the country is pursuing appropriate policies and is making a good-faith effort to reach a collaborative agreement with its creditors. This policy came into play in the case of Ecuador’s 1999 default, mentioned above.

Debt Relief. Finally, reform efforts have also included addressing the debt burdens of the poorest countries. After some gradual efforts in the late 1980s and early 1990s, the IMF and World Bank executive boards, at the request of the G-7, agreed in 1996 to launch the Heavily Indebted Poor Countries (HIPC) initiative. This initiative marked the first time that multilateral, Paris Club, and other official bilateral and commercial creditors joined in an effort to reduce the external debt of the world’s poorest and most debt-burdened countries. (The Paris Club is the voluntary gathering of governments of creditor countries willing to treat in a coordinated way the bilateral debt due them by developing-country borrowers.) The HIPC initiative is funded by both bilateral and multilateral creditors. Originally, 41 countries were identified as candidates for the program, and so far 24 of these have debt relief agreements in place. To qualify for assistance under the HIPC initiative, a country must meet three conditions: it must have a low enough income per capita to qualify for concessional lending from the IMF and the World Bank; it must have an unsustainable debt burden even after the exhaustion of available debt-relief mechanisms; and it must have demonstrated a commitment to economic reform and poverty reduction with a track record of good performance and drawn up a Poverty Reduction

Strategy Paper (PRSP) showing how the country intends to use debt relief to improve living standards for its poor.

The first 3 years of the initiative did not prove as productive as had been hoped: only seven countries qualified during that time. In September 1999 the program was enhanced to provide deeper and faster debt reduction. The HIPC initiative will allow 24 countries to reduce the net present value of their debt by a total of \$22 billion—roughly half of what they owe—and when combined with traditional debt relief and additional bilateral debt forgiveness, it will reduce their debt by almost two-thirds. The IMF and the World Bank expect average social spending in the HIPC countries to increase by 45 percent in 2001-02 from 1999 levels, with savings from HIPC debt relief accounting for a sizable proportion of this increase. In 2001-02 these countries are expected to spend three times more on social services than debt service.

Critiques of Reform Efforts

As the above discussion makes clear, many changes have been made to the international financial system over the past 7 years in an effort to improve its stability and performance. However, fundamental problems remain, and new proposals have been put forward by both private sector and public sector entities. Critiques of the efforts to date can be broken down into the same four key areas discussed above: transparency and accountability, strengthening national financial systems, management of crises, and debt relief.

Reform efforts appear to have made the most progress in enhancing transparency and accountability and strengthening national financial systems. Nevertheless, several complaints have been raised. With regard to accountability, critics often raise objections to “mission creep” on the part of the IFIs, which can lead to an overlap of efforts that hinders accountability. Without a precise understanding of each IFI’s responsibilities, it is difficult to judge the degree to which each IFI is accomplishing its objectives. The IMF draws on its expertise to consult and provide helpful advice on such matters as the appropriate stance of monetary and fiscal policy as well as the related choice and operation of an exchange-rate regime. At the same time, the MDBs have considerable expertise in development issues, both at the individual project level and in providing fundamental public goods such as health and education. Most recently, the MDBs have contributed substantial sums to programs for such middle-income economies in crisis as Argentina and Turkey, which, until their crises broke, had benefited greatly from private capital inflows. The MDBs should not be used as a source of immediate emergency financing. Rather, their role in crisis countries is to provide support to address longer term policies and institutional capacity building, to help cushion the impact of crises on the poor.

Thus almost all observers have argued for a clearer delineation of the IFIs' responsibilities, allowing each institution to focus on its core mission and expertise. Mission creep into other areas only serves to divert scarce expertise away from its best use. The IFIs have responded to this criticism and have taken steps to better coordinate their assistance, most noticeably through joint participation in the preparation of ROSCs and FSAP reports.

Progress on transparency has also been uneven, both on the part of borrowing countries and on the part of the IFIs. As mentioned earlier, the IFIs have made great strides in making information available to the public; nonetheless, market participants remain critical of what they regard as the scant and untimely release of information from the official sector during crisis resolution and negotiations. These criticisms have been directed toward the IFIs and even more pointedly toward the Paris Club. Without sufficient information and coordination, private creditors worry that their claims on a borrowing country will be treated less favorably than the claims of government and other official creditors. The Paris Club has begun taking steps to improve information flow, with the launch of a website disclosing the terms of debt restructurings and other information. The Paris Club has also initiated a dialogue with private sector creditor organizations in an effort to improve communication.

Efforts to strengthen national financial systems have focused on using agreed standards and codes aimed at implementing best practice in financial regulation. This effort has been judged quite promising, although implementation remains an area of concern. In particular, it may be expensive for developing economies to find and develop the expertise necessary to observe the standards and codes. For example, recruiting, training, and retaining skilled bank examiners may be difficult. The standards also require certain supporting institutions. In a country where the rule of law is weak, it may be difficult for financial examiners to make a real difference in financial institutions' practices. Finally, there has been some concern over the appropriate body to judge an economy's compliance with a standard. Local authorities may be too prone to find their own country's institutions in compliance, and the same might be true for IFIs that happen to be lenders to the country. There is no reason why private markets could not provide the necessary evaluation of compliance; indeed, this option has been advocated by many but has not yet been fully realized.

Efforts to reform the management of financial crises have generated the most criticism and the most additional proposals. The criticisms have focused on essentially two areas: the structure of IFI programs, and mechanisms for facilitating private sector involvement. Much attention has been paid to the conditions imposed on borrowing countries as part of IFI lending programs, called "conditionality." Some observers have argued that such conditions have too often involved overly restrictive austerity policies, which

have deepened economic slumps and postponed recovery. IMF programs during the East Asian crisis, which required fiscal austerity of economies, are often cited in this context. Critics have also argued that IMF programs should have allowed for more accommodative monetary policies, on grounds that high interest rates made it harder for debtors to service their debt, heightening investors' concerns and worsening the economic downturn. However, the IMF still argues that high interest rates, in relation to both expected inflation and interest rates on U.S. dollar-denominated assets, were necessary to stabilize currencies, whose depreciations also made it difficult for debtors to service their foreign currency-denominated debt.

According to another view, IFI programs too often went beyond macroeconomic (fiscal and monetary) conditions to impose unnecessary structural economic reforms. This view claims that the problems of debtor countries largely require macroeconomic solutions, and that therefore it is reasonable for the IFIs to insist on macroeconomic performance criteria to be met as a condition for loan disbursements. But in the late 1990s, some observers feel, the IMF often overstepped these bounds—and its own expertise—by placing too much emphasis on micromanagement of the recipient economies. An often-cited example is the Indonesian program, which required the elimination of the Clove Marketing Board and changes in the structure of the sugar, flour, and cement markets. Defenders of the existing approach have responded that, without a change in structural conditions, changes in macroeconomic policies are likely to have little effect. They also note that involvement of the MDBs in crisis lending provides whatever microeconomic and structural expertise is required. In any case, in response to these criticisms, the IMF has recently sought to streamline the conditionality attached to its lending programs, and to focus that conditionality on core macroeconomic and financial concerns.

Frustration with a lack of progress in some countries, as evidenced by repeated IMF programs over a prolonged period, raises another issue concerning the structure of these programs. For example, since 1980 the Philippines has been under six IMF programs, with disbursements made in 17 of the past 21 years. This example raises the concern that more attention should be paid to the nature of the crisis facing an economy. It may be necessary to tailor program lending differently for liquidity crises than for insolvency crises. In a liquidity crisis, where an otherwise healthy borrower is incapacitated by a cutoff in private financing, programs would appropriately involve short-term lending at penalty interest rates, to encourage and facilitate the borrower's quick return to private capital markets. In the case of an insolvent borrower, in contrast, where private funds are cut off because of poor economic prospects, the IFIs should not provide financing to avoid a debt restructuring. However, in such cases the IMF may still have a role in helping to support the country and facilitate the rebuilding of reserves, as

happened in Ecuador (see Box 7-2). Although the IFIs have different types of lending facilities for each of these two purposes, the repeated occurrence of “crises” in some economies suggests that sufficient attention was not paid to the possibility that recipients were insolvent rather than illiquid.

The issue of private sector involvement in the resolution of crises remains the most contentious, as evidenced by a recent flurry of proposals and analysis. Proposals to enhance private sector involvement range from the very modest (limiting involvement to the voluntary modification of sovereign bond contracts), to somewhat structured proposals involving standstills (temporary suspension of debt service), to formal proposals calling for an international recognition of standstills in a manner similar to an international bankruptcy proceeding.

Many observers, including the IMF, continue to urge that new sovereign bond issues include collective action clauses. One type of clause allows for a majority or supermajority of creditors to make changes in the financial terms of a bond’s contract; bonds issued under United Kingdom law typically contain such provisions. These clauses attempt to foster an orderly negotiation process that would allow the debtor country to reach agreement with its creditors on a restructuring that permits a return to a sustainable situation. However, many sovereign bonds are issued under jurisdictions, including that of New York, where collective action clauses are not customary. These bonds often require the unanimous approval of creditors to modify the payment terms. In this situation, a single holdout creditor, in hopes of obtaining more favorable treatment than the other creditors, can block a restructuring that is in the best interest of both the creditors and the debtor. It remains a bit of an economic mystery why more recently issued bonds do not include less restrictive collective action clauses; empirical work finds that borrowers do not face a higher interest rate on instruments that have this flexibility. One explanation may be simple inertia.

The modification of sovereign bond contracts in a sense represents an attempt to facilitate restructuring of private debt by creating an appropriate legal framework. Two other ideas have been advanced along the same lines. One proposal calls for more widespread use of rollover clauses in lending contracts, representing a precommitment by lenders that could be invoked during a crisis. This proposal would make automatic the rollover of bank loans like that negotiated in the case of Korea in 1997. Another recent proposal would generate private sector involvement before a crisis, by taxing the stock of cross-border claims to create a fund that could then be used for lending in the event of a crisis. All cross-border investors would thus contribute to the resolution of a country’s crisis.

A recent joint proposal from the Bank of Canada and the Bank of England advocates the use of standstills by insolvent debtor economies. The proposal calls for tight limits on IMF lending for all but exceptional cases, in an

attempt to force a distinction between insolvent and illiquid borrowers. A borrower that could not meet its obligations through this limited IFI support would declare a payment standstill and begin negotiations with its creditors on a debt restructuring. This would put the borrower in violation of the payment terms of its loan agreement, opening the door to legal action by creditors that might disrupt the negotiations. However, the proposal argues that fears of such disruption are overstated. Private creditors find it difficult to execute judgments against a sovereign borrower, especially when the borrower does not have readily identifiable assets, such as those of state-owned enterprises, outside its borders. Critics of the proposal counter that the cloud of legal action could nevertheless weigh on negotiations during the standstill, especially if cooperative creditors fear that any new payment arrangements agreed to could be subject to attachment by holdout creditors. The recent experience with the holdout creditor Elliott Associates in the case of Peru is cited in this regard (Box 7-3).

At roughly the same time that the Bank of England/Bank of Canada proposal was announced, the First Deputy Managing Director of the IMF called for a framework that would create the analogue of bankruptcy at the sovereign level, providing legal protection for a necessary restructuring. The proposal cites specifically the troubling implications of the Peruvian case. Legal protection from holdout creditors would be offered under two conditions: the country must be negotiating in good faith with its creditors to restructure its debt burden, and it must agree to follow sound policies to avoid similar problems in the future. The proposal also envisions that participating borrowing countries would likely impose temporary exchange controls, to ensure that capital did not flee the country while negotiations with creditors were under way. The protection from litigious creditors, in effect a formal standstill, would be sanctioned by the IMF and would have legal standing in national courts.

Implementation of the IMF proposal might take many years, because the IMF's Articles of Agreement would have to be amended, as might national legal codes around the globe. Some criticism of the proposal has focused on the impracticality of implementing these changes. Other critics argue that because the IMF might well be one of the creditors in the case, an IMF-sanctioned standstill would create a potential conflict of interest. (In domestic bankruptcy cases, the judge who presides over the resolution may not be one of the creditors of the troubled firm.) Other observers, however, note that any internationally sanctioned proceeding would not be able to remove the "management" of the debtor economy (that is, its government), also unlike in domestic bankruptcy proceedings. In that case involvement of an official creditor, such as the IMF, that can impose conditions on new lending programs may make sense. In any event, the IMF proposal has generated a great deal of interest and calls for further study.

Box 7-3. Elliott Associates versus Peru

In October 1995 Peru announced an arrangement under the Brady Plan (see Box 7-2) to restructure loans extended to two Peruvian banks that had been guaranteed by the government in 1983. The plan culminated in November 1996 with 180 creditors agreeing to exchange the old debt for a combination of Brady bonds and cash. Under the agreement, coupon payments on the new Brady bond were to begin in March 2000, with the second coupon to be paid in September 2000.

From January through March 1996, as details of the plan were being negotiated, Elliott Associates, an investment fund specializing in the purchase of securities of distressed debtors, bought Peruvian bank loans with a face value of \$20.7 million for \$11.4 million. After sending a formal notice of default on the bank loans, and shortly before the Brady exchange, Elliott Associates filed suit in New York State's Supreme Court seeking payment. Elliott did not participate in the Brady exchange, thus becoming a "holdout creditor." Elliott's suit was removed to Federal district court where, after a trial, the claim was dismissed in August 1998.

In dismissing Elliott's claim, the district court ruled that Elliott had purchased the Peruvian bank debt with the intent and purpose of bringing suit. This was found to be a violation of Section 489 of the New York Judicial Law, which is based on the long-standing legal concept of champerty. (Champerty is defined as maintaining a suit primarily in return for a financial interest in the outcome.) However, in October 1999 the U.S. Second Circuit Court of Appeals overturned the district court's ruling. The case was remanded to the Federal district court, which in June 2000 awarded Elliott a judgment of \$55.7 million, representing principal and past-due interest on the bank claims.

To enforce this judgment, Elliott sought to attach the September 7, 2000, coupon payment that was to be made to the creditors that had participated in the Brady exchange. Elliott obtained a restraining order to prevent the New York fiscal agent for the Brady bond from making the coupon payment, and the firm tried to obtain a similar order against the European fiscal agent. After arguing in the Belgian courts without Peru's attorneys present, Elliott was granted the restraining order on appeal on October 5, 2000. By this time Peru was close to defaulting on the Brady bond, as the 30-day grace period for the coupon due on September 7 had almost expired. Rather than default, Peru settled with Elliott by paying the firm \$56.3 million (the judgment amount of \$55.7 million plus interest). Thus the case was not litigated to a conclusion, leaving market participants uncertain about any precedents that the case might have set.

continued on next page...

Box 7-3.—*continued*

In issuing the restraining order, the Belgian court accepted the argument that, by paying the Brady bondholders but not paying Elliott, Peru would violate the *pari passu* clause in the bank loans held by Elliott. (The Latin phrase *pari passu* means “with equal step” or “side by side.”) The court interpreted the *pari passu* clause as meaning that if a debtor does not have enough money to pay its creditors in full, they all should be paid on a *pro rata* basis. This interpretation has proved controversial, however, with some legal scholars arguing that the clause relates only to the act of subordinating one class of creditors to another and should not be interpreted so as to force *pro rata* payments. These scholars base their arguments on the interpretation of *pari passu* clauses in domestic corporate bankruptcies.

This case is economically important for the effects it might have both on other developing economies’ attempts to restructure their debt and on future capital flows to these economies. The incomplete resolution of the case leaves open the possibility that other creditors might follow the example of Elliott Associates in holding out on future debt restructurings by developing economies—and that they might succeed. In particular, some argue that the Belgian court’s acceptance of Elliott’s *pari passu* argument could complicate Argentina’s current effort to restructure its debt. Creditors may hesitate to participate in any restructuring offers if they believe that holdout creditors might be able to attach payments or even get paid in full. Most observers argue that the relative balance of power between creditors and distressed sovereign borrowers would have been unchanged had the *pari passu* argument failed.

With regard to future capital flows, the concern is that if Peru had prevailed in the case on its champerty defense, it could have made it easier for sovereign countries to default on their debt. In that event, creditors might have contemplated curtailing lending to developing economies, or charging a higher interest rate. The Second Circuit Court of Appeals decision cited these concerns in overturning the district court’s champerty finding. In any event, both market participants and legal scholars agree that a final legal resolution of the issues raised in this case would eliminate a source of uncertainty now complicating transactions in the market for developing-country debt.

Finally, with regard to debt relief, although the HIPC initiative has already provided significant relief, it will not ensure a lasting exit from debt problems unless the countries receiving relief sustain growth far in excess of their historical averages. Real GDP growth in 22 eligible HIPCs averaged only 3.1 percent from 1990 through 1999, yet the IMF projects that they will grow at an annual rate of 5.6 percent from 2000 through 2010. Skeptics find little reason to be so optimistic, as many of these countries were already on IMF programs and receiving disbursements to begin with. If growth falls well short of the IMF's projections, it could be difficult for these countries to reduce their debt burden, even with HIPC debt relief. Most of the HIPCs depend heavily on exports of a narrow base of primary commodities, such as coffee or cotton, to service their external debt. Commodity prices can be quite volatile, leaving these countries vulnerable to price shocks. What might help this situation is if the industrial economies, which now spend \$360 billion a year on subsidies to protect their own agricultural sectors, lowered these barriers to trade, thereby allowing the HIPCs and other developing countries to diversify their export base.

Advancing International Financial System Reform

The need for continued reform of the international financial system has generated a rich debate. Clearly, the benefits of global economic integration must be made available to all the world's citizens, and the support of the official sector is key to ensuring the smooth operation of the global trading and financial systems that underpin continued integration. At the same time, it must be recognized that official sector resources are finite and do not come out of thin air. Resources may be provided in the form of loans to developing economies, but these resources still come from public funds. As such, they are obtained from taxpayers across the globe and have an opportunity cost in terms of other governmental priorities. Both of these considerations argue for a careful assessment of costs and benefits when designing and using the international financial system.

With these ideas in mind, a set of principles for the IFIs can be identified. First, all of the above arguments and examples point to the need to differentiate between those countries that are temporarily illiquid and those that are insolvent. Although this distinction can be difficult in practice, it is crucial for good stewardship of official sector resources. Shortening the maturity of official loans may help make this distinction. Some observers have claimed that short maturities for official loans are too constraining, arguing that it is hard to help an economy by extending a loan that must be repaid in 12 to 18 months. However, if it is clear that such a loan is unlikely to be repaid, then it is more likely that the economy is insolvent rather than just illiquid. An illiquid economy should be able to regain access to capital markets in this

period of time; an insolvent economy will not be able to. Insolvent economies require more drastic treatment, such as a restructuring of debt obligations coupled with limited and longer term official sector lending once the restructuring is well under way.

Official funding can also be leveraged with private sector involvement. Future design changes to the international financial system must continue to focus on incentive mechanisms that encourage involvement of the private sector. Financing that is dedicated to encouraging a voluntary restructuring is one example of such a mechanism. Such financing can serve as a catalyst in returning a troubled economy to a sustainable footing.

In the first half of the 1990s, a set of International Development Goals were developed from agreements and resolutions adopted at world conferences hosted by the United Nations. The goals found a new expression in the Millennium Declaration of the United Nations in September 2000. Most of the world's poorest countries, particularly those in Sub-Saharan Africa, are falling well behind in achieving these International Development Goals in basic education, health, and poverty reduction. The President has called for a bolder move away from loans toward grants for the poorest countries. This approach, coupled with the progress under the HIPC initiative, holds the promise of higher living standards for the least fortunate, as it would facilitate productivity-enhancing investments without adding to their debt burden. In addition, grants to the poorest economies should be targeted toward those basic needs, such as education and health, that are vital to a growing and vibrant economy. In particular, grants can lead toward a redirection of resources to combating scourges such as HIV/AIDS that tear at the very fabric of society.

Consistent with the Administration's efforts to shift the MDBs' emphasis toward grants for low-income countries is its continued efforts to make these institutions more efficient and more focused on productivity growth in developing countries as a core objective. Careful selection of programs and a greater attention to results are the two key principles underpinning the U.S. MDB reform exercise. This means that the MDBs must do a much better job in sharpening the focus of their activities, concentrating on basic development work and working collaboratively among themselves and with other donors to ensure a development framework that is consistent and efficient.

The United States has also accorded particular importance to a comprehensive review of the pricing of MDB loans, to explore the possibility of greater differentiation of lending terms. Price differentiation is crucial to achieve greater lending selectivity based on differences in the development impact of individual operations and in borrowers' income per capita and creditworthiness, with preferential treatment for priority core social investments.

Finally, tying official support to efforts at creating trade can dramatically leverage any financial assistance provided to illiquid economies. As this chapter has made clear, trade is a powerful engine for economic growth and improvements in living standards. If assistance packages allow an economy both to regain access to capital flows and to invigorate trade flows, all of the developing world will share in the improvement of world living standards.

Conclusion

International flows of resources, goods, and services have played an increasingly important role in the world economy. The citizens of the United States, living in one of the most open economies in the world, have seen their well-being improve dramatically with this increased economic integration. So have the citizens of many other countries that were willing to open their borders to flows of goods, services, and capital. The gains from trade are the result of an improved allocation of resources. A more efficient global allocation of productive inputs such as capital and labor translates into an increase in global output and consumption.

To ensure that economic integration continues, constant attention must be devoted to the institutional infrastructure that supports market-based exchanges of goods, services, and capital. The past year has witnessed signs of a slowing global economy, as well as violent threats to the freedom that is essential to a well-functioning economic system. These dangers make it more important than ever to ensure continued progress toward the free flow of resources and output across national borders.

It is therefore critical that the United States remain an active leader in the continued liberalization of trade in goods and services, both on a bilateral and on a multilateral basis. At the same time, the United States must continue to encourage efforts to strengthen the international financial system that supports production-enhancing cross-border flows of capital. Strong U.S. leadership on both these fronts will help safeguard and enhance both our own economic prospects and those of the rest of the world.