

12077

[REDACTED]

[REDACTED]

MICROFICHE COPY

Report No.: 12047 Type: (PUB)
Title: PROCEEDINGS OF THE WORLD
Author: WORLD BANK
Ext.: Room: X Dept.: X
BOOKSTORE MARCH 1993

Supplement to THE WORLD BANK ECONOMIC REVIEW
and THE WORLD BANK RESEARCH OBSERVER

- KEYNOTE ADDRESS The State and Economic Development: A Mexican Perspective, Pedro Aspe and José Angel Gurria
- Toward a Counter-Counterrevolution in Development Theory, Paul Krugman
- Two Strategies for Economic Development: Using Ideas and Producing Ideas, Paul M. Romer
- Labor Market Institutions and Policies: Help or Hindrance to Economic Development? Richard B. Freeman
- Labor Market Adjustment in Transitional Economies, Jan Svejnar
- Information Flows and Discrimination in Labor Markets in Rural Areas in Developing Countries, Andrew D. Foster and Mark R. Rosenzweig
- Knowledge, Property, and the System Dynamics of Technological Change, Paul A. David
- Accumulating Technological Capability in Developing Countries, Martin Bell and Keith Pavitt
- Technology Gaps between Industrial and Developing Countries: Are There Dividends for Latecomers? Howard Pack
- International Capital Flows to Latin America: What is the Promise? Pedro-Pablo Kuczynski
- Capital Flows to Developing Countries: Implications from the Economies in Transition? Susan M. Collins
- Conversion of Official Bilateral Debt: The Opportunities and the Issues, Stephany Griffith-Jones
- ROUNDTABLE DISCUSSION The Health of the Public: A Public Responsibility? Julio Frenk, Eyitayo Lambo, W. Henry Mosley, and Uwe Reinhardt

PROCEEDINGS OF THE WORLD BANK
ANNUAL CONFERENCE
ON
DEVELOPMENT ECONOMICS
1992

Supplement to THE WORLD BANK ECONOMIC REVIEW
and THE WORLD BANK RESEARCH OBSERVER

EDITORS

Lawrence H. Summers and Shekhar Shah

EDITORIAL CONSULTANT

Elinor Berg

BANK

The World Bank Annual Conference on Development Economics is a forum for discussion and debate of important policy issues facing developing countries. The conferences emphasize the contribution that relevant policy, empirical, and basic economic research can make to understanding development processes and to formulating sound development policies. Conference papers are written by researchers in and outside the World Bank. The conference series was started in 1989.

The *Proceedings of the World Bank Annual Conference on Development Economics* is published as a special supplement to the World Bank's professional economics journals, *The World Bank Economic Review* and *The World Bank Research Observer*. All current subscribers receive this supplement to the journals. The Editorial Boards of the *Review* and the *Observer* do not review the *Proceedings*, nor are the papers selected from the conference and published here subject to the peer review process through which regular submissions to these journals must pass. Summaries are included of selected floor discussions and the roundtable discussion; they attempt to convey the sense and substance of what was discussed, interventions by participants from the floor, and responses by panelists. They have not been reviewed by the authors, the discussants, or the participants.

The views and interpretations expressed in this *Proceedings* volume do not necessarily represent the views of the World Bank or its member countries and should not be attributed to the World Bank or its affiliated organizations.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. Manufactured in the United States of America.

First printing March 1993

ISSN 1014-7268

ISBN 0-8213-2211-7

Material in this journal is copyrighted. Requests for permission to reproduce articles should be addressed to the Office of the Publisher, the World Bank, Washington, DC 20433, U.S.A. The World Bank encourages dissemination of its work and will normally give permission promptly and, when the intended reproduction is for noncommercial purposes, without asking a fee. Permission to copy portions for classroom use is granted through the Copyright Clearance Center, 27 Congress Street, Salem, MA 01970, U.S.A.

The World Bank Economic Review is indexed in *Current Contents / Social & Behavioral Sciences*, the *Social Sciences Citation Index*, the *Journal of Economic Literature*, and the *Public Affairs Information Service*. *The World Bank Research Observer* is indexed in the *Journal of Economic Literature*, the *Standard Periodical Directory*, the *Public Affairs Information Service, Inc.*, and, online, by the *Economic Literature Index* and *DIALOG*. It is available in microform through *University Microfilms, Inc.*, 300 North Zeeb Road, Ann Arbor, MI 48106, U.S.A.

PROCEEDINGS OF THE WORLD BANK
ANNUAL CONFERENCE
ON
DEVELOPMENT ECONOMICS
1992

Supplement to THE WORLD BANK ECONOMIC REVIEW
AND THE WORLD BANK RESEARCH OBSERVER

Introduction <i>Lawrence H. Summers and Shekhar Shah</i>	1
Opening Remarks <i>Lewis T. Preston</i>	7
KEYNOTE ADDRESS The State and Economic Development: A Mexican Perspective <i>Pedro Aspe and José Angel Gurría</i>	9
GROWTH AND DEVELOPMENT THEORIES	
Toward a Counter-Counterrevolution in Development Theory <i>Paul Krugman</i> Comment, <i>Joseph E. Stiglitz</i> 39 Comment, <i>Lal Jayawardena</i> 51 Floor Discussion 59	15
Two Strategies for Economic Development: Using Ideas and Producing Ideas <i>Paul M. Romer</i> Comment, <i>Kaushik Basu</i> 93 Comment, <i>Marcelo Selowsky</i> 99 Comment, <i>T. N. Srinivasan</i> 103 Floor Discussion 111	63

LABOR MARKETS AND DEVELOPMENT

**Labor Market Institutions and Policies: Help or Hindrance
to Economic Development? 117**

Richard B. Freeman

Comment, *Victor E. Tokman* 145

Comment, *Kwadwo A. Tutu* 151

Floor Discussion 153

Labor Market Adjustment in Transitional Economies 157

Jan Svejnar

Comment, *Marek Góra* 169

**Information Flows and Discrimination in Labor Markets
in Rural Areas in Developing Countries 173**

Andrew D. Foster and Mark R. Rosenzweig

Comment, *Nancy Birdsall* 205

Floor Discussion of Svejnar and Foster-Rosenzweig Papers 209

TECHNOLOGY

**Knowledge, Property, and the System Dynamics
of Technological Change 215**

Paul A. David

Comment, *Ashok Desai* 249

Comment, *Morris Teubal* 253

**Accumulating Technological Capability in Developing
Countries 257**

Martin Bell and Keith Pavitt

**Technology Gaps between Industrial and Developing
Countries: Are There Dividends for Latecomers? 283**

Howard Pack

Comment, *John Page* 303

Comment, *Jorge M. Katz* 307

Comment, *Carl Dahlman* 313

Floor Discussion of Bell-Pavitt and Pack Papers 317

INTERNATIONAL CAPITAL FLOWS

International Capital Flows to Latin America: What Is the Promise? 323

Pedro-Pablo Kuczynski

Comment, *Albert Fishlow* 337

Comment, *Mario I. Blejer* 341

Floor Discussion 345

Capital Flows to Developing Countries: Implications from the Economies in Transition? 349

Susan M. Collins

Comment, *András Inotai* 371

Comment, *Isbrat Husain* 379

Conversion of Official Bilateral Debt: The Opportunities and the Issues 383

Stephany Griffith-Jones

Comment, *Jonathan Eaton* 403

Floor Discussion of Collins and Griffith-Jones Papers 409

ROUNDTABLE DISCUSSION 413

The Health of the Public: A Public Responsibility?

Julio Frenk, Eyitayo Lambo, W. Henry Mosley, and Uwe Reinhardt

Introduction

Lawrence H. Summers and Shekhar Shah

The papers collected in this volume were presented at the World Bank's Annual Conference on Development Economics in Washington, D.C., April 30 and May 1, 1992. Now in its fourth year, the conference series is aimed at expanding the flow of ideas among development policy researchers and practitioners around the world. The Bank is an institution in which ideas are as important as finance. It not only disseminates current knowledge about development but is also engaged in furthering this understanding. As the largest international agency concerned with economic development, the Bank seeks to share with the development community the issues and concerns that staff members face on a daily basis, to expose Bank staff to recent research in economics, and to improve interaction with policymakers and researchers in developing and industrial countries. The annual conferences, and the wider dissemination of their deliberations in the proceedings volumes, contribute to these objectives.

The 1992 conference addressed four themes: theories of growth and development, the operation of labor markets, technology and its role in development, and international capital flows. The format of the conference allowed each theme to be explored at some length, first by the authors, then by comments from invited panelists, and finally through floor discussions. The 1992 conference continued the tradition of holding a roundtable discussion related to the subject of the next *World Development Report*: health policy in developing countries.

In his opening remarks, Lewis T. Preston, the president of the World Bank, emphasized the importance of development research at the Bank. Noting that the search for knowledge about development is vital, he emphasized the need to present such knowledge in clear, pragmatic terms that can be easily communicated and implemented in the diverse economic, social, and cultural circumstances of individual developing countries.

The keynote address by Pedro Aspe, Mexico's minister of finance and public credit, and José Angel Gurría, Mexico's undersecretary for international financial affairs, focused on the public sector and on Mexico's structural reform

Lawrence Summers is the World Bank's chief economist and vice president, Development Economics, and Shekhar Shah is deputy administrator of the World Bank's Research Advisory Staff.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

2 Introduction

program. Gurría, who delivered the address, provided an insider's view of the reform process and the transformation that has made the government leaner but has also strengthened its role in providing social infrastructure, health and education services, public safety, and the administration of justice.

How have theories of growth and development shaped our thinking on development policy? In a reassessment of what he terms "high development theory," Paul Krugman traced the ideas of Paul Rosenstein-Rodan, Albert Hirschman, and others who emphasized the increasing returns and pecuniary external economies that arise from the effects of market size. Krugman argued that these important ideas faded from mainstream economics not because of intrinsic logical flaws but because the authors failed to codify their ideas in internally consistent, formal models. These ideas also lost popularity in the face of the neo-classical counterrevolution against interventionist development models. New, more rigorous developments in international economics, growth theory, and industrial organization, Krugman argued, suggest that the time is now ripe for a "counter-counterrevolution" that would seek a middle ground between the extremes of a free-market orthodoxy and state control. Krugman's reading of the history of these ideas was challenged in spirited retorts by Joseph Stiglitz and Lal Jayawardena.

Paul Romer's paper incorporated "ideas" as a factor of production and contrasted alternative development strategies based on using existing ideas, as in Mauritius, and on producing ideas, as in Taiwan (China). He argued that it is crucial to distinguish ideas from human capital because ideas as economic goods often differ from both public and private goods. Moreover, explicit attention to the economic role of ideas can serve as a powerful tool for analyzing growth theories. Romer was ultimately pessimistic about the ability of countries to follow an interventionist industrial policy aimed at producing ideas rather than using existing ideas. He noted that until adequate national institutions for fostering the production of ideas come into existence, the gains from international integration offer the best hope for development.

In his paper on the role of labor market interventions such as minimum wages, job security, and collective bargaining, Richard Freeman contrasted the positions of the "distortionists" and the "institutionalists" regarding the desirability of such interventions. After assessing the evidence available to him, he concluded that neither side had a clear-cut case. The distortionist argument, he argued, suffers from the absence of a convincing and robust measure of labor market distortions and from its reliance on weak cross-country time-series data rather than detailed investigations of specific interventions in particular countries. Freeman suggested that the decline in real wages during the past decade in several developing countries may indicate that labor markets are much more flexible than the distortionists generally assume. Turning to the institutionalists, Freeman also found little evidence for the value of socialized pacts and consultative modes of labor market adjustment. Against this backdrop, Freeman proposed a different model of labor market institutions as serving the political

economy purpose of shifting attitudes and expectations of relative rewards in the labor market, and hence moving the perceived short- and long-run costs in favor of reform programs. He argued that such a pragmatic approach, based more on understanding the role of interventions and institutions in specific settings, may yield better insights for the design of adjustment programs.

Two other papers addressed the theme of labor markets. Andrew Foster and Mark Rosenzweig emphasized the importance of information flows in rural labor markets with heterogenous labor, on the basis of their examination of workers' actual productivity and the a priori productivity perceived by employers. Using a rich data set on time and piece-rate wages in casual labor markets, they presented empirical evidence on the significant effects of information barriers on employment discrimination and outlined the policy interventions that might reduce the cost of such information barriers. They pointed in their policy conclusions to the implications for worker mobility, permanency of labor contracts, and private returns to human capital (and therefore investments in such capital) and to the role that such interventions as public employment programs can play in influencing the structure of information flows. They also suggested that information on the form of worker payment (piece-rate versus time wages) can be very important and that inferences that ignore this distinction can yield misleading conclusions.

Jan Svejnar's paper, which surveyed labor market developments in Eastern Europe—particularly Poland, Hungary, and Czechoslovakia—asked why the microeconomic transformation had so far proved to be more difficult than the process of macroeconomic stabilization in these countries. Svejnar discussed the desirability of the wage norm—arguing that profitable enterprises in the tradables sector should be permitted to operate without a strict norm—and recommended the imposition of hard budget constraints. He argued for the use of social rather than private welfare criteria to deal with redundant labor in state enterprises. Finally, he noted that labor unions may not have been sufficiently involved in the design of reforms and could become strong opponents of the transition, particularly in the face of rapidly rising unemployment.

On the theme of technology, Paul David used a systems-theoretic approach to explore the relationships between technological change and the acquisition of knowledge. He argued against a linear, unidirectional scheme of technological change that placed fundamental scientific discoveries and experimental findings at one end, and imitation and diffusion of technology at the other end. He suggested, for example, that even if the interests of industrial and developing nations are identified, respectively, with these earlier and later stages, that in itself was not sufficient to clarify the current North-South disagreements on intellectual property rights. David called instead for an evolutionary approach to the analysis of interactions between science and technology, with attention to the dynamic feedbacks and nonlinearities that are often overlooked. In this context he examined the appropriability problem that affects the production and diffusion of information.

4 Introduction

In his paper on the role of technology in developing countries, Howard Pack examined the issue of "convergence" and asked whether low-income developing countries can benefit from being latecomers by using the stock of knowledge of more advanced countries, thereby achieving rates of per capita income growth that are higher than expected on the basis of their accumulation of physical and human capital. Pointing to several important empirical problems with the convergence hypothesis when it is tested using time-series data for total factor productivity growth, he argued that much of the unexplained growth can be shown to have come from the shift from agriculture toward manufacturing and services. Pack discussed the experiences of the Republic of Korea and Taiwan (China)—economies that have successfully shifted to a higher production function through technology and development.

Keith Pavitt and Martin Bell explored the lessons for developing economies from the experience of the industrial countries with technology development. They argued for a view of technological innovation and diffusion that is much more firm- and industry-specific and that takes into account the learning capabilities of firms in different policy settings. They also made the point that growing specialization and vertical disintegration have weakened the link between production capabilities and technological accumulation.

Turning to the final conference theme, international capital flows, Pedro-Pablo Kuczynski discussed the reasons for—and the promise of—the resurgence of capital into many parts of Latin America. He traced the internal and external causes for these reflows, emphasizing the significant changes in economic policies in virtually all of these countries, and called for the development of fixed-income capital markets to channel the inflows into essential infrastructural projects. Kuczynski highlighted the need for political and social stability in Latin America and observed that if such stability is to be achieved, it is imperative that the economic conditions of the poor improve rapidly.

In her paper, Susan Collins asked whether developments in Eastern Europe and the former U.S.S.R. are likely to divert capital flows from developing countries. Collins maintained that demand from Eastern Europe is unlikely to materially affect developing countries' prospects for private capital flows, which in any case appear limited. She examined the implications for all other aid recipients of the sustained increases in official development assistance to four countries (China, Egypt, Israel, and Bangladesh) since 1970 and found significant diversion in that specific case. To the extent that there are parallels between aid allocation for these four countries then, and Eastern Europe now, Collins suggested that some diversion of aid may occur.

Stephany Griffith-Jones, in her paper on the conversion of official bilateral debt, looked for parallels between such debt conversion, particularly by low-income developing countries, and the ongoing process of converting commercial debt, mainly in the middle-income countries. She examined the potential benefits and costs of official debt conversion and suggested that unlike commercial debt conversions that have focused on debt-equity swaps, official debt conver-

sion could profitably examine the scope for directing greater funding to social sectors, to poverty alleviation, and to the environment, on the grounds of high social rates of return.

As in previous years, the planning and organization of the 1992 conference was a joint effort. We both benefited from the guidance and support of Gregory Ingram, administrator of the World Bank's Research Advisory Staff, the organizational home of the conference. We would also like to thank other staff members, in particular the conference coordinator, Jean Gray Penchamni, whose superb organizational skills kept the conference on track. She was ably assisted by Mantejwinder Jandu. Finally, we thank the editorial staff, especially our consultant, Elinor Berg, as well as Sandra Gain and Patricia McNees.

Opening Remarks

Lewis T. Preston

I am pleased to be here this morning to inaugurate the Bank's Fourth Annual Conference on Development Economics. To all of you, a warm welcome.

I particularly want to welcome those of you who are attending the conference from other parts of the world. I believe it is imperative that the Bank remain an open institution, constantly looking for opportunities to strengthen its links with the development community.

This conference is an important opportunity for us to share our varied perspectives on the problems of development. The economics of development is central to the mission of the World Bank. Economic development and the alleviation of poverty have been the Bank's unchanged business for twenty years.

I want to stress that development research—the analysis of why development takes place in some settings, and why poverty persists in others—is also central to the Bank's mission.

As Larry Summers emphasized in his address to last year's conference, knowledge is essential for effective action in a world of finite resources. It is the search for ideas which could make a difference to development that has kept research at the center stage in this organization.

In this pursuit of worthwhile ideas, we must continue to examine development policy in pragmatic, operational terms that we can discuss with our member countries and that they can implement and sustain under diverse social and political circumstances.

Over time our research topics have broadened to include issues in health, education, and environmental management. Yet we have not abandoned our past focus on capital accumulation, industrialization, and the role of government. Even now we are hard at work improving our understanding of how these last three elements have successfully transformed Japan and the newly industrialized countries in East Asia.

We have learned that it is unrealistic to believe that the invisible hand of the market is all that is needed to promote development. Governments must lend the market a helping hand—by doing well those things that only the public sector can do and by avoiding those activities best left to the private sector.

Development economics is at once the oldest and the newest branch of economics. Adam Smith began, after all, with an inquiry into the nature and causes

Lewis T. Preston is president of the World Bank.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

8 *Opening Remarks*

of the wealth of nations, and yet there is much we do not know. Overall, advances in developing countries in the past twenty-five years have been impressive. Life expectancy has risen by ten years, infant mortality rates have been halved, and per capita income in low-income countries has nearly doubled.

At the same time, country-by-country experience has been highly variable. Thirty-six countries were poorer at the end of the 1980s than they were twenty-five years ago, and incomes declined on two continents in the past decade. We need to understand the reasons for this decline if we are to reduce poverty and improve living standards.

Increasing our knowledge about the determinants of development will require ingenuity and resources as well as insights about what policies to pursue. We also have much to learn from the different perspectives and experiences that each of us brings to the common pursuit of such knowledge.

Development needs the help of practitioners, policymakers, and scholars to identify the ideas and areas that should be pursued and to seek the approaches that are likely to yield the highest returns. It is only through an open dialogue that allows us to test each other's premises and assumptions about development and challenges past practices that we can hope to shape the intellectual foundations for effective development.

As several of the papers at this conference will tell us, ideas are seminal to successful development. We can generate new ideas, and we can adapt and use existing ones. I invite all of you to join both of these ongoing endeavors at the World Bank.

KEYNOTE ADDRESS

The State and Economic Development: A Mexican Perspective

Pedro Aspe and José Angel Gurría

Some time ago, Albert Hirschman (1958) said that development is like a jigsaw puzzle: it is easier to fit in a particular piece when the adjoining pieces are already in place; the pieces that are hard to find are those with only one neighbor in place. This clever analogy evokes two very important economic principles that both researchers and policymakers are rediscovering as we move from the decade of adjustment to a new period of reform and growth. The first one is that during the early phases of development, when an economy is no more than a collection of fragmented markets and regions, the establishment of government institutions, the construction of infrastructure, and the direct participation of the state in some areas of the economy are not only desirable but indispensable preconditions for the growth process.

The second principle is more in line with the recent theories of endogenous economic growth (see Scott 1991; Romer 1989; Lucas 1988; and Uzawa 1965). It reflects the notion that the opening-up of investment opportunities through changes in the environment where individuals work, save, and invest both creates and reveals new investment opportunities. In Hirschman's example, once the difficult parts of the puzzle have been solved, the remaining pieces begin to fall into place almost automatically. What this means for the role of the state in economic development is that after an initial period of protection and government intervention, growth no longer responds as strongly to further involvement as it did during the very first stages of industrialization. Furthermore, this analogy conveys the notion that once the basic institutional framework has been implemented, the public will be better served by indirect support of economic activity through deregulation, privatization, trade liberalization, and a competitive environment than by direct government participation in production activities.

With these two principles in mind, we would like to briefly elaborate on the consequences of structural reform of the public sector, using the Mexican experience as an example.

Pedro Aspe is Mexico's minister of finance and public credit, and José A. Gurría is Mexico's under-secretary for international financial affairs. The address was delivered by Mr. Gurría.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

From a long-term perspective, Mexico's years of low output growth and high inflation are more the exception than the rule. For example, between 1950 and 1970 Mexico had a remarkable macroeconomic record: gross domestic product (GDP) expanded at an average rate of nearly 6.6 percent a year, while annual inflation was below 4.5 percent. This era was in many ways the result of the ripening of institutions brought into being by the Mexican Revolution (1910–20), combined with the more stable performance of world markets. The development strategy was aimed at linking markets throughout the country by investing heavily in infrastructure, telecommunications, energy, and transport. Agricultural production was stimulated by land distribution, and industrialization was promoted through import substitution. These structural elements were backed by strict fiscal and monetary discipline. Throughout these years public sector borrowing requirements were consistently less than 3 percent of GDP, while the available instruments of monetary control emphasized price and exchange rate stability.

These results suggest that there was no reason to argue that Mexico's economic strategy—based on protection, financial repression, and the strong presence of the state—was not adequate. At the same time, some problems revealed that there was a limit to how far and how long this strategy could continue. For instance, in the late 1960s it was already apparent that domestic and foreign private investment were slowing down. One important reason for this was that protected oligopolies had already achieved strong market presence and had little incentive to keep on growing by increasing employment and productivity. Thus in the 1970s, pressed to offer work for a rapidly increasing population, Mexico had two alternatives: shift to an export-oriented economy, as the Republic of Korea did in 1965, or stay the course, filling the gap created by the declining role of private investment with government spending. At the time the authorities opted for the latter.

By 1972 Mexico was on the road to higher inflation. The public sector deficit, which was slightly more than 2 percent of GDP in 1971, began to grow, financed by an inflation tax and external borrowing, and reached 9.1 percent of GDP in 1976. Facing its first financial crisis since 1940—and the first devaluation of the peso against the dollar in twenty-two years—the government conceded that Mexico had to undertake a structural reform to restore growth and stability. However, the discovery of large oil reserves and the possibility of continued access to international financial markets not only postponed the needed reform but convinced the authorities that the economy could grow even faster. For some time GDP increased at about 9 percent a year, as inflation crept into the double-digit range, until 1982, when Mexico reached the limit of external borrowing capacity.

The 1982 crisis was the worst since the Great Depression. The event was not a temporary slowdown but the reflection of fundamental disequilibria that seriously jeopardized the long-term viability of the economy. The cumulative effects of previous fiscal and external imbalances created major hurdles. Overall, the

public sector borrowing requirement was 16.9 percent of GDP. Net external transfers in the following years amounted to 6 percent of GDP. On the structural side, the financial sector lacked the flexibility to halt capital flight and to intermediate savings efficiently; excessive reliance on oil exports made the fiscal and external accounts extremely vulnerable to terms of trade shocks; the tax system was seriously impaired by evasion and conspicuous subsidies; and the economy remained shielded from international competition by prohibitive tariffs and numerous quantitative restrictions. Trade policy, in particular, fostered inefficient private enterprises in the industrial sector, many of which had to be rescued by the government.

In retrospect, the purpose of government intervention during the 1970s and early 1980s was to maintain employment and economic activity while responding to the most pressing social needs of the population. The years of adjustment and crisis proved that by trying too hard, the authorities failed to meet both goals. A growth strategy based on expanding inefficient parastatals could not be sustained with a hard external budget constraint. Covering the losses of the parastatals diverted resources from investments in health, education, and infrastructure.

For example, Sidermex, a steel complex, accumulated losses of more than \$10 billion in a decade as a result of poor management, bad investment decisions, and severe union problems. Those same resources could have provided drinkable water, sewerage, and health and education facilities to all the marginal communities in the southeastern part of the country.

For the past ten years Mexico has followed an economic strategy that combines stabilization and structural change. President de la Madrid emphasized the permanent correction of macroeconomic imbalances, while the government of President Salinas de Gortari has stressed eliminating the institutional distortions that had inhibited the participation of the civil society in the development process.

On the macroeconomic side, the adjustment has been extraordinary. The budget deficit, which had reached 17 percent of GDP in 1982, recorded a surplus of 6 percent last year. The public sector balance registered a deficit of slightly more than 1 percent of GDP last year and is expected to record a surplus of 1 percent in 1992 (without the proceeds from privatizations). But far more important than the figures themselves is the way in which adjustment was implemented in the broader context of the reform of the state.

To make adjustment permanent, it was necessary to reduce the size of the general government; to privatize, shut down, or transfer parastatal enterprises; to renegotiate the external debt; and to put in place fiscal reforms that would lower tax rates and increase compliance. In addition, to reinforce the anti-inflationary thrust of adjustment, nontariff barriers were eliminated and tariffs were lowered and rationalized.

The outcome of the stabilization effort is encouraging so far. Inflation fell from an annualized rate of nearly 200 percent in January 1988 to 18.8 percent in 1991. After zero growth between 1982 and 1988, GDP has been growing at twice the rate of population growth for the past three years.

The old literature on development and the new views on growth point to the importance of government efforts to bridge missing markets and open opportunities for civil society. From the perspective of Mexico's experience, we should add a third element to this list: the social dimension of a leaner—but stronger—government.

Today, state participation in economic activity is very different from a decade ago. Growth policies are geared toward creating a propitious climate for the participation of individuals. Instead of subsidies and nationalization, the emphasis is on the elimination of institutional constraints on competition, the creation of new markets, and the generation of opportunities for all members of the population. Let me illustrate with some examples.

In the financial sector, the authorities have introduced new instruments to make possible intertemporal and intergenerational transfers and to hedge against inflation and exchange rate risks. Transactions that were impossible only three years ago are now encouraged through inflation- or exchange-rate-indexed bonds, flexible-rate and variable-maturity treasury bills, or specially designed individual retirement accounts.

The apparatus of development financing has also been revamped. In 1987 the National Industrial Development Bank (Nacional Financiera) allocated 94 percent of its total credit program to large (and inefficient) parastatal enterprises and the remaining 6 percent to small and medium-size enterprises. In 1991 it applied only 6 percent of total credits to parastatals and the rest to the private sector. Credit policy aims to provide access, not subsidies, to small entrepreneurs who are ineligible for financing through normal commercial bank lending because they lack a credit history.

Fiscal reform, apart from the international harmonization of rates and the effort to increase compliance, also focuses on integrating marginal producers into the formal economy through a transitional regime that does not require small firms to maintain complex accounting standards but obliges them to register revenues and expenses so that taxes can be calculated. Aside from the direct revenue implications, this reform helps small firms change from inefficient entities to competitive enterprises. The financial information they record for taxes also helps them to track their profit performance and to compare it against their competitors'.

In addition to the divestiture or closing of more than 900 of the 1,155 enterprises under state control in 1982, Mexico has deregulated ground and air transport, telecommunications, fisheries, textiles, automobiles, and foreign investment and has passed new laws on technology, patents, and trademarks. A recent constitutional reform gives legal title to very small holdings (known as *ejidatarios*), to communal workers, and to small farmers so that they are free to sell or lease the land or use it as collateral.

In the external sector the unilateral dismantling of trade barriers has been followed by a series of bilateral and multilateral negotiations for free trade areas. An agreement has already been signed with Chile, negotiations are now

under way with Canada and the United States to form a North American Free Trade Area, and talks are being held with Venezuela, Colombia, and the Central American countries. The new opportunities to conduct business have renewed enthusiasm for foreign and domestic investment. Foreign investment flows in 1991 reached \$12 billion, the largest amount ever recorded and almost equal to total foreign investment in Mexico prior to 1982.

The reform of the state, however, is not exclusively aimed at increasing economic efficiency and promoting faster growth. It also has a vigorous social dimension. The correction of fiscal disequilibria has gradually strengthened the state's capacity to provide basic social infrastructure, health and education, public safety, and the administration of justice. In 1982 the federal government transferred an amount equivalent to 12.7 percent of GDP to the parastatals; in 1991 that figure was 2.5 percent. In contrast, social spending, which in 1988 was 32 percent of total programmable spending and 6 percent of GDP, increased to almost 8 percent of GDP and 45.5 percent of programmable spending in 1991. (In pesos, this translates into a 60 percent real increase between 1988 and 1992.)

Decentralized social participation has been encouraged. The National Solidarity Program to fight poverty is an example of the new approach. The action program of Solidarity begins with the creation of a committee in a given town or locality. This committee provides the government with a proposal, and the project is undertaken cooperatively. The raw materials, such as bricks, cement, and wood, are provided by the government, the engineering skills by university students, and the labor by the local population. The program has disbursed \$1 billion a year since 1989 (0.4 percent of GDP) but has already achieved remarkable results because it has actively involved large and diverse sectors of the population. In three years 6 million people have been given access to health services facilities; 8 million people have been provided with drinking water; 11 million people in more than 10,000 communities now have electricity; a million and a half children are receiving education in the new school facilities refurbished by parents and teachers; and a quarter of a million children have been granted cash scholarships to complete their primary education. Hundreds of new projects and initiatives take shape every day across the nation. In 1992 the budget for this program is more than \$2 billion.

Our generation was educated in a context in which the greater presence of the state was considered a synonym of social justice. For many years when it was necessary to promote industrialization, the state created enterprises and public entities to channel resources and subsidies. It bought companies in financial difficulties to save jobs and support production. We now see that these were not always the right decisions. We now know that a bigger state is not necessarily a more capable state. From the perspective of the past twenty years in Mexico, a bigger state eventually caused lower growth, a diminished ability to respond to social needs, and, in the end, a weaker state. The reform of the state has meant a change from a paternalistic government to one that insists on market-driven measures, without forgetting how much is due to those who need more.

In the past decade the people and the government of Mexico have come a long way, but there is still a long way to go. Inflation must be brought down to international levels. Improvements are necessary in reregulation, deregulation, and the modernization of government institutions to encourage growth. Reform will only be completed, however, when all Mexicans have a decent standard of living, access to training, and job and business opportunities.

The new views of economic development challenge both government and civil society. They are based on the importance of the citizen and the proper organization and operation of the state. The ultimate goal is to establish an economy in which access to opportunities is the same for everyone within an open and fully democratic society.

REFERENCES

- Hirschman, Albert. 1958. *The Strategy of Economic Development*. New Haven, Conn.: Yale University Press.
- Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22(July):3-42.
- Romer, Paul. 1989. "Increasing Returns and New Developments in the Theory of Growth." NBER Working Paper 3098. Cambridge, Mass.: National Bureau of Economic Research.
- Scott, Maurice FG. 1991. *A New View of Economic Growth: Four Lectures*. World Bank Discussion Paper 131. Washington, D.C.
- Uzawa, Hirofumi. 1965. "Optimum Technical Change in an Aggregative Model of Economic Growth." *International Economic Review* 6(1):18-31.

Toward a Counter-Counterrevolution in Development Theory

Paul Krugman

During the 1940s and 1950s a distinctive set of ideas emerged in development economics that stressed the importance of increasing returns and pecuniary external economies arising from the effects of market size. Unfortunately, the economists who proposed these ideas were at first unable, and later unwilling, to codify them in clear, internally consistent models. At the same time the expected standard of rigor in economic thinking was steadily rising. The result was that development economics as a distinctive field was crowded out of the mainstream of economics. Indeed, the ideas of "high development theory" came to seem not so much wrong as incomprehensible.

This paper argues that in light of new developments in industrial organization, international economics, and growth theory, the old development economics now looks much more sensible than it seemed during the "counterrevolution" against interventionist development models. While development economics has been used to justify some highly destructive economic policies, there is a valid and useful set of core ideas that can be usefully resurrected. Thus this paper calls for a "counter-counterrevolution" that restores some of the distinctive focus that characterized development economics before 1960.

Once upon a time there was a field called development economics—a branch of economics concerned with explaining why some countries are so much poorer than others and prescribing ways for poor countries to become rich. In the field's glory days in the 1950s, the ideas of development economics were regarded as revolutionary and important and commanded both great intellectual prestige and substantial real-world influence. Moreover, development economics attracted creative minds and was marked by a great deal of intellectual excitement.

That field no longer exists. There are, of course, many excellent people who work on the economics of developing countries. Some of the problems they work on are essentially generic to all countries, but there are also issues uniquely characteristic of poorer countries, and in this sense there is a field that focuses on the economics of underdevelopment. But it is a diffuse field: those who work on

Paul Krugman is professor of economics at the Massachusetts Institute of Technology.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

the economics of agriculture in the developing countries, for example, have little (if any) overlap with those who work on developing country trade in manufactures, who in turn hardly talk to those who focus on the macroeconomics of debt and hyperinflation. And very few economists would now presume to offer grand hypotheses about why poor countries are poor or what they can do about it. In effect, a counterrevolution has swept development economics away.

This paper argues that the counterrevolution went too far. I will argue that during the 1950s a central core of ideas emerged regarding external economies, strategic complementarity, and economic development that remains intellectually valid and that may continue to have practical applications. This set of ideas, which I will refer to as "high development theory,"¹ anticipated in a number of ways the cutting edge of modern trade and growth theory. But high development theory was virtually buried, essentially because the founders of development economics failed to make their points with sufficient analytical clarity to communicate their essence to other economists, and perhaps to themselves. Recent changes in economics now make it possible to reconsider what the development theorists said and to regain the valuable ideas that have been lost. In other words, this paper calls for a counter-counterrevolution in development theory.

It is somewhat awkward to present a paper that is to some extent a history of thought at a conference on current research. But development is one of those fields (growth, trade, and regional economics are others) in which recent concepts can lead to a rediscovery of the validity of discarded insights. And by linking old ideas with the new, we may avoid repeating old mistakes and get a chance to make new ones instead.

I. A MOTIVATING EXAMPLE: THE BIG PUSH REVISITED

The Big Push paper of Rosenstein-Rodan (1943) has inspired many interpretations. Some economists read it as essentially Keynesian, a story about interactions between the multiplier and the accelerator. Rosenstein-Rodan himself seems to have had some more or less Keynesian idea about effective demand in mind, with (as we will see) considerable justification. Other economists saw it as an assertion that growth must be somehow "balanced" in order to be successful. Indeed, Hirschman (1958) cast *The Strategy of Economic Development* as an argument with—and refutation of—Rosenstein-Rodan (and others of the balanced-growth school), which I will argue was both a misunderstanding and self-destructive. Yet other economists tried to generate low-level equilibrium traps by invoking such mechanisms as interactions among income, savings, and population growth (see, for example, Leibenstein 1957; Nelson 1956). Such

1. It will become apparent that what I identify as "high development theory" is essentially the nexus among the external economy/balanced growth debate, the concept of linkages, and the surplus labor doctrine. This theory's golden age began with Rosenstein-Rodan (1943) and more or less ended with Hirschman (1958). Obviously this nexus does not cover all of what was happening in the field of development economics even at that time, but it is the core of what I believe needs to be recaptured.

mechanisms can also justify a Big Push, but they are very far from the spirit of the original story.

Fairly recently, however, Murphy, Shleifer, and Vishny (1989) offered a formalization of the Big Push that is quite close to the original spirit and also quite revealing about the essential aspects of high development theory. I offer a slightly streamlined presentation of their model and then turn to what it tells us.

The Big Push Model of Murphy, Shleifer, and Vishny

The model examines an economy that is closed to international trade. The economy can be described by assumptions about factor supply, technology, demand, and market structure. This model can serve as a motivating example to explain both the elements of high development theory and why that theory failed to establish a secure place in the mainstream of economics.

Factor supply. The economy is endowed with only a single factor of production—labor—in fixed total supply L . Labor can be employed in either of two sectors: a “traditional” sector, characterized by constant returns, or a “modern” sector, characterized by increasing returns. Although the same factor of production is used in the traditional and modern sectors, it is not paid the same wage. Labor must be paid a premium to move from traditional to modern employment. We let $w > 1$ be the ratio of the wage rate that must be paid in the modern sector to that in the traditional sector.

Technology. It is assumed that the economy produces N goods, where N is a large number. We choose units so that the productivity of labor in the traditional sector is unity in each of the goods. In the modern sector unit labor requirements are decreasing in the scale of production. For simplicity, decreasing costs take a linear form. Let Q_i be the production of good i in the modern sector. Then if the modern sector produces the good at all, the labor requirement will be assumed to take the form

$$(1) \quad L_i = F + cQ_i$$

where $c < 1$ is the marginal labor requirement. Note that for this example it is assumed that the relationship between input and output is the same for all N goods.

Demand. Demand for the N goods is Cobb-Douglas and symmetric. That is, each good receives a constant share $1/N$ of expenditure. The model will be static, with no asset accumulation or decumulation; expenditure thus equals income.

Market structure. The traditional sector is assumed to be characterized by perfect competition. Thus, for each good there is a perfectly elastic supply from the traditional sector at the marginal cost of production; given our choice of units, this supply price is unity in terms of traditional sector labor. By contrast, a single entrepreneur is assumed to have the unique ability to produce each good in the modern sector.

How will such a producer price? Given the assumption of Cobb-Douglas demand and a large number of goods, she will face unit-elastic demand. If she were an unconstrained monopolist, she would therefore raise her price without limit. But potential competition from the traditional sector puts a limit on the price: she cannot go above a price of 1 (in terms of traditional labor) without being undercut by traditional producers. So each producer in the modern sector will set the same price—unity—as would have been charged in the traditional sector.

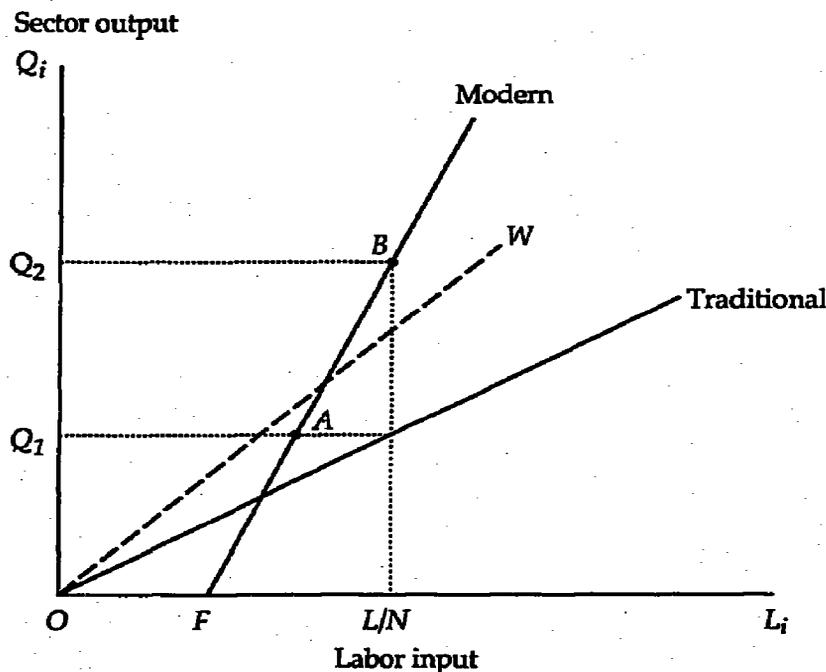
We can now ask the question, will production actually take place in the traditional or the modern sector? To answer this, it is useful to draw a simple diagram (figure 1). On the horizontal axis is the labor input, L_i , used to produce a typical good. On the vertical axis is that sector's output, Q_i . The two solid lines represent the technologies of production in the two sectors: a ray from the origin for the traditional sector, a line with a slope of $1/c$ for the modern sector.

From this figure it is immediately possible to read off what the economy would produce if all labor were allocated to either the modern or the traditional sector. In either case L/N workers would be employed in the production of each good. If all goods are produced traditionally, each good would have an output Q_1 . If they are all produced using modern techniques, the output is Q_2 . As drawn, $Q_2 > Q_1$; this will be the case provided that

$$(2) \quad \frac{(L/N) - F}{c} > L/N$$

that is, as long as the marginal cost advantage of modern production is suffi-

Figure 1. *The Big Push*



ciently large, or fixed costs are not too large, or both. Since this is the interesting case, we focus on it.

But even if the economy could produce more using modern methods, this does not mean that it will. Production must be profitable for each individual entrepreneur in the modern sector, taking into account the necessity of paying the premium wage w , as well as the decisions of all other entrepreneurs.

Suppose that an individual firm starts modern production while all other goods are produced using traditional techniques. The firm will charge the same price as that charged by traditional makers and hence will sell the same amount. Because there are many goods, we may neglect any income effects and suppose that each good continues to sell Q_1 . Thus, this firm would have the production and employment illustrated by point A .

Is this a profitable move? The firm uses less labor than would be required for traditional production but must pay that labor more. Draw in a ray from the origin whose slope is the modern relative wage w ; OW in the figure is an example. Then modern production is profitable given traditional production elsewhere if and only if OW passes *below* A . As drawn, this test is of course failed: it is not profitable for an individual firm to start modern production.

In contrast, suppose that all modern firms start simultaneously. Then each firm will produce Q_2 , leading to production and employment at point B . Again, this will be profitable if the wage line OW passes below B . As drawn, this test is satisfied.

Obviously there are three possible outcomes.² If the wage premium $w - 1$ is low, the economy always "industrializes"; if it is high, it never industrializes; and if it takes on an intermediate value, there are both low- and high-level equilibria.

One would hardly conclude from this model that the existence of multiple equilibria is likely; even given the assumptions, such multiple equilibria will occur only for some parameter values. And it is easy to critique the plausibility of the assumptions. Yet the model can serve as a useful jumping-off point for thinking about development models.

Some Analytical Implications

The Big Push model may be viewed as a minimalist demonstration of the potential role of pecuniary external economies for development, of the necessary conditions for such external economies, and of what a model of external economies must include.

External economies. It is clear that when there are two equilibria in this model, the movement from one to the other involves meaningful external economies. This is true even if one takes the wage premium for the modern sector to represent payment for the disutility of modern life; that is, if one

2. Actually four, if one counts the case in which (2) is not satisfied, so that the economy actually produces less using modern techniques. In this case it clearly stays with the traditional methods.

regards the gain in wages when workers move from traditional to modern jobs as having no welfare significance. Even in that case the industrialized equilibrium leaves workers indifferent while generating profits that would otherwise not exist. If one instead offers some kind of efficiency wage or surplus labor argument that places at least partial value on the rise in wages, the case is that much stronger.

But there are no technological external economies in the model. Why do pecuniary external economies matter here?

Necessary conditions. Two conditions are necessary to generate external economies in this model. First, there must be economies of scale in production. This is obvious from the geometry: if there were no fixed costs in the modern sector, the profitability of modern firms would not depend on how many other firms were using modern techniques.

Second, the modern sector must be able to draw labor out of a traditional sector that pays lower wages. I would like to stretch the point a bit here and think of the essence of the condition as the availability of an elastic supply of labor to the modern sector—labor that would not be employed in equally productive occupations otherwise. (This is what gives the model its vaguely Keynesian feel.) It is thus the interaction between internal economies of scale and elastic factor supplies that gives rise to *de facto* external economies.

Modeling. A final point is crucial. In writing a coherent model of the Big Push, it is necessary to deal with the problem of market structure. As long as there are unexhausted economies of scale in the modern sector, which are crucial to the whole argument, one must face up to the necessity of modeling the modern sector as imperfectly competitive. In the Murphy, Shleifer, and Vishny (1989) formulation, imperfect competition is dealt with by assuming a set of limit-pricing monopolists. This works well here, although, as we will see, it is not always an adequate device. The point is, however, that one must deal with the issue somehow. To the extent that there is anything to high development theory, it is intimately bound up with imperfect competition. If one tries to fudge that issue, as many economists have, one ends up with mush.

There are, unfortunately, no general or even plausible tractable models of imperfect competition. The tractable models always involve some set of arbitrary assumptions about tastes, technology, behavior, or all three. This means that in order to do development theory one must have the courage to be silly, writing down models that are implausible in the details in order to arrive at convincing higher-level insights.

This is not a new lesson. Trade theorists learned it more than a dozen years ago, when they realized that a reconstruction of trade theory to take account of increasing returns would necessarily involve abandoning all pretense of generality; growth theorists learned the same lesson a few years later. High development theory faltered because it did not take the same leap.

II. THE ELEMENTS OF HIGH DEVELOPMENT THEORY

This section offers an interpretive summary of the main elements of high development theory. It is not a literature survey: I base the discussion on only a few authors and must admit that I am imposing more coherence on their views, both across authors and within the work of an individual author, than an unbiased reader is likely to discover on reading their works. So this is, in a way, a statement of what high development economics could or should have been, rather than a portrait of what it was.

Economies of Scale and External Economies

A casual reading of the development literature suggests that there is a dividing line around 1960. Before 1960 writers on development generally assumed that economies of scale were a limiting factor on the ability to establish profitable industries in developing countries and that in the presence of such economies of scale pecuniary external economies assumed real welfare significance. They seem, however, to have been unaware of the degree to which economies of scale raise problems for explicit modeling of competition and of the extent to which the drive for formalism was pushing economics toward explicit models.³ After 1960, by contrast, economists working on development had been trained in the formalism of constant-returns general equilibrium; they did not so much reject the possibility that economies of scale might matter as simply fail to notice it.

The Big Push model presented above is one in which economies of scale at the plant level and an elastic supply of factors of production interact to yield pecuniary external economies with real welfare significance. In retrospect, it is remarkable how clearly similar stories were presented in many papers from the era of high development theory—and also how unaware many of the authors seem to have been of the extent to which their conclusions depended crucially on the non-neoclassical assumption of significant unexploited scale economies.

We begin with Rosenstein-Rodan (1943). In his seminal paper he illustrated his argument for coordinated investment by imagining a country in which 20,000 (!) “unemployed workers . . . are taken from the land and put into a large new shoe factory. They receive wages substantially higher than their previous income *in natura*.” Rosenstein-Rodan then goes on to argue that this investment is likely to be unprofitable in isolation but profitable if accompanied by similar investments in many other industries. Both crucial assumptions are clearly present: the assumption of economies of scale, embodied in the assertion that the factory must be established at such a large scale, and the assumption that these workers can be drawn elastically from among the unemployed or poorly paid agricultural workers.

3. In the field of economic geography, in which some of the tradition of high development theory survives, it appears that the problems raised by economies of scale for market structure are still not appreciated. See, for example, Dicken and Lloyd (1990), an excellent and clear-headed survey that nonetheless blithely ignores the problems of market structure.

Some—although not all—subsequent development writers invoked economies of scale as crucial to external economies. In the best papers the basic story comes through clearly. Fleming (1955) presented an analysis of the nature of external economies in development that focused on the interaction between factor supply and scale economies and that also, unlike Rosenstein-Rodan, points out clearly that the case for coordination falls apart without both assumptions.

Hirschman (1958) is not usually thought of as a thinker preoccupied with non-convexities. Nevertheless, his explanation of the concept of backward linkages explicitly invokes the importance of achieving minimum economic scale, and his discussion of forward linkages alludes to the role of scale as well, although more vaguely. I would argue, then, that a central concept of high development theory circa 1958 was the idea that economies of scale at the level of the individual plant translated into increasing returns at the aggregate level through pecuniary external economies. Admittedly some of the literature of the time does not seem to agree with my argument that scale economies were a key element of the theory. Nurkse (1952), while accepting that indivisibilities play a role in virtuous circles of development, denies that they are essential. Scitovsky (1954), in making the clear distinction between technological and pecuniary external economies, makes the now classic point that in competitive equilibrium it is actually efficient to ignore pecuniary external effects. When he searches for reasons to soften this conclusion, he provides only a single paragraph on scale effects, then turns to an extended discussion of expectational errors. Lewis's (1955) text on economic growth seems fairly innocent of the whole idea of external economies; indeed, the term does not even appear in the index. And Myrdal's (1957) exposition of the role of "circular and cumulative causation" sounds as if it must surely include a major role for economies of scale, but I have been unable to find in his work a single—even indirect—reference to their role. Indeed, when he offers an example of the process of circular causation, the external economies occur through the tax rate rather than through any private market spillover.

So it may be giving too much credit to our intellectual forerunners to think of 1950s development theory as involving a general appreciation of the way economies of scale at the level of the individual plant can aggregate to strategic complementarity at the level of the economy. But at least some theorists seem to have understood the point quite clearly.

Factor Supply

Probably the most famous paper in the literature of development economics is Arthur Lewis's (1954) "Economic Development with Unlimited Supplies of Labor." In retrospect it is hard to see exactly why. One interpretation of Lewis's argument is that the shadow price of labor drawn from the agricultural sector in developing countries is zero—or at least low—so that the social return to investment in industry exceeds its private return. It was pretty obvious even early on, however, that this was a fragile basis for the justification of protection and the promotion of industry.

Why then was Lewis so influential? One reason was probably that the surplus labor story, unlike many of the other development stories emerging at the time, could be formalized with relative ease; thus it gave economists a way to follow the mainstream's increasing emphasis on rigor and formalism while continuing to work on development. But even though Lewis himself made no reference to the external economy/development literature, his defense of the surplus labor idea also helped shore up one of the key ideas of that literature. The assumption that additional labor in the manufacturing sector could come out of rural underemployment was, as already noted, central to Rosenstein-Rodan (1943), and Fleming (1955) observed that in the absence of such an assumption industrial investments would be substitutes instead of complements.

Rosenstein-Rodan and Lewis stressed the elasticity of labor supply as a key factor in development. Other authors, such as Nurkse (1952), stressed the elasticity of capital supply, which also has a potential role in Murphy, Shleifer, and Vishny's (1989) two-period version of the Big Push. Krugman (1990) raises the possibility of multiple equilibria, and Hirschman (1958) emphasized at length the extent to which investment opportunities could elicit additional savings. Again, it may be stretching the point, but many development theorists in the 1950s seem to have been aware that elasticity of factor supply was also crucial to an external economy story of development.

It may be worth pointing out that in regional economics and economic geography, it is entirely natural to assume high elasticity of factor supply to a particular region, since factors of production may be attracted from other regions. This is one reason why the tradition of high development theory remained alive much longer among geographers than among economists; development stories such as those of Pred (1966) continued to seem natural and plausible.

Backward and Forward Linkages

The idea of linkages is one of the greatest sources of confusion in thinking about both the theory of development and development in practice. Hirschman (1958) introduced the term and presented it as something quite new. Later commentators have taken him at his word. Thus Little (1982) insists that since other authors had already explored at some length (if with some confusion) the possible role of pecuniary external economies, Hirschman's linkage concept must have crucially involved a nonpecuniary element. Yet in Hirschman's definition of backward linkages, as already mentioned, the role of pecuniary externalities linked to economies of scale is quite explicit: an industry creates a backward linkage when its demand enables an upstream industry to be established at minimum economic scale. The strength of an industry's backward linkages is to be measured by the probability that it will push other industries over the threshold.

Forward linkages are also defined by Hirschman as involving an interaction between scale and market size; in this case the definition is vaguer, but it seems to involve the ability of an industry to reduce the costs of potential downstream

users of its products and thus, again, push them over the threshold of profitability.

Seen in this way, the concepts of forward and backward linkages seem quite straightforward—and also less distinctive to Hirschman. Fleming (1955), in particular, argued that the “horizontal” external economies of Rosenstein-Rodan were less important than the “vertical” external economies that result when intermediate goods are produced subject to scale economies, which sounds awfully close to linkage theory.

It is also possible to offer simple formal models illustrating the concepts of forward and backward linkages. Indeed, the Murphy, Shleifer, and Vishny Big Push model can be seen as essentially driven by the backward linkages among goods; each good produced in the modern sector enlarges the markets for all other goods.

Forward linkages are a little more difficult to model. They ordinarily arise in the context of industries producing intermediate goods (although not always, as described below); this means that a more complex structure than that presented in the model above is required. Also, the limit-pricing assumption that makes imperfect competition easy in the Big Push model immediately rules out any forward linkage, since cost savings are never passed on to downstream consumers.

There are, however, slightly but not much more difficult models in which both forward and backward linkages do appear. In particular, in models of economic geography (Krugman 1991) it is quite natural to try to model concentrations of population by assuming that factor mobility interacts with economies of scale at the plant level to generate external economies.

In the simplest such model, a geographic concentration of manufacturing is held together for two reasons: firms want to locate close to the large market provided by other firms’ workers (a backward linkage), and workers want to live close to the supply of goods provided by other firms (a forward linkage). The appendix to this paper presents this simplest model (again in a somewhat streamlined form compared with its initial version) and shows how the algebra naturally gives rise to terms that can be identified with forward and backward linkages.

As in the case of Lewis, it is slightly puzzling that Hirschman’s work had such an impact. What he seems to have offered by way of distinct analysis were two hints about development planning. First, the focus on linkages involving intermediate goods rather than final demand suggested that development efforts could focus on a few strategic industries rather than seek an economywide Big Push; hence Hirschman’s view that he was an opponent of Rosenstein-Rodan and Nurkse, even though they were far closer to one another in world view than any of them was to the emerging views of mainstream economics. Second, Hirschman’s discussion seemed to suggest that appropriate critical industries could be identified by examining input-output tables, an exciting suggestion for the quantitatively oriented planner.

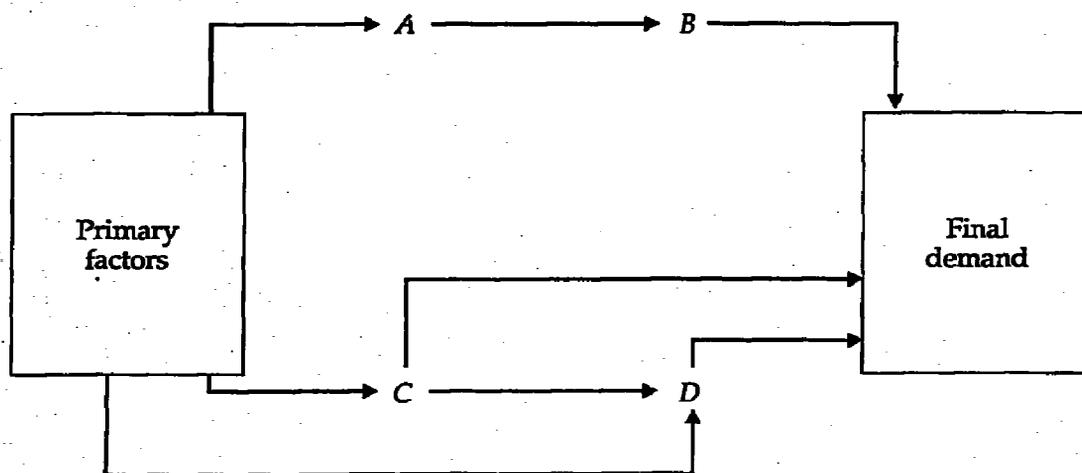
In fact, the concept of linkages, even as Hirschman presented it, implied no such thing. Consider the input-output structure illustrated by the arrows in figure 2. Imagine that there are four industries—*A*, *B*, *C*, and *D*. *A* is a pure intermediate-good industry, and *B* makes no direct use of factors of production. By contrast, *C* sells part of its output directly to final demand, while *D* uses primary inputs as well as inputs from *C*. Any classification using the input-output table will suggest that *A* has stronger forward linkages than *C* and that *B* has stronger backward linkages than *D*. But what if *A* and *B* are characterized by constant returns, while *C* and *D* are characterized by economies of scale? Then it could easily be that there is a coordination problem for *C* and *D*—that neither industry will emerge unless assured of sufficient scale of the other—without any corresponding problem for *A* and *B*. In other words, lots of entries in the input-output table tell the analyst little about which industries might actually play a catalytic role for the economy.

In general, it seems best to regard “linkages” as simply a particularly evocative phrase for the strategic complementarities that arise when individual goods are produced subject to economies of scale. This, in effect, argues that Hirschman’s distinctive contribution was more one of style than of substance, a point to which I will return below.

Summary

I have argued that a number of works in development economics written during the 1950s contained, more or less explicitly and more or less self-consciously, a theory in which strategic complementarity played a key role in development, in which external economies arose from a circular relationship whereby the decision to invest in large-scale production depended on the size of the market, and in which the size of the market depended on the decision to invest. Whatever the practical relevance of this theory, it made perfectly good logical sense.

Figure 2. *Input-Output Linkages among Four Industries*



Yet this development theory was subsequently abandoned to such an extent that classic papers in the field began to seem, as the physicist Wolfgang Pauli used to say, "not even wrong"—simply incomprehensible. We next turn to the reasons for that abandonment.

III. THE DECLINE OF HIGH DEVELOPMENT THEORY

Why did development economics fade away? One can, with some justification, offer the cynical explanation that the field waned with its funding. After all, development economists were most often consulted or given positions of influence in connection with the disbursement of foreign aid. As foreign aid became increasingly unpopular with the electorates of rich nations and as the real value of such aid not only failed to keep pace with gross world product but actually declined, development economics became a much less exciting career. One may also argue that development economics was discredited by a lack of practical success. After all, compared with the hopes of the 1950s and the 1960s, the performance of most developing countries has been dismal. (Indeed, the polite phrase "developing country" itself has become an embarrassment when it must be used in such sentences as "Per capita income in the developing countries of Sub-Saharan Africa has declined steadily since the mid-1970s.") It is unfair to blame Western economists for more than a small fraction of this failure but, as described below, the ideas of development economics were too often used to justify policies that in retrospect impeded rather than encouraged growth. Where rapid economic growth did occur, it happened in ways that were not anticipated by the development theorists.

Yet neither declining external demand for development economists nor their practical failures fully explain the collapse of the field. Purely intellectual problems were also extremely important. During the years when high development theory flourished, the leading practitioners failed to turn their intuitive insights into clear-cut models that could serve as the core of an enduring discipline.

Failures of Formalism

From the point of view of a modern economist, the most striking feature of the works of high development theory is their adherence to a discursive, non-mathematical style. Economics has, of course, become vastly more mathematical over time. Nonetheless, development economics was archaic in style even for its own time. Of the four most famous works, Rosenstein-Rodan's was approximately contemporary with Samuelson's formulation of the Heckscher-Ohlin model, while Lewis, Myrdal, and Hirschman were all roughly contemporary with Solow's initial statement of growth theory.

The problem was not that development economists were peculiarly incapable mathematically. Hirschman made a significant contribution to the formal theory of devaluation in the 1940s, while Fleming helped create the still-influential Mundell-Fleming model of floating exchange rates. Moreover, development practitioners were at the same time generating mathematical planning models—

first Harrod-Domar-style growth models, then linear programming approaches—that were actually quite technically advanced for their time.

So why wasn't high development theory expressed in formal models? Almost certainly for one basic reason: the difficulty of reconciling economies of scale with a competitive market structure.

The examples of the Big Push model and the linkage model show that models in the spirit of high development theory need not be very complicated. They must, however, somehow deal with the problem of market structure. This essentially means making some peculiar assumptions that allow one to exploit the bag of tricks that industrial organization theorists developed for thinking about such issues in the 1970s. In the 1950s, although the technical level of development economists was actually high enough to allow them to do the same thing, the bag of tricks wasn't there. So development theorists were placed in an awkward bind, with essentially sensible ideas that they could not quite express in fully worked-out models. And the drift of the economics profession made the situation worse. In the 1940s—and even in the 1950s—it was still possible for an economist to publish a paper that made persuasive points verbally without tying up all the loose ends. After 1960, however, the response to a paper like Rosenstein-Rodan's would have been: "Why not build a smaller factory (for which the market is adequate)? Oh, you're assuming economies of scale? But that means imperfect competition, and nobody knows how to model that, so this paper doesn't make any sense." It seems safe to say that such a paper would have been unpublishable any time after 1970, if not earlier.

Some development theorists responded by getting as close to a formal model as they could. This is to some extent true of Rosenstein-Rodan, and certainly of Fleming (1955), which gets painfully close to being a full model. But others at least professed to see a less formal, less disciplined approach as a virtue. It is in this light that one needs to see Hirschman and Myrdal. These authors are often cited today (by me among others) as forerunners of the recent emphasis in several fields on strategic complementarity. Their books, however, actually marked the end, not the beginning, of high development theory. Myrdal's central thesis was the idea of "circular causation." But the idea of circular causation is essentially already there in Young (1928), not to mention Rosenstein-Rodan, or Nurkse, who in 1952 referred repeatedly to the circular nature of the problem of getting growth going in poor countries. So Myrdal was, in effect, encapsulating an already extensive and familiar set of ideas rather than a new departure. Similarly, Hirschman's idea of linkages was more distinctive for the effectiveness of the term and the policy advice that he derived loosely from it than for its intellectual novelty; in effect Rosenstein-Rodan was already talking about linkages, and Fleming very explicitly had both forward and backward linkages in his discussion.

What marked Myrdal and Hirschman was not so much the novelty of their ideas but their stylistic and methodological stance. Until their books were published, economists doing high development theory were trying to be good mainstream economists. They could not develop full formal models, but they got as

close as they could to the increasingly model-oriented mainstream. Myrdal and Hirschman abandoned this effort and eventually took stands on principle against any effort to formalize their ideas.

One imagines that this was initially liberating for them and their followers. Yet in the end it was a vain stance. Economic theory is essentially a collection of models. Broad insights that are not expressed in model form may temporarily attract attention and even win converts, but they do not endure unless codified in a reproducible and teachable form. You may not like this tendency; certainly economists tend to be too quick to dismiss what has not been formalized (although I believe that the focus on models is basically right). Like it or not, however, the influence of ideas that have not been embalmed in models soon decays. This was the fate of high development theory. Myrdal's effective presentation of the idea of circular and cumulative causation and Hirschman's evocation of linkages were stimulating and immensely influential in the 1950s and early 1960s. By the 1970s (when I was a student of economics) they had come to seem not so much wrong as meaningless. What were these guys talking about? Where were the models? High development theory was not so much rejected as simply bypassed.

The exception proves the rule. Lewis's surplus labor concept was the model that launched a thousand papers—even though surplus labor assumptions were already standard among development theorists, the empirical basis for assuming surplus labor was weak, and the idea of external economies/strategic complementarity is surely more interesting. The point was, of course, that precisely because he did not mix economies of scale into his framework, Lewis offered theorists something they could model using available tools.

Practical Failures

Development theorists were unable to formulate their ideas with the precision required by an increasingly model-oriented economic mainstream and were thus left behind. Although I believe this to be the main explanation of what went wrong, it is also true that practical failures and empirical evidence had something to do with the decline of development economics.

Little (1982) has pointed out that development theorists in the 1950s were by and large not optimists. Indeed, they were generally more pessimistic than turned out to be warranted (except in Africa). But their pessimism was largely based on doubts about the ability of countries to carry out the coordinated effort that they regarded as central to industrialization.

What actually happened was that most developing countries were quite successful at developing industrial bases, but since these industries were highly inefficient, industrialization turned out not to have much to do with development. This called into question the whole idea that the problem of coordinating investments in the face of external economies was a major part of the underdevelopment story.

The successes of development have also been an embarrassment for high development theory. Most versions of that theory tacitly assumed that firms would

produce only for the domestic market, a view that seemed sensible given the trade pessimism of the time. The great success stories have, however, involved export-oriented industrialization.

Another major embarrassment for development has been the realization of the extent to which the rhetoric of development theory has been used to cover poorly conceived or even corrupt policies. This is a familiar subject, acerbically surveyed in Little (1982). The calculation of effective rates of protection, whatever the method's flaws, revealed levels that were often absurdly high, some cases of negative value added at world prices, and highly variable rates of protection across industries that were difficult to justify. Country studies of trade policy revealed a heavy preference for complex administrative regulations that were evidently more costly than tariffs and, moreover, failed to yield revenue (Little, Scitovsky, and Scott 1970). Studies of repressed financial systems showed similar irrationalities (see, for example, McKinnon 1973). And it became apparent that the incentives provided by administratively generated rents were becoming major objectives of both legal and illegal economic activity (Krueger 1974).

These observations of bad policies based on old development economics provided a key argument in what amounted to the formation of a new orthodoxy. Little (1982) has argued that in the 1950s what he calls a "structuralist" theory of economic development was in effect an orthodoxy. I find it hard to link the ideas of high development theory as described above with his definition of the structuralist view, which is that developing country markets are rigid and need to be pushed into action by government. But it is certainly true that in the 1950s market failures were seen as pervasive, and the case for intervention was taken to be not so much an empirical observation as an obviously true conclusion from obviously true theory. By about 1980 a belief in the efficacy of free trade and free markets for developing countries had similarly taken hold, its intellectual credibility underpinned by the demonstration of market efficiency in neoclassical general equilibrium theory. This orthodoxy also effectively denies that there is anything special about the situation of developing countries compared with those of richer nations. The poor are no different from you and me—they just make less money.

What was ironic was that a competitive neoclassical orthodoxy settled in on the development front just as that orthodoxy was breaking up in other fields.

IV. RECENT THEORETICAL DEVELOPMENTS AND HIGH DEVELOPMENT THEORY

We can now see that whatever bad policies may have been implemented in the name of high development theory, the theory itself makes quite a lot of sense. Indeed, in some ways it was a remarkable anticipation of ideas that would come to analytical fruition thirty years later in the fields, for example, of international trade and economic growth. This section reviews these developments and asks where we go from here: how do we recapture the insights of development economics, and what good might they do for policy formulation?

International Trade

One of the main empirical criticisms of models of high development theory is their emphasis on domestic rather than international markets and their neglect of international trade. So it is somewhat ironic that the initial rehabilitation of the idea that economies of scale, external economies, and strategic complementarity are important outside the field of industrial organization came from the theory of international trade.

The literature on the "new trade theory" is much too extensive to cite. It is, however, possible to summarize the main lines of thought. The initial models, many of which are summarized in Helpman and Krugman (1985), invoke economies of scale at the level of individual firms to explain national specialization in individual products and hence the observed large volumes of intra-industry trade. In these initial models, economies of scale at the firm level did not aggregate to yield any kind of strategic complementarity, and thus there were few parallels with what I call high development theory.

In subsequent developments, however, many of the themes of high development theory have in effect reentered. Trade theorists noted (Ethier 1982) that increasing returns in the production of intermediate goods generate external economies at the level of final goods. If intermediate goods are nontradable, these de facto external economies are country-specific, and one easily generates examples of multiple equilibria (Helpman and Krugman 1985, ch. 11).

Nor need one assume nontradability. Models in which there are nonprohibitive transport costs can also yield external economies; the model in the appendix to this paper is an example. And indeed in some such models the likelihood and welfare significance of multiple equilibria can actually increase when transport costs fall and trade increases (Krugman 1991, ch. 3). It is often argued that Big Push stories could only be valid in an autarkic economy. What the models suggest is that as long as some things (perhaps only factors of production) are costly to trade, the possibility of country-specific external economies remains.

But are such external economies of real practical importance for international trade? Systematic empirical testing is difficult. At least among business observers, however, external economy stories of international specialization are widely accepted. Porter's (1990) analysis of international competition can be viewed in large part as a return to traditional development themes in the context of an advanced country.

In the academic field of international trade, recent work has increasingly shifted to the analysis of long-run growth (Grossman and Helpman 1991). This move in trade is part of the broader movement known as "new growth theory."

Growth

The new growth theory was created by Romer (1986) and Lucas (1988). It attempts to get away from the conventional Solow result that most long-term

per capita growth arises from exogenous technical progress, through three related hypotheses: (1) social returns to investment are higher than private returns because of external economies; (2) capital broadly defined is a much larger share of input than conventionally measured; and (3) technical progress is largely an endogenous result of market-driven research and development.

Of these three hypotheses, the first is clearly in the same spirit as that of high development theory. In Romer's initial formulation, external economies were treated as purely technological in form. (Fashions in plausibility change: Scitovsky did not hesitate to ridicule Marshall's "bucolic" examples and proclaim the "scarcity of pure technological external economies," and indeed most 1950s theorists seem to have regarded pecuniary externalities as the only interesting case.) In later papers, however, Romer imported techniques from the trade literature to show how growth driven by external economies could arise from economies of scale at the plant.

Given the immense interest in new growth theory, is there any point in revisiting old development theory? Or to put it another way, even if modern theorists have reinvented the wheel, didn't they get it better this time around?

I think the answer is no. New growth theory has been preoccupied with a different question than high development theory: how to explain the persistence of growth rather than how to get it started. And it is notable that new growth models tend to assume that the economy has only one sector, or that all sectors are symmetric. By contrast, high development theory had a core preoccupation with the difference between modern sectors that were presumed to be characterized by economies of scale and traditional sectors that were not. Even within the modern sector, the linkage concept implied a search for key industries. So while the philosophy of new growth theory is in essence a rediscovery of high development theory, it has not returned to the same questions.

The most essential difference, of course, is that despite its abstract approach, high development theory was intended as a guide to policy, while new growth theory was not. The latter theorists only try to explain the world, while their predecessors thought they could change it. The point is that if the underlying economic assumptions of high development theory have once again been legitimized, one can ask the same questions again.

Policy Implications

It is not my intention to offer a clarion call for interventionist trade and industrial policies, much less for a return to import substitution. Indeed, a proper interpretation of the insights from old development theory (and new trade and growth theories) suggests that the costs of fragmenting markets are high precisely because markets are not characterized by constant returns and perfect competition.

Instead, I want to offer two more modest suggestions. First, this survey of intellectual developments may serve as a caution against carrying a free-market

orthodoxy too far. It makes considerable sense for the World Bank and other multilateral agencies to push very hard for liberal policies in developing countries, given the demonstrated tendencies of these clients to engage in economically irrational interventions. But in the back of our minds we should remember that it is not true that economic theory "proves" that free markets are always best: there is an intellectually solid case for some government promotion of industry—one that has often seemed empirically plausible to sophisticated observers. In other words, don't get caught up too much in the orthodoxy of the moment.

It is worth recalling that in the course of a generation, the ideas of high development theory went from being regarded as self-evident to being regarded as logically meaningless. They were actually ideas that were sensible and coherent, but they were of more limited application than their creators imagined. The lesson here is that both casual empiricism and intellectual narrow-mindedness will lead one badly astray.

The second suggestion is that we need a reorientation of research. Research on trade and industrial policy in developing countries is still dominated by the agenda of the counterrevolution that began in the 1960s: the horrors of import substitution and the distortions imposed by government policy. This is valuable work, and governments continue to give it opportunities to be useful. Yet it is probably time once again to focus on market as well as government failures.

APPENDIX. A SIMPLE MODEL OF FORWARD AND BACKWARD LINKAGES

The Murphy, Shleifer, and Vishny (1989) Big Push model essentially captures, in a minimalist way, the idea of backward linkages. The way it solves the problem of oligopoly pricing, however (by limit pricing to the level set by a traditional, constant-returns technology), rules out any forward linkages. The purpose of this appendix is to show in a minimalist way how a somewhat different formulation can produce both backward and forward linkages.

The example uses regional concentration rather than the transition from traditional to modern production. This is largely because in Krugman (1991) I have already worked out this model (although this is a streamlined exposition). One may also argue, however, that the regional version is more plausible than the application to the whole transition to modernity. Indeed, Myrdal (1957) chose to begin his discussion with the case of relative regional development within countries rather than international disparities, presumably because he too regarded that case as easier to sell. High development theory lived on in the work of regional scientists such as Pred (1966) well after it had declined to obscurity in mainstream economics. And as a practical matter, regional disparities within developing countries are significant issues. The main point of this appendix is, however, to provide an illustrative model on which to hang the discussion in the text.

Assumptions of the Model

We consider a country that has two regions, East and West, and produces two kinds of goods, agricultural and manufactured. Agricultural production is homogeneous, produced under constant returns and perfect competition. Manufactures consist of a number of differentiated products, each produced subject to economies of scale, with a monopolistically competitive market structure.

Everyone in the economy is assumed to share the same tastes. Welfare is a Cobb-Douglas function of consumption of agricultural goods and a manufactures aggregate:

$$(A.1) \quad U = C_M^\pi C_A^{1-\pi}.$$

Note that given this functional form, π is the share of expenditure that falls on manufactures.

The manufactures aggregate is in turn a CES function of consumption of individual manufactured goods, of which there is a large number, not all of them actually produced:

$$(A.2) \quad C_M = \left[\sum_i c_i^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}.$$

As long as a large number of manufactured goods is produced, this functional form ensures that the elasticity of demand for any individual good is simply σ .

There are two factors of production, each of which is specific to a particular sector. "Farmers" produce agricultural goods and "workers" produce manufactured goods; farmers cannot become workers or vice versa. To save notation we choose units so that there is a total of $1 - \pi$ farmers and π workers (this choice of units leads to the result that the wages of farmers and workers are equal in equilibrium).

The geographic distribution of farmers is taken as fixed, with $(1 - \pi)/2$ farmers in each region. Workers move to whichever region offers them a higher real income.

Farmers produce their goods under constant returns to scale. The economies of scale in manufacturing take the form of a linear cost function, in which a fixed cost in manufacturing labor must be incurred in order to produce any individual variety of manufactures:

$$(A.3) \quad L_{Mi} = \alpha + \beta x_i.$$

Finally, we assume that there are costs of transporting manufactured goods between the two regions. These take Samuelson's "iceberg" form, in which only a fraction of a good that is shipped arrives (so that transport costs are incurred in the good shipped). We let $\tau < 1$ be the fraction of a manufactured good shipped that actually arrives. Transport of agricultural goods is assumed to be costless, an assumption made for analytical convenience; it ensures that the wage rates of farmers and the prices of agricultural goods are the same in the two regions.

Pricing and Competition

Since there is a large number of potential manufactured goods, each of them produced subject to economies of scale, there is no reason for any two firms to try to produce the same good; the market structure of manufactures will therefore be one of monopolistic competition.

The producer of any one good will face an elasticity of demand σ . Her profit-maximizing price is therefore a constant markup over marginal cost,

$$(A.4) \quad p_i = \frac{\sigma}{\sigma - 1} \beta w$$

where w is the wage rate of manufacturing workers.

If there is free entry, however, profits will be driven to zero. The zero-profit condition may be written

$$(A.5) \quad (p - \beta w)x = \alpha w.$$

Note that with zero profits, price equals average cost. But this means that the ratio of average cost to marginal cost—which is one measure of economies of scale—is simply $\sigma/(\sigma - 1)$. Thus equilibrium economies of scale are a function only of σ , so that σ , even though it is a parameter of tastes rather than technology, nonetheless acts as a sort of inverse index of the importance of increasing returns.

The zero-profit and pricing conditions together imply that the output of a representative manufacturing firm is

$$(A.6) \quad x = \frac{\alpha(\sigma - 1)}{\beta}.$$

Consider a region with a resident labor force of L_M workers; the number of manufactured goods that region will produce is

$$(A.7) \quad n = \frac{L_M}{\alpha + \beta x} = \frac{L_M}{\alpha \sigma}.$$

Sustainability of a Core-Periphery Pattern

We now ask the following question: is a situation in which all manufacturing is concentrated in one region, leaving the other region with only agriculture, an equilibrium? Since it doesn't matter which region we choose, we examine the sustainability of an equilibrium with East as the manufacturing core and West as the agricultural periphery.

As we will see in a moment, there are two "centripetal" forces tending to keep a manufacturing core in existence and one "centrifugal" force tending to pull it apart. Holding the core together are the desire of firms to locate close to the larger market and the desire of workers to have access to the goods produced by other workers. It is these two forces that may be thought of as corresponding, respectively, to backward and forward linkages.

These forces can only be seen analytically by examining their tension against the third force, which tends to break the core apart: the incentive of firms to move out to serve the peripheral agricultural market. What we will do is derive a criterion that determines whether the backward and forward linkages are strong enough to sustain an established core.

We begin by noting that given our choice of units, the wage rates of workers and farmers will be equal. That is, a share of expenditure π is spent on manufactured goods (including those goods that "melt" in transit) and (since profits are zero) ends up as wages of workers; but we have also chosen units so that a fraction π of the population consists of workers. Given this choice of unit, the wage rates will necessarily be equal.

Now ask how the incomes of the two regions compare. East has half of the farmers, who receive a share $(1 - \pi)/2$ of total income, plus all of the workers, who receive a share π . Let total income be unity; then the income of East is

$$(A.8) \quad Y^E = (1 + \pi)/2.$$

West has only its immobile farmers, who receive a share $(1 - \pi)/2$ of income; so the income of West is

$$(A.9) \quad Y^W = (1 - \pi)/2.$$

This situation, in which all manufacturing is concentrated in East, will be sustainable if it is unprofitable for any firms to enter in West. So we must determine whether it is profitable for an individual firm to "defect" by commencing production in West.

The brief answer is that it will be profitable to defect if a firm that moves to West can afford to pay workers a higher real wage than they are currently receiving in East.

Let n be the (large) number of firms currently producing in East. Then the sales of each of these firms will be

$$(A.10) \quad s^E = \pi/n.$$

If a firm were to try to start production in West, it would need to attract workers. Suppose that to do this it must pay a wage that is w times the wage that East's firms are paying. Then, because the profit-maximizing price is a constant markup over marginal labor cost, such a firm will charge an f.o.b. price that is also w times as high as that of East's firms.

To find the sales of the defecting firm, we note that in West's market, the c.i.f. price of East's good will be $1/\tau$ times its f.o.b. price. Since the elasticity of substitution is σ , this means that in West's market, the value of the sales of the firm will be $(w\tau)^{-(\sigma-1)}$ times the sales of a representative Eastern firm. By similar reasoning, in East's market the sales ratio will be $(w/\tau)^{-(\sigma-1)}$. Thus the relative sales of a Western firm will be

$$(A.11) \quad S_W/S_E = S = \frac{1}{2} w^{-(\sigma-1)} [(1 + \pi) \tau^{\sigma-1} + (1 - \pi) \tau^{1-\sigma}].$$

How high a wage can such a firm afford to pay? It is easiest to think of the firm as using an operating surplus to cover fixed costs. The operating surplus,

in the Dixit-Stiglitz model, is proportional to sales. The fixed cost is incurred in labor, so the maximum relative wage that the firm can afford to pay is defined by

$$(A.12) \quad S = w.$$

This gives us the wage equation

$$(A.13) \quad w = \left[\frac{1 + \pi}{2} \tau^{\sigma-1} + \frac{1 - \pi}{2} \tau^{1-\sigma} \right]^{1/\sigma}.$$

The final step is to ask whether this maximum nominal wage implies a real wage higher or lower than the real wage in East. Since this is a lone defecting firm, all manufactured goods (except for its own negligible contribution) would have to be imported. Recall that only a fraction τ of a good that is shipped arrives. The price of manufactured goods in West will therefore be $1/\tau$ times as high as that in East. The overall price index, which is a geometric average of manufactures and agricultural goods, will thus be $\tau^{-\pi}$ times as high. The maximum relative real wage a defecting firm can pay is thus

$$(A.14) \quad \omega = w\tau^\pi = \tau^\pi \left[\frac{1 + \pi}{2} \tau^{\sigma-1} + \frac{1 - \pi}{2} \tau^{1-\sigma} \right]^{1/\sigma}.$$

A concentration of all manufacturing in East is an equilibrium only if this relative real wage is less than one (that is, a defecting firm could not profitably induce labor to move). We can immediately note that if manufacturing were a small part of the economy (π close to zero), concentrated manufacturing would never be an equilibrium. Equation A.14 would reduce to

$$(A.15) \quad \omega = \left[\frac{1}{2} \tau^{\sigma-1} + \frac{1}{2} \tau^{1-\sigma} \right]^{1/\sigma}$$

which is always greater than one because of Jensen's inequality. This result arises because of the attraction of dispersing production toward the dispersed rural market.

A comparison of (A.14) and (A.15) reveals immediately the potential role of a large manufacturing sector to generate linkages that lead to concentration. In the case with large π , the first term becomes less than one and is smaller the larger is π . This first term represents the attraction to workers of locating close to the existing manufacturing concentration to have access to the goods it produces; in effect it measures a forward linkage. At the same time, a large π shifts the weights on the average inside the brackets from the second term, which exceeds one, toward the first term, which is less than one, and therefore reduces the size of this second term as well. This effect occurs because the larger the manufacturing sector, the larger the market in the region that gets the manufacturing. That is, this effect arises because of the attraction of producing near the market provided by other firms—in effect a backward linkage.

If the forward and backward linkages are strong enough, which they will be if π is large enough, there will be enough strategic complementarity to lead to concen-

trated manufacturing. Krugman (1991) shows that this depends on the degree of economies of scale; concentration is also more likely when scale economies are stronger.

Finally, we may note that if concentration in East is an equilibrium, so is concentration in West, because the regions are symmetric. Thus strong forward and backward linkages imply multiple equilibria.

This model is, of course, very simplified compared with the ideas of linkage that were in the minds of most development theorists. In particular, there are no intermediate goods and thus no input-output structure. But it does show in the simplest fashion that these concepts do at least make sense.

REFERENCES

- Dicken, Peter, and Peter Lloyd. 1990. *Location in Space*. London: Harper and Row.
- Ethier, Wilfred. 1982. "National and International Returns to Scale in the Modern Theory of International Trade." *American Economic Review* 72:389-405.
- Fleming, J. M. 1955. "External Economies and the Doctrine of Balanced Growth." *Economic Journal* 65 (June):241-56.
- Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press.
- Helpman, Elhanan, and Paul Krugman. 1985. *Market Structure and Foreign Trade*. Cambridge, Mass.: MIT Press.
- Hirschman, Albert O. 1958. *The Strategy of Economic Development*. New Haven, Conn.: Yale University Press.
- Krueger, Anne. 1974. "The Political Economy of the Rent-Seeking Society." *American Economic Review* 64(3):291-303.
- Krugman, Paul. 1990. "Endogenous Innovation, International Trade, and Growth." In Paul Krugman, ed., *Rethinking International Trade*. Cambridge, Mass.: MIT Press.
- . 1991. *Geography and Trade*. Cambridge, Mass.: MIT Press.
- Leibenstein, Harvey. 1957. *Economic Backwardness and Economic Growth*. New York: Wiley.
- Lewis, W. A. 1954. "Economic Development with Unlimited Supplies of Labor." *Manchester School of Economic and Social Studies* 22 (2):139-91.
- . 1955. *The Theory of Economic Growth*. London: Allen and Unwin.
- Little, I. M. D. 1982. *Economic Development*. New York: 20th Century Fund.
- Little, I. M. D., Tibor Scitovsky, and Maurice Scott. 1970. *Industry and Trade in Some Developing Countries*. Oxford, U.K.: Oxford University Press.
- Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22:2-42.
- McKinnon, Ronald. 1973. *Money and Capital in Economic Development*. Washington, D.C.: Brookings Institution.
- Murphy, Kevin M., Andrei Shleifer, and Robert Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97 (October):1003-26.
- Myrdal, Gunnar. 1957. *Economic Theory and Under-developed Regions*. London: Duckworth.
- Nelson, R. R. 1956. "A Theory of the Low-Level Equilibrium Trap in Underdeveloped Economies." *American Economic Review* 46 (December):894-908.

- Nurkse, Ragnar. 1952. "Some International Aspects of the Problem of Economic Development." *American Economic Review* 42 (May):571-82.
- Porter, Michael. 1990. *The Competitive Advantage of Nations*. New York: Free Press.
- Pred, Allan. 1966. *The Spatial Dynamics of US Urban-Industrial Growth*. Cambridge, Mass.: MIT Press.
- Romer, Paul. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94:1002-37.
- Rosenstein-Rodan, Paul N. 1943. "Problems of Industrialization of Eastern and South-Eastern Europe." *Economic Journal* 53 (June-September):202-11.
- Scitovsky, Tibor. 1954. "Two Concepts of External Economies." *Journal of Political Economy* 62 (April):143-57.
- Young, Allyn. 1928. "Increasing Returns and Economic Progress." *Economic Journal* 38 (December):527-42.

COMMENT ON "TOWARD A COUNTER-COUNTERREVOLUTION
IN DEVELOPMENT THEORY," BY KRUGMAN

Joseph E. Stiglitz

I wholeheartedly agree with the main points that Professor Krugman raises so eloquently. I would like to raise two related issues. First, I do not think Krugman's interpretation of the intellectual history of development economics is quite right. Second, although Krugman has identified two factors that represent important critiques of the neoclassical paradigm and form the basis for the construction of a "new view," his vision is too narrow: there are equally important factors that he has ignored. In brief, Krugman argues that:

- High development theory left the mainstream of economics.
- The reason for this was that "development theorists were unable to formulate their ideas with the precision required by an increasingly model-oriented economic mainstream, and were thus left behind."
- Attention was diverted by ideas like Lewis's (1955) surplus labor model that could be easily formalized.
- Real-world events, such as the failure of industrialization, "called into question [the idea that] coordinating investments in the face of external economies was a major part of the underdevelopment story."
- The resurrection of high development theory can be attributed to the development of simple models of increasing returns.

Each of these propositions is debatable. To take the first, whether an idea is or is not in the mainstream depends on what river you are sitting beside. The mainstream looks quite different depending on whether one is viewing it from the banks of the Charles (that is, from the Massachusetts Institute of Technology), the Cam (Cambridge), or the Cherwell (Oxford), let alone from the shores of Lake Lagunita (Stanford University). At these institutions—and others—scholars never stopped talking about the importance of externalities, returns to scale, imperfect competition, and technological change and the relationships among them. Research continued on modeling not only the endogeneity of market structure but also "endogenous growth," with theoretical and empirical

Joseph E. Stiglitz is professor of economics at Stanford University. He is indebted to Joshua Gans, Mark Gersovitz, and Andrés Rodríguez for helpful comments on the issues raised here.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

work aimed at understanding the determinants of the transfer, absorption, development, and adaptation of new technologies. These ideas were and continue to be a major focus of academic research and a standard part of the graduate curriculum.

At Cambridge, for instance, throughout the 1950s, 1960s, and 1970s Kaldor (1970, 1972) emphasized three of the elements that Krugman stresses—increasing returns, imperfect competition, and technological change. With Mirrlees (Kaldor and Mirrlees 1969), he provided a formal model of growth theory that captured some of his ideas. Kaldor recognized the profound policy implications of these ideas, and they provided the theoretical foundations for the selective employment tax enacted by the Labour government during his tenure as economic adviser.¹

At Stanford, Arrow (1962) developed one of the central versions of what would later be called a model of endogenous growth. Uzawa (1963, 1965) developed another, at Stanford and Chicago.² Many other formal models were constructed and published, including the well-known paper by Inada (1969), which actually used the word “endogenous.”

To be sure, we were not satisfied with the models offered. The results, particularly those pertaining to steady states, were highly sensitive to the special parameterizations, and one of the objectives of the research program was to explore these sensitivity issues. (Inada 1969 illustrates this line of analysis.) And we were aware that with increasing returns, markets would be imperfectly competitive, and we needed to model those imperfections. Krugman is right in identifying the advances in the theory of imperfect competition of the 1970s as providing a crucial building block. But he fails to mention the other problem, raising the interesting question of the extent to which progress can be attributed to a lowering of standards—a willingness to work with special (should I say ad hoc?) consequential parameterizations, which generated results that were not robust.

The 1970s and 1980s were marked by advances in the modeling of externalities, technological progress, and returns to scale. Major strands of research on evolutionary modeling were associated with Nelson and Winter (1982) and Dosi and others (1988); the analysis of network externalities was undertaken by David (1987) and Arthur (1985, 1988, 1989) and the work on the microeconomics of technological progress by Dasgupta and Stiglitz (1980a, 1980b) and Stiglitz (1988). Aoki (1970) formalized the concept of Marshallian externalities, and Greenwald and Stiglitz (1986) provided a general framework for the analysis of externalities. Although they focused on incomplete markets and

1. Kaldor (1970) explores the implications of these ideas in the context of regional development. See also Kaldor (1972).

2. Modeling “endogenous” technical change was a major thrust of research in this period. See, for instance, the collection of essays by Shell (1967); Atkinson and Stiglitz (1969); Bardhan (1970); and Teubal (1967).

imperfect information, their framework was equally applicable to economies with tax distortions and imperfect competition. They showed that what might be thought of as pecuniary externalities essentially always mattered, as long as the economy was not (constrained) Pareto efficient, and that in these circumstances the economy was essentially *never* constrained Pareto efficient.

Indeed, not only did Krugman ignore major strands of theoretical work; he also ignored major empirical research projects that were exploring some of the central issues of high development economics, such as the Economic Commission for Latin America (ECLA) program under Jorge Katz (see Katz 1987).

Not only was research on these ideas under way, but policies were also informed by these perspectives. I have already referred to the selective employment tax in Great Britain. Certainly current writings on the policies pursued in Japan (see Komiya, Okuno, and Suzumura 1988), the Republic of Korea (see Amsden 1989; Pack and Westphal 1986), and Taiwan (China) (see Wade 1990) suggest that these economic theories were an important part of the intellectual background for those programs.

In short, in my reading of intellectual history, high development economic theory never died; it was alive and well, and the rest of the world may have taken little note of its absence on the banks of the Charles.

I would like to agree with Krugman concerning the importance of theory and models for shaping the direction of the profession. Yet I remain unpersuaded of the dominant role assigned by Krugman, for several reasons.

The first is perhaps a normative rather than a positive argument: that we can write down a model of a phenomenon proves almost nothing. It does not make the idea right or wrong, important or unimportant. It is—at most—a test of certain logical relations, of the consistency of certain ideas. Formalizing ideas is extremely important for quite another set of reasons: it leads to better and more concise debates and to precise and more useful questions!

Second, there were formal models available. Many of us had published models with all the characteristics that Krugman would like—simplicity, elegance, and rigor. The lack of such models simply cannot account for the temporary demise of high development theory—if that had happened.

Conversely, had Rosenstein-Rodan (1943) succeeded in formalizing his ideas, I doubt that those ideas would have been any more palatable. In his model the income effects associated with increasing returns leave the economy stuck in a low-level equilibrium. As Krugman points out, the problem arises from a lack of demand, but once we open the economy to international trade, this argument loses its force.³

I also take issue with Krugman's contention that what accounts for the central role of surplus labor in the 1970s is not the importance of surplus labor but the

3. There are contexts in which a more subtle version of the argument might be relevant: income effects are obviously important for nontraded goods, and there may be spillovers between the returns to scale for nontraded intermediate goods used to produce traded and nontraded final goods.

ease of modeling it. The model was successful because it described central aspects of the development process, including the reallocation of labor from the low-productivity rural sector to the high-productivity urban sector and the high rates of capital accumulation that were facilitated by low wages. These are still important aspects of the development process, although they are far from the whole story.⁴

I would submit that a far more plausible explanation for the seeming demise of high development theory is that the same currents that led to the dominance of free market ideology in the United Kingdom and the United States were reflected—at least in the United States—in the dominance of those ideas in certain intellectual circles. In short, it was as much the market demand for ideas as the supply of models that was crucial.

Krugman is correct in his contention that real world events, such as the failure of the planning paradigm, reinforced these currents, but they do not fully account for them. I say this for two reasons. First, the critique of the neoclassical paradigm was far broader than its omission of increasing returns and externalities. In the 1970s we realized not only that the informational assumptions that underlay that model were implausible but that *all* the results of the model were highly sensitive to these assumptions (see, for instance, Stiglitz 1985). But a careful analysis of the implications of imperfect and costly information provided a critique of both the free market and the planning paradigms (see Stiglitz 1992). Krugman seems to suggest that once the planning paradigm was rejected, the only alternative was the free market paradigm. There were alternatives available, and to explain which alternatives the profession focused on, one has to look elsewhere.

Second, not only is Krugman's view of the intellectual alternative incorrect; his analysis ignores the debates about the success of the East Asian economies, which was based, according to some interpretations, on selective government intervention, consistent with the new insights of microeconomic analysis (see Amsden 1989; Komiya, Okuno, and Suzumura 1988). Amsden cites Kaldor (1970), while Itoh and others (1991) cite papers from the 1970s and early 1980s, well before the formal models that Krugman would like to credit with the resurgence of high development economics were written.⁵

4. Krugman seems to be unaware of the work that originally established Lewis's reputation as an economist (see, for example, Lewis 1949): the importance of overheads (nonconvexities and increasing returns), which he stressed throughout the 1970s and 1980s in courses on development economics at Princeton.

5. This is not the only evidence that intellectual developments outside economics help us understand the dominant ideas in economics. How else could we account for the prevailing fashion of the time: the emphasis on models assuming full employment? Surely memories are not so short as to relegate the Great Depression to ancient history. Were economists so confident about the new era that the economic downturns in 1982 and 1991, accompanied by rising unemployment, came as a total surprise? What about the persistent unemployment in Europe in the 1980s? Here was an area in which simple models with alternative explanations were available.

THE VISION

Krugman takes far too narrow a view of the development process and of what is wrong with both the standard neoclassical and the planning paradigms. I have already illustrated one limitation of his vision: If the central problems were those of externalities and increasing returns, the planning process would have been an appropriate remedy. But that assumption ignored information problems, which are now recognized to be central. Evidently, governments are not well equipped to identify projects and motivate project managers. But these were not the issues on which the planning mechanism focused, and, not surprisingly, it did not resolve them.

Financial Institutions

Indeed, the question of who gets funding and how it is used is the essential problem addressed by financial institutions in capitalist economies. They provide the institutional "solution" to the information problem. How, when, and whether they work is certainly part of the development story. Recent research in macroeconomics has emphasized the markedly different consequences of debt and equity for risk; it has identified failures in both aspects of the capital market (the presence of debt and equity rationing). There is here another link between an elastic labor supply and economic growth. Earlier literature emphasized the importance of capital accumulation; the new literature emphasizes the form in which capital is accumulated—equity versus debt (see Greenwald, Kohn, and Stiglitz 1990). Equity is viewed as being more powerful. Low wages result in high profits and the accumulation of *equity* capital, thus facilitating the growth process. Krugman's failure to mention the importance of these institutions in the growth process is perhaps the best example of what I mean when I say that a broader vision is required.

Political Economy

In interpreting the general problem of government interventions to correct market failures, Krugman refers to problems of political economy. To be sure, these problems are important. But his analysis of the issues is both incomplete and misleading. As noted earlier, political economy problems are not the only source of the failure of the planning paradigm. Moreover, rent-seeking behavior is, at the very least, an incomplete explanation for the failure of public sector enterprises. Krugman fails to note the existence—let alone the importance—of rent-seeking in modern managerial capitalism (see Shleifer and Vishny 1989; Edlin and Stiglitz 1992). And finally, ascribing to political problems the failure to develop does not explain the differences in regional development that have characterized virtually all countries at various stages of their growth. (See Greenwald, Levinson, and Stiglitz 1992 for a discussion of how localized knowledge of capital markets can explain patterns of regional development.) Nor can the allusion to political economy problems explain the many successful govern-

ment enterprises. They may represent a minority of all such enterprises, but there are enough successes to make it plausible that success is not just a matter of luck.

Externalities and Increasing Returns

There is no single explanation of why countries grow or fail to grow. Increasing returns, externalities, and learning by doing may be—and undoubtedly are—important, but modeling them in a way that provides insights into the development process requires more care than has typically been taken, and many of the models formulated to date simply miss the essential issues.

Consider, for instance, the modern rendition of the Big Push argument, at least as interpreted by Murphy, Shleifer, and Vishny (1989). I have already suggested that those arguments, based on income effects, have dubious plausibility (in their present formulations) when applied to economies that face trading opportunities.⁶

Or consider the argument originally modeled by Aoki (1970) and incorporated in Romer's (1986) growth model—that we can reconcile learning by doing with competitive behavior when learning is external to the firm (and internal to the country). If the spillover to other firms is less than 100 percent (and it is hard to believe that those outside the firm learn everything) any time there is learning by doing, competition will be imperfect (see Dasgupta and Stiglitz 1988).

Or take the argument that what is important are “aggregate increasing returns.” That suggests that large economies have a distinct advantage over small economies; it does not explain how a small economy could grow into a big economy. The essential problem—from both an analytic and a policy perspective—is to identify the nature of the externalities that are not internalized by markets and the sources of the returns to scale.

Coase (1960) went too far when he (or his disciples) asserted that all externalities could be internalized; yet many can be. Indeed, a primary theme of Chandler's (1977) classic study is that firms are an alternative to markets and succeed in internalizing certain externalities to solve failures of coordination. (See Sah and Stiglitz 1989 for a discussion of “diffuse externalities” that are relatively unamenable to internalization; see also Stiglitz 1991.)

Similarly, it makes a great deal of difference whether the locus of increasing returns is within an industry or within the broader economy. In the former case even a small economy can, by specializing, avail itself of increasing returns; surely there are industries in which the minimum efficient scale of production is relatively small.

6. Or consider the argument that because early innovators get to choose the product in which they then specialize, they can choose a product with a better learning curve. In an international context these effects are essentially undone by changes in relative prices (Skeath 1989). Indeed, if we focus, for simplicity, on the case of unitary price elasticities, price effects will precisely undo output effects, so that income rates of growth will be the same in all countries.

When the economies of scale are spread more broadly, one must ask how they arise. And offsetting these economies of scale are diseconomies of scale—congestion economies. Regional economics provides some insight into these issues. We see agglomerations, economic activity that is not dispersed. Yet we also see viable communities, with high per capita income, that are relatively isolated geographically and are relatively small, certainly under a million population. These communities are, of course, part of larger communities, but what are the effective barriers? If there were none, everyone would be equally a part of the world economy and could take advantage of whatever economies of scale were relevant at this highly aggregate level. But costs of communication and transport help delimit the scope of communities. These costs, in turn, have implications for patterns of development; at certain stages of development and for certain products, they may be larger. Unfortunately, models with aggregate increasing returns to scale give us absolutely no insight into the relevant issues.

One attraction of models with economies of scale and externalities is that using models with nonconvexities and externalities makes it easy to construct multiple equilibria, as Krugman effectively illustrates. (See also Sah and Stiglitz 1989; Stiglitz 1987, 1991; Murphy, Shleifer, and Vishny 1989). And it is tempting to try to interpret the differing situations in which industrial and developing countries find themselves as reflecting these different equilibria. But again, we hardly need nonconvexities and externalities to generate multiple equilibria. Solow (1956) showed us how we could do that with his simple model; all we need is to have savings rates or reproduction rates depend (in a particular way) on the capital-labor ratio. These models were inadequate because some of their central implications—such as convergence in the rate of growth of income per capita and equalization of factor prices⁷—seemed counterfactual.

Differences in Technology

There is, fundamentally, only one way to resolve the paradox that all factors receive lower returns: the “effective” technologies in the two countries are different. There are two reasons that this might be so. If economies of scale are significant, larger economies are better off. For reasons already cited, I find this

7. Stiglitz (1970) and Inada (1968) extend the standard theory to the context of growth. The implications for factor prices across countries remain even after human capital is introduced; they are simply a consequence of the negative slope of the factor price frontier. For instance, if interest rates are equalized, it must be the case that if unskilled wages are lower in one country, skilled wages are higher. The critical assumption, of course, is that all countries face the same technology. By the same token, in international trade models with factor price equalization, such as that cited by Helpman and Krugman (1985)—as in earlier models of local public goods with free migration of labor and goods (for example Stiglitz 1977)—one can easily obtain asymmetric multiple equilibria; yet welfare of all those of a given ability is identical in all communities and countries. Such models, accordingly, have little to contribute to our understanding of the development problem.

explanation—at least as it is usually presented—at best incomplete and at worst misleading or wrong.⁸

The second reason that technology may be less effective in a developing country is simply differential knowledge. To economists who are used to assuming that everyone has access to best-practice technology, this explanation is anathema; it is too simple, it is ad hoc, or it leaves unexplained why countries lack access to best-practice technology. Yet once we recognize that information is costly to obtain and disseminate, that firms in industrial countries may have strategic reasons for withholding their most advanced technology, and that local conditions make necessary adaptation of the technology for the particular country, the explanation of differential knowledge makes perfect sense (see Gans 1989). That it is common sense is a virtue, not a vice.

The developing countries provide a rich set of facts and phenomena to be explained. The challenge for economic theory is to devise models that accommodate as many of these as possible. Doing so will, as Krugman rightly says, take us back to what he calls high development economics, but it is a vision of high development economic theory which, although it incorporates externalities and nonconvexities, is richer and more complex than one that incorporates those features alone.

REFERENCES

- Amsden, A. E. 1989. *Asia's Next Giant: South Korea and Late Industrialization*. New York: Oxford University Press.
- Aoki, Masahiko. 1970. "A Note on Marshallian Process under Increasing Returns." *Quarterly Journal of Economics* 84(1) (February):100–12.
- . 1987. "An Incentive-Compatible Approximation of Nash-like Solution under Nonclassical Environments." In Thomas Graves, Ray Radner, and Stanley Reiter, eds., *Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*. Minneapolis: University of Minnesota Press; London: Blackwell.
- Arrow, K. J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29 (June):155–73.
- Arthur, W. B. 1985. "Competing Techniques and Lock-in by Historical Events: The Dynamics of Allocation under Increasing Returns." Stanford University.

8. One variant of the model attempts to identify the source of the returns by focusing explicitly on issues of nontradability (Rodríguez 1992). It shows how a *small* economy—open to trading many, but not all, goods and services—may be caught in a low-level equilibrium in which there are no incentives for capital to flow into the country and in which both skilled and unskilled laborers receive low wages. It rests on the reasonable hypothesis that there are nonconvexities in the production of intermediate goods, some of which (such as services) are essentially nontradable, and that the range of the intermediates that are available depends on the pattern of production of final goods. Countries that produce to their current comparative advantage (based on their current supply of these intermediate goods, not the underlying factor prices) may produce final goods that do not generate demand for the large variety of intermediate goods needed to produce complex goods at competitive costs. But these intermediate goods form the basis of industrialization. What is attractive about this kind of modeling is that it goes well beyond an appeal to aggregate economies of scale or externalities.

- . 1988. "Competing Technologies: An Overview." In Giovanni Dosi, ed., *Technical Change and Economic Theory*. London: Frances Pinter.
- . 1989. "Competing Technologies, Increasing Returns and Lock-in by Historical Events." *Economic Journal* 99:116-31.
- Atkinson, A. B., and J. E. Stiglitz. 1969. "A New View of Technological Change." *Economic Journal* 79:573-78.
- Bardhan, Pranab. 1970. *Economic Growth, Development, and Foreign Trade: A Study in Pure Theory*. New York: Wiley Interscience.
- Chandler, Alfred. 1977. *The Visible Hand*. Cambridge, Mass: Belknap Press of Harvard University Press.
- Coase, Donald. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3:1-44.
- Dasgupta, Partha, and J. E. Stiglitz. 1980a. "Industrial Structure and the Nature of Innovative Activity." *Economic Journal* 90 (June):266-93.
- . 1980b. "Uncertainty, Market Structure and the Speed of R&D." *Bell Journal of Economics* 11(1) (Spring):1-28.
- . 1988. "Learning by Doing, Market Structure and Industrial and Trade Policies." *Oxford Economic Papers* 40:246-68.
- David, Paul. 1987. "Some New Standards for the Economics of Standardization in the Information Age." In Partha Dasgupta and Paul Stoneman, eds., *Economic Policy and Technological Performance*. Cambridge, U.K.: Cambridge University Press.
- Dosi, Giovanni, ed. 1988. *Technical Change and Economic Theory*. London: Frances Pinter.
- Edlin, Aaron, and J. E. Stiglitz. 1992. "Discouraging Rivals: Managerial Rent Seeking and Economic Inefficiencies." Stanford University, Department of Economics, Stanford, Calif.
- Gans, Joshua S. 1989. "Knowledge of Growth and the Growth of Knowledge." *Information Economics and Policy* 4:201-24.
- Greenwald, Bruce C., and Joseph E. Stiglitz. 1986. "Externalities in Economies with Imperfect Information and Incomplete Markets." *Quarterly Journal of Economics* 101:229-64.
- Greenwald, Bruce C., Meir Kohn, and Joseph E. Stiglitz. 1990. "Financial Market Imperfections and Productivity Growth." *Journal of Economic Behavior and Organization* 13 (June):321-45.
- Greenwald, Bruce C., A. Levinson, and J. E. Stiglitz. 1992. "Capital Market Imperfections and Regional Economic Development." Paper prepared for CEPR conference on Finance and Development in Europe, Santiago, Spain, December 1991.
- Helpman, Elhanan, and Paul R. Krugman. 1985. *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press.
- Inada, K-I. 1968. "Free Trade, Capital Accumulation and Factor Price Equalization." *Economic Record* 44 (September):28-37.
- . 1969. "Endogenous Technical Progress and Steady Growth." *Review of Economic Studies* 36(1):99-107.
- Itoh, Motoshige, and others. *Economic Analysis of Industrial Policy*. San Diego, Calif.: Harcourt Brace Jovanovich.
- Kaldor, Nicholas. 1970. "The Case for Regional Policies." *Scottish Journal of Political Economy* 17:337-48.
- . 1972. "The Irrelevance of Equilibrium Economics." *Economic Journal* (December):1237-55.
- Kaldor, Nicholas, and J. A. Mirrlees. 1969. "A New Model of Economic Growth." *Review of Economic Studies* 82 (June):174-92.

- Katz, Jorge. 1987. "On the Microeconomics of Technical Progress." *Technology Generation in Latin American Manufacturing Industries*. Basingstoke, U.K.: Macmillan.
- Komiya, Ryutaro, Masahiro Okuno, and Kotaro Suzumura. 1988. *Industrial Policy of Japan*. San Diego, Calif.: Academic.
- Lewis, W. A. 1949. *Overhead Costs*. New York: Holt Rinehart.
- . 1954. "Economic Development With Unlimited Supplies of Labor." *Manchester School of Economics and Social Studies* 22(2):139–91.
- . 1955. *The Theory of Economic Growth*. London: Allen and Unwin.
- Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97(5):1003–26.
- Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Pack, Howard, and L. E. Westphal. 1986. "Industrial Strategy and Technological Change: Theory versus Reality." *Journal of Development Economics* 22:87–128.
- Rodríguez, Andrés. 1992. "The Big Push in a Small Open Economy." Stanford University. Department of Economics, Stanford, Calif.
- Romer, Paul. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94(5):1002–37.
- Rosenstein-Rodan, Paul N. 1943. "Problems of Industrialization in Eastern and South-eastern Europe." *Economic Journal* 53 (June–September):202–11.
- Sah, Ray, and J. E. Stiglitz. 1989. "Sources of Technological Divergence between Developed and Less Developed Countries." In Guillermo A. Calvo and others, eds., *Debt, Stabilization and Development: Essays in Memory of Carlos Diaz-Alejandro*. Oxford, U.K.: Blackwell.
- Shell, Karl, ed. 1967. *Essays on the Theory of Optimal Economic Growth*. Cambridge, Mass.: MIT Press.
- Shleifer, Andrei, and Robert W. Vishny. 1989. "Management Entrenchment: The Case of Manager-Specific Investments." *Journal of Financial Economics* 25:123–39.
- Skeath, Susan. 1989. "Learning, Price Effects, and Income Growth." Princeton University, Princeton, N.J.
- Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70(1):101–8.
- Stiglitz, J. E. 1970. "Factor Price Equalization in a Dynamic Economy." *Journal of Political Economy* 78 (May–June):456–89.
- . 1977. "Theory of Local Public Goods." In M. S. Feldstein and R. P. Inman, eds., *The Economics of Public Services*. New York: Macmillan.
- . 1985. "Information and Economic Analysis: A Perspective." *Economic Journal* 95, supplement (March):21–41.
- . 1987. "Learning to Learn: Localized Learning and Technological Progress." In Partha Dasgupta and Paul Stoneman, eds., *Economic Policy and Technological Performance*. Cambridge, U.K.: Cambridge University Press.
- . 1988. "Technological Change, Sunk Costs, and Competition." *Brookings Papers on Economic Activity* 3:883–947.
- . 1991. "Social Absorption Capability and Innovation." Paper prepared for Korean Development Institute, 20th Anniversary Symposium, Seoul, June.
- . 1992. "Whither Socialism?: Perspectives from the Economics of Information." Wicksell Lectures presented in Stockholm, May.
- Teubal, Morris. 1967. "Inflation, Unemployment, and Economic Growth." Ph.D. thesis. University of Chicago, Chicago, Ill.

- Uzawa, Hirofumi. 1963. "On a Two-Sector Model of Economic Growth II." In J. E. Stiglitz and Hirofumi Uzawa, eds., *Readings in the Modern Theory of Economic Growth*. Cambridge, Mass.: MIT Press.
- . 1965. "Optimal Technical Changes in an Aggregative Model of Economic Growth." *International Economic Review* 6 (January):12-31.
- Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton, N.J.: Princeton University Press.
- Wright, Gavin. 1987. "The Economic Revolution in the American South." *Journal of Economic Perspectives* 1 (Summer):161-78.

COMMENT ON "TOWARD A COUNTER-COUNTERREVOLUTION
IN DEVELOPMENT THEORY," BY KRUGMAN

Lal Jayawardena

Professor Krugman seeks to explain the disappearance of "high development theory," which spanned the period between 1943 and 1958 and was associated with such names as Rosenstein-Rodan, Fleming, Nurkse, Myrdal, Scitovsky, and Hirschman. He argues that its eclipse, despite the continuing relevance of its insights to the ever-present problem of accelerating development, occurred because its pioneers were unable "to turn their intuitive insights into clear-cut models that could serve as the core of an enduring discipline."

Krugman draws out the policy implications that follow from his legitimization of the underlying economic assumptions of high development theory in a manner that would carry conviction to mainstream economic theorists. There seem to be two implications: first, a caution against carrying a free-market orthodoxy too far because there is an intellectually solid case for some government promotion of industry and, second, a need to shift the research on trade and industrial policy away from its focus on government failure or the horrors of import substitution and the distortions imposed by government policy and toward a concern with market failure.

The difficulty confronting anyone asked to comment on Krugman's paper is that there is not a great deal in it with which one can disagree. My principal reservation is whether the failure of high development theory to model its results was the sole reason for its disappearance. I therefore faxed one of the few surviving members of that generation, Albert Hirschman, to check out my intuition. His faxed response is worth quoting:

My reaction to the Krugman paper is, of course, that I am delighted: delighted to be rehabilitated and to be present, unlike most other contributors to high development theory, at our collective rehabilitation at the hands of Krugman. It certainly is true that the failure of the proponents of "high development theory" to speak the modeling language contributed to the loss of influence of their ideas in the 1970s and 1980s. But here I agree

Lal Jayawardena is assistant secretary-general, United Nations, and director, World Institute for Development Economics Research, Helsinki.

© The International Bank for Reconstruction and Development / THE WORLD BANK

with you that this is not the whole story. His is essentially what is called an "internalist" account in the history and sociology of science. I tried to supply the "externalist" side of the story in my article on "The Rise and Decline of Development Economics."

Hirschman (1981), in providing his "externalist" explanation, argues that development economics as formulated in the period of high development theory was characterized by two basic ingredients: a rejection of the *monoeconomics* claim, which argued that the same economic laws were applicable to both developing and industrial countries, and an assertion of the *mutual benefit* claim, which held that economic relations between these two groups of countries could be handled in ways that would yield gains to both. The available competing approaches to economics were situated differently with respect to each of these ingredients. Orthodox economics asserted both the monoeconomic and mutual benefit claims, while the neo-Marxist economic theories, which gained ground later, rejected both claims. Hirschman traces the decline of development economics to the consequences of the "strange alliance of neo-Marxism and monoeconomics." In brief, the neoclassical right faulted it for having forsaken the true principles of efficient resource allocation prescribed by monoeconomics in that it, for example, promoted inefficient import-substituting industrialization in developing countries. And for the neo-Marxist left, exemplified by economists such as Paul Baran (1952), development economics was not sufficiently radical.

Even more fatal, in Hirschman's view, were the political disasters that struck many developing countries after the 1960s. It was the resulting self-doubt and failure to mount a counterattack against the "unholy alliance of neo-Marxists and neoclassicists" that doomed development economics. At the same time, growing disenchantment with development assistance undermined interest in the "mutual interests" idea—in spite of attempts by such international groups as the Brandt Commission (Independent Commission on International Development Issues 1980) to keep the theme alive. (Paradoxically, the current concern about the global environment might help reinstate the mutual benefit claim for adequate resource transfers to developing countries and lead to a revival of interest in development economics.)

Krugman's thesis has also been underlined in Williamson's (1990) paper, which states the "Washington consensus" on policy reform. Having outlined the five elements that constitute the consensus—a balanced budget, relative price correction (principally a competitive exchange rate), liberalization of trade and foreign investments, privatization, and domestic market deregulation—Williamson adds:

A striking fact about the list of policies on which Washington does have a collective view is that they all stem from classical mainstream economic theory, at least if one is allowed to count Keynes as a classic by now. None of the ideas spawned by the development literature, such as the "big push,"

balanced or unbalanced growth, surplus labour, or even the two-gap model, plays any essential role in motivating the Washington consensus. This raises the question as to whether Washington is correct in its implicit dismissal of the development literature as a diversion from the harsh realities of the dismal science.

What is significant is that the only reaction to this view during the conference came from Stanley Fischer, chief economist of the World Bank at the time, who argued against an "active industrial policy" in developing countries despite its success "on balance . . . in parts of East Asia" (Williamson 1990).

INDUSTRIAL POLICY IN EAST ASIA

Any lack of rigor in articulating high development theory did not prevent its determined and fruitful application in East Asia, particularly in Japan and the Republic of Korea, whose planning models embodied its insights. Unlike the cases of South Asia and Latin America, there is little evidence to support the view that high development theory was invoked in East Asia in support of *inefficient* import-substituting industrialization. *Efficient* import substitution, in contrast, endeavored to promote selected infant industries with the potential to compete as exporters in international markets. The success of East Asian interventionist strategies in "picking winners" has sparked an intense debate within the World Bank as to whether the standard trade liberalization component of the Washington consensus is appropriate for developing countries or should be modified to provide both "functional" (nondiscriminatory) and "selective" (discriminatory) government intervention.

The key insight of high development theory was that the coordination implicit in the communication of price signals was not sufficient where significant economies of scale existed. Thus, investment decisions required explicit coordination by a planning authority to achieve efficient allocation of resources. The implication of this insight was that while the price mechanism could be relied on to take care of the *production* problems of an economy, the allocation of *investment* could not be left to the price mechanism but would require state intervention.

This insight was incorporated at a very early stage in East Asian (specifically, Korean) planning and was implemented through planning models (Westphal 1979) for those sectors in which substantial economies of scale existed. These essentially took the form of numerically specified general equilibrium models. With allowance for economies of scale and following a modified input-output approach, these programming models were used to determine the optimal pattern of investment when the composition of final demand depends only on per capita income and the main choices are between domestic production and imports. Note that from a policy perspective a methodology *was* available to arrive at meaningful conclusions on the problem of vertical (intertemporal) and horizontal (intersectoral) interdependence within the framework of an econ-

omywide model that explicitly allowed for economies of scale in specified sectors.

The rapid development of the Republic of Korea, Taiwan (China), Singapore, and Hong Kong has led development economists to draw quite opposite policy conclusions as to the underlying reasons. Those concerned with market failure point to successful industrial policy and trade protection. Those concerned with government failure point to the adoption of sensible macroeconomic policies—low inflation, competitive exchange rates, a commitment to small government, and an insistence on guiding (rather than supplanting) a market mechanism with an educated labor force and entrepreneurial skills. The traditional World Bank view cited earlier falls within the second group, although there are indications in the current internal debate that the first view is gaining prominence (The issues in this debate are summarized in World Bank 1991.)

The general World Bank approach to industrial policy qualifies this stance only to the extent of being moderately neoclassical; in other words, because factor and product markets are not fully efficient in developing countries, there is a role for a modest degree of government intervention. The strong preference is for functional, as opposed to selective, intervention. In sum, the general approach concentrates on getting prices right and on limiting the range of interventions in specific industrial sectors and firms.

In its microlevel approach, in contrast, the Bank takes a different view. In India, for example, it recommends “selective” interventions to support each industry, including specific measures for improving design, technology, equipment, management, and marketing. The export strategy study recommends “picking specialized and high-performance exporters” for “selective” support. The capital goods study lists institutional and technological measures to boost competitiveness. These are not neutral interventions; they can all be considered forms of “picking winners.”

The upshot of the debate was a compromise that called for selective interventions together with traditional across-the-board incentives for industrialization. Ultimately, there does not appear to be any uniform prescription for industrialization (see Taylor 1991).

Japan’s view of industrial policy is consistent with its belief in the protection of immature industries. An important recent paper by Japan’s aid agency, the Overseas Economic Cooperation Fund (OECF), challenges the conditionality of rapid trade liberalization in structural adjustment loans.

If imports are liberalized too quickly, is it possible to develop industries that will play leading roles in the next stage of economic development? Is it not necessary to protect domestic industry to some extent for a certain period of time in order to allow a viable export industry to develop? (OECF 1991)

The paper argues that conventional trade liberalization is limited to capturing the static comparative advantage typical of a developing country, which is to be

found mainly in primary products and light industry with low value added. Since the objective of developing countries is to move toward high value added production, with more sophisticated technologies and a more substantial growth potential, the OECF concludes that "sticking to simple trade liberalization based on static comparative advantage may have a negative impact on the possibility of economic development." The study argues that it is too optimistic to expect that industries with high value added will automatically emerge from the private sector in the absence of deliberate measures to foster such industries, including protection. The OECF's preferred policy prescription is to protect chosen industries for a specified period, taking precautions to avoid the harmful effects of protection.

The concern about too rapid a liberalization is also an important issue in Eastern Europe's economic transformation. Katz (1991) suggests that instead of subjecting industry to the full force of international competition, Eastern Europe might learn from East Asia's approach.

The discussion above shows that, quite independently of the formal legitimization of the underlying economic assumptions of high development theory, the successful application of its policy prescriptions in East Asia provides an empirical basis for its validity. These empirical considerations (now reinforced by Krugman) constitute a powerful case for revising the trade liberalization components of the Washington consensus, or at least excising it from adjustment loan conditionality.

The necessary revision would seem to require a somewhat less tentative approach to interventionist trade and industrial policy than is suggested in Krugman's paper. In other words, instead of merely cautioning against "carrying a free-market orthodoxy too far," the World Bank and other multilateral agencies might incorporate in structural adjustment lending specific ways of developing sound long-term industrial policies (including both functional and selective interventions) and ensuring that trade liberalization does not impede these interventions.¹

What remains to be considered is Krugman's suggested reorientation of the research agenda to focus on market failure. The World Institute for Development Economics Research (WIDER) has already made a beginning in this direction, based in part on Krugman's work on new trade theory (Helleiner 1992). A challenge to conventional wisdom is contained in a WIDER volume edited by Banuri (1991). A third area of WIDER research has begun to examine whether

1. The OECF (1991) suggests a continuing policy dialogue on industrial development between donors and developing countries. The institutional location for such a dialogue could be World Bank aid groups or United Nations Development Programme (UNDP) roundtables, which are all under a certain amount of pressure to take a longer-term view of development. The case has been reinforced both by the social costs of short-term adjustment during the 1980s and by the need to incorporate long-term environmental considerations into a country's development strategy. (See Jayawardena 1991 for a discussion of this approach and of the financing requirements for achieving a minimum rate of growth in developing countries in the 1990s.)

markets can be relied on to bring about adjustment and sustainable growth within a politically sensible timeframe without incurring extensive social costs—and, if not, what supporting mechanisms are needed to offset market failure (Taylor 1988). Other areas of research have attempted to bring together neo-classical and nonneoclassical viewpoints on the specific issue of the economic transformation of centrally planned economies (see Blanchard and others 1991; Kornai 1990; and Hansson 1992).

I have tried to do no more than outline WIDER's current work on the kind of reoriented research agenda that Krugman has suggested. I would hope that, as a follow-up, interested research institutions could pursue some form of pooling and coordination of effort in the most promising research areas.

REFERENCES

- Banuri, Tariq, ed. 1991. *Economic Liberalization: No Panacea. The Experience of Latin America and Asia*. WIDER Studies in Development Economics. Oxford, U.K.: Clarendon Press.
- Baran, Paul A. 1952. "On the Political Economy of Backwardness." In Robert I. Rhodes, ed., *Imperialism and Underdevelopment: A Reader*. New York: Monthly Review Press.
- Blanchard, Oliver, and others. 1991. *Reform in Eastern Europe. The 1990 Report of the WIDER World Economy Group*. Cambridge, Mass.: MIT Press.
- Hansson, Ardo. 1992. "Report on the conference 'Transformation from a System of Central Planning to a Market Economy,' jointly sponsored by the Government of Finland and WIDER, Helsinki, February 6–7."
- Helleiner, Gerald K. 1992. *Trade Policy, Industrialization, and Development*. WIDER Studies in Development Economics. Oxford, U.K.: Clarendon Press.
- Hirschman, Albert O. 1981. "The Rise and Decline of Development Economics." In *Essays in Trespassing: Economics to Politics and Beyond*. New York: Cambridge University Press.
- Independent Commission on International Development Issues ("Brandt Commission"). 1980. "Mutual Interests." In *North-South: A Program for Survival*. London: Pan.
- Jayawardena, Lal. 1991. "A Global Environmental Compact for Sustainable Development: Resource Requirements and Mechanisms." *WIDER Research for Action*. Helsinki.
- Katz, S. Stanley. 1991. "East Europe Should Learn from Asia." *Financial Times*. April 24.
- Kornai, János. 1990. *The Road to a Free Economy: Shifting from a Socialist System: The Example of Hungary*. New York: W. W. Norton.
- OECD (Overseas Economic Cooperation Fund). 1991. "Issues Related to the World Bank's Approach to Structural Adjustment—Proposal from a Major Partner." Occasional Paper 1. Tokyo.
- Taylor, Lance. 1988. *Varieties of Stabilization Experiences: Towards Sensible Macroeconomics in the Third World*. WIDER Studies in Development Economics. Oxford, U.K.: Clarendon Press.
- . 1991. *Income Distribution, Inflation, and Growth: Lectures on Structuralist Macroeconomic Theory*. Cambridge, Mass.: MIT Press.

- Westphal, Larry E. 1979. "Economies of Scale and Investment over Time." In Hollis Chenery, ed., *Structural Change and Development Policy*. New York: Oxford University Press.
- Williamson, John, ed. 1990. *Latin American Adjustment: How Much Has Happened?* Washington, D.C.: Institute for International Economics.
- World Bank. 1991. "Industrialization in Newly Industrializing Countries: Case Studies of Korea, India and Indonesia." Vol. 1. World Bank, Operations Evaluation Department, Washington, D.C.

FLOOR DISCUSSION OF THE KRUGMAN PAPER

Responding to Joseph Stiglitz (discussant), Krugman adapted a saying: "Those who can, do, and those who can't, worry about definitions." He and Stiglitz, being amateur economic historians, seemed to be locked into defining what constitutes "mainstream" and deciding whether things had left it or not. People used to say that everything Milton Friedman saw made him think about money; so everything Krugman saw made him think of increasing returns, and many things Stiglitz saw made him think of asymmetric information.

In response to Lal Jayawardena (discussant), Krugman said it was an open question how much the success of East Asian economies could be used to support high development theory. Many people believe that they represent a success for interventionist policies; many others believe that interventionist policies are second order and that the East Asian economies represent a success for outward-oriented policies. Either way, their success is a reproach to classical models, which can explain neither avenue to success.

One participant welcomed Krugman's rescue of important elements from development theory that might still play even a partial role in strategies for economic development. He felt that the counter-counterrevolution in development theory should include not only growth theory and the new theoretical developments emphasized in Krugman's paper but also the new themes that microeconomic research, including case studies, have shown to be important in the past twenty years. No bag of tricks yet existed to formalize these new themes.

A participant said that he fully endorsed what Stiglitz said about Krugman's selective and dated reading of the development literature. The Rosenstein-Rodan Big Push paradigm had not failed or been abandoned because of lack of formal modeling, the speaker contended, but because it was totally irrelevant once you allowed that the world was an open, not a closed, economy. Rescuing it now was not useful unless you addressed the two reasons it failed. World trade was thriving in the 1950s and 1960s, so it was pointless to talk about whether domestic demand restricted total demand or economies of scale; the minimum efficient scale for many products developing countries could produce was so small in the context of world demand that models which emphasized indefinite increases in returns were irrelevant as a basis for policymaking. Similarly, it was

The session was chaired by Vinod Thomas, chief economist, East Asia and Pacific Regional Office, the World Bank.

difficult to find indefinite externalities of the pecuniary or technological variety. Even now, the most important avenue to development is world trade, not the inward-oriented focus that Krugman's rescued development theory would suggest.

Another participant asked what in this growth theory was relevant today, when developing countries—including China and India—were more open than at any time in recent memory.

Krugman responded that many people believe that because of the potential of world trade, anything having to do with market size within a country is now irrelevant. We must keep some perspective, he said, about what is and is not trade. The Federal Reserve Bank of Boston did a systematic study of the tradable and nontradable sectors in the Massachusetts economy, characterizing sectors as serving either the local market or the national or global market. That study showed that less than 30 percent of employment was in what one would consider tradables. Even perfectly open economies such as the state of Massachusetts are engaged primarily in producing for the domestic market, he said, and these issues are relevant where the domestic market matters. But increasing returns do apply domestically—often to intermediate inputs (both intermediate goods and, more important, pools of skilled labor).

Krugman admitted that there was some fallacy of misplaced concreteness in using these models to get at their importance. He said that he would be happy to widen the range of possibilities, but he did not think one could simply dismiss their relevance. He did not believe that in the late 1960s people decided that they understood what Rosenstein-Rodan said but it was no longer relevant because they had open economies.

Larry Summers shared the speakers' general sense that things were shifting toward a more activist era, with good reason. He asked if anything in the Big Push, returns-to-scale concept did—or should—have anything to do with that shift toward activism. He said that the principal advocate of the idea that small producers in small markets are inefficient and that producers can be more efficient in large markets was Stalin, who sought to enlarge enterprises at the expense of differentiation. The idea of making markets bigger so there would be increasing returns, he suggested, was not as important a theme to activists as three others: the importance for industrial policy of "learning-by-doing" external economies; the importance of using command-and-control approaches to manage change until you have incentives in place, because you can't create or change markets if the right institutions are not in place; and the need to work around barriers that stop things from going from where they are less productive to where they are more productive (the old dualism theme).

Krugman said he was getting used to the bit about Stalin—when one deviates from the "Washington consensus," one is immediately accused of being a Stalinist, which he was not. As a matter of intellectual strategy, Krugman contended that it was important to lean against the wind regarding the tendency to emphasize somewhat "cosmic" concepts, such as the generation of knowledge and

technology. It was not that those futuristic topics were not important but that they were so much more tempting to explore than the grungy details of economies of scale in a plant or the greater likelihood of finding specialized skills in thick labor markets, which was part of the same story. We have an obligation, he said, to lean toward the less glamorous aspects of the problem because there is a strong tug the other way. In his opinion, technology as usually discussed was overrated; the crucial externalities often lay in more mundane types of interactions. As for dualism, he did not mean to preclude other options, and perhaps his bias was too narrow and concrete.

A participant from the World Bank's East Asia region said that Krugman had ignored one element in the history of economic thought: the sociology of the economics profession. In the 1970s and 1980s the agenda for development economics had been heavily influenced by the dominant ideologies of institutions such as the World Bank, not only through their funding of research but also, and more insidiously, because their resources and their grant of legitimacy affected the economics profession, especially in developing countries. A prime example is the World Bank's capacity-building initiative in Africa.

A World Bank participant made a distinction between a mathematical model and verbal or logical models, including sociological models, which tend to be forgotten because they are difficult to carry over historically. A second participant said that Krugman—and to some extent Stiglitz—stressed the importance of formal modeling as a legitimization of economic knowledge. He asked if they would agree that, given Krugman's definition, legitimate knowledge would include Euclidian geometry but would not include Darwin's theory (a nonformal, nonmathematical model that couldn't make predictions) or a great deal of nineteenth century physics and chemistry.

A participant asked Jayawardena to comment on Krugman's statement that it made sense for the World Bank and other multilateral agencies to push hard for liberal policies in developing countries, given their demonstrated tendency to engage in economically irrational interventions. Was it premature to start thinking about a counter-counter-counterrevolution? the participant asked.

Krugman closed the session by referring to the quotation about consistency being the hobgoblin of little minds. There is internal consistency, and there is consistency across different issues, he said, and nobody could accuse him or Stiglitz of having a consistent model of the world. He referred, finally, to another quote, this time from an economist, that the most reckless and dangerous theorist is the man who professes to let the facts speak for themselves. It is important to discipline oneself, he said. What he meant by a model was not necessarily a mathematical model but putting oneself, at least briefly, in an intellectual straitjacket just to be sure of actually staying in the same place for a little while.

Two Strategies for Economic Development: Using Ideas and Producing Ideas

Paul M. Romer

The key step in understanding economic growth is to think carefully about ideas. This requires careful attention to the meaning of the words that we use and to the metaphors that we invoke when we construct mathematical models of growth. After addressing these issues, this paper describes two different ways in which ideas can contribute to economic development. The history of Mauritius shows how a poor economy can benefit by using ideas from industrial countries within its borders. The history of Taiwan (China) shows how a developing economy can be pushed forward into the ranks of those that produce ideas for sale on world markets.

The central claim of this paper is that the difference between the economics of ideas and the economics of objects is important for our understanding of growth and development. A subsidiary claim is that academic and policy discussions in these areas might be more fruitful if we spent less time working out solutions to systems of equations and more time defining precisely what the words we use mean. The notion that ideas are different from objects is both familiar and obvious. In the economic analysis of patents, for example, there is a long tradition of recognizing the unique characteristics of ideas as economic goods. The content of the claim lies in the assertion that these differences are more subtle than some presentations suggest and that they matter for aggregate-level policy analysis.

All too often, economists concerned with the economy as a whole have been willing to treat the economics of ideas as a footnote to the rest of economic analysis—important for understanding some of the details but not something that changes how we think about big policy questions. A neoclassical model with perfect competition and exogenous technological change continues to frame many, if not most, policy discussions of growth and development. Ideas are routinely ignored.

Paul M. Romer is a professor at the University of California, Berkeley. The author would like to thank conference participants and the discussants for much useful feedback. Detailed comments by George Akerlof, Curtis Eaton, Richard Lipsey, and David Romer were also very helpful. This work was supported by National Science Foundation grant SES 9023469.

In what follows, two kinds of support are offered for the claim that ideas should be our central concern: abstract arguments about the economic attributes of ideas and descriptions of the role that ideas played in two cases of successful economic development. The abstract arguments presented in the first part of the paper proceed from the observation that ideas are extremely important economic goods, far more important than the objects emphasized in most economic models. In a world with physical limits, it is discoveries of big ideas (for example, how to make high-temperature superconductors), together with the discovery of millions of little ideas (better ways to sew a shirt), that make persistent economic growth possible. Ideas are the instructions that let us combine limited physical resources in arrangements that are ever more valuable.

As economic goods, ideas differ from objects in ways that are more subtle than traditional aggregate models allow. The familiar description of an idea as a public good like a lighthouse beacon overlooks the fact that many ideas are controlled by private individuals who respond to market incentives. An equally misleading approach is to equate ideas with human capital and to treat them as conventional private goods. This misses the notion, correctly suggested by the public good analogy, that an idea can be used by many people at the same time. Adding external effects or spillovers to human capital and physical capital comes no closer to capturing the essential attributes of ideas. Externalities suggest incomplete control or appropriability, but they do not capture the absence of opportunity costs that is the key characteristic of an idea. The combination of some degree of private control and an absence of opportunity costs means that ideas are neither public goods nor private goods—nor a mixture of the two.

To address a frequently expressed objection to this description of an idea, the abstract discussion acknowledges that ideas are used in fixed proportions with objects that do have an opportunity cost. For example, the bit string representing a computer program can be used by an unlimited number of people with no loss of functionality for the first user. In practice, however, it takes a floppy disk to make a copy of the bit string for another user, and the floppy disk does have an opportunity cost. The distinction between capital and labor is useful even though both are needed to produce output; so is the distinction between the bit string and the floppy disk.

The abstract analysis of ideas has special relevance for poor countries. Industrial countries possess a stock of ideas that could yield large increases in standards of living if they were put to use throughout the world. Moreover, since the use of an idea by one person does not limit its use by someone else, the large potential gains available to developing countries need not come at the expense of industrial countries.

In the few instances in which ideas have been controlled by international aid organizations, we can see just how large the gains from worldwide dissemination can be. The idea behind the smallpox vaccine has now eliminated this disease from every country on earth. The simple idea behind oral rehydration

therapy has saved the lives of millions of children who would otherwise have died from diarrhea.

These examples suggest the magnitude of the gains that ideas can offer, but they are not typical. Most ideas with economic value are not controlled by a charitable organization willing to bear the costs of dissemination. Instead, they are controlled by people who will not incur the costs needed to share what they know unless they have a monetary incentive to do so. As a result, the gains from the dissemination of ideas will not be realized if distortions, weak institutions, and bad political structures prevent the holders of ideas from sharing in the gains that accrue when the ideas are brought to a new geographic area.

For this reason, the logic behind the economics of ideas supports the new development orthodoxy that a policy of openness with few distortions offers the potential for large gains in poor countries. The experience in Mauritius, the first of the two economies considered below, supports this view. A poor island whose prospects for development once seemed very bleak, Mauritius successfully exploited a development strategy that consisted almost entirely of trying to make use of ideas that already existed in industrial countries by encouraging foreigners to produce there. This is the first of the two strategies noted in the title of this paper: using ideas. Judging from the increased receptivity of many poor countries to direct foreign investment, it appears that this strategy is gaining favor.

Yet as important as this strategy can be during the early stages of development, the analysis also suggests that there may be limits to how far it can take an economy. Both the experience in the second economy discussed, Taiwan (China), and the logic behind the analysis of ideas suggest that some interventions may encourage growth at intermediate stages of development. Taiwan pursued the second strategy described in the title—producing ideas—and intervention by the government seems to have contributed to the strategy's success.

Most economists would acknowledge that some kinds of intervention to support the production of ideas are appropriate. Few would challenge the assertion that governments should subsidize education and some forms of research. If one follows the logic of the economics of ideas, one sees that there is no basis in economic theory (as opposed to political theory) for restricting government intervention to support for education and research.

Many economists are also convinced that restrictions on trade and direct foreign investment are bad policy instruments for encouraging development. Yet the evidence from Taiwan (China) suggests that those policies were useful there. These observations can be reconciled by adding a political and institutional analysis to the economic analysis. An economic analysis based on the economics of ideas suggests that trade and investment restrictions can be growth-enhancing only if they are complemented by other policies such as support for education and rigorous standards of performance for protected firms. To be effective, these policies must be implemented by a government that is immune to the political pressures associated with rent-seeking and that possesses a competent and relatively honest bureaucracy. Because these conditions are not met in most

of the world, restrictions on trade and investment will almost always be counter-productive. The challenge is to find better forms of government intervention, ones that have better economic effects and pose fewer political and institutional risks.

The temptation for economists, however, has always been to duck the complicated political and institutional issues that this kind of analysis raises and instead to work backward from a desired policy conclusion to a simple economic model that supports it. According to this approach, if we want to discourage counter-productive restrictions on trade and foreign investment in most countries of the world, then the right model is one with perfect markets so that intervention can be shown to be everywhere and always a mistake.

The motivation behind this paper comes from a belief that the poor people of the world will be better served if we resist this temptation and meet head-on the intellectual challenge presented by the economics of ideas. In industrial countries a number of government interventions have evolved to encourage the production and dissemination of ideas. Examples include patents, copyrights, laws to protect trade secrets, subsidies for education at all levels, peer-reviewed research grants, and agricultural extension services. These interventions have been relatively free of political manipulation and, despite some weaknesses, are generally regarded as having had a positive effect.

The same arrogance that made people at the turn of the century think that almost everything had already been invented sometimes leads us to think that there is nothing left to discover about the institutions that can encourage economic development. It is conceivable that the institutions now present in industrial countries exhaust the list of beneficial interventions that would fit the circumstances of low- and middle-income countries. It is far more likely that there are undiscovered institutional arrangements that would work even better. We will never know if we always look at the evidence through a theoretical lens that does not let us consider this possibility.

I. MODELS AND METAPHORS

Most theoretical discussions of economic growth revolve around a few mathematical equations built from abstractions such as an aggregate production function. Because of its simplicity, theorists sometimes call such a system of equations a toy model to distinguish it from a multiequation simulation or forecasting model. The label is apt because a good theoretical model should be as easy to manipulate in one's head as the mental image of a child's toy. That is, a successful model invokes a metaphor, and the metaphor has a subtle but pervasive effect on the reader's understanding of the principles behind the equations and his or her belief in the accuracy of its description of the world. To show how one model can obscure our vision of the role of ideas and how a different model can highlight their role, it helps to start by being unusually explicit and concrete about the metaphors behind the math.

A Toy Model

One of the great successes of neoclassical economics has been the elaboration and extension of the metaphor of the factory that is invoked by a production function. To be explicit about this image, recall the child's toy called the Play-Doh Fun Factory. To operate the Fun Factory, a child puts Play-Doh (a form of modeling compound) into the back of the toy and pushes on a plunger that applies pressure. The Play-Doh is extruded through an opening in the front of the toy. Depending on the particular die used to form the opening, out come solid Play-Doh rods, Play-Doh I-beams, or lengths of hollow Play-Doh pipe.

We use the Fun Factory model or something just like it to describe how capital (the Fun Factory) and labor (the child's strength) change the characteristics of goods, converting them from less valuable forms (lumps of modeling compound) into more valuable forms (lengths of pipe). In most applications we imagine that the characteristics being changed are physical characteristics such as shape, chemical composition, or connections with other objects. We push the model slightly when we recognize that the date and location at which a good is available are also relevant characteristics, ones that can be changed by storage and transport. We push the model much further by extending the list of characteristics that can be changed to include a description of who holds property rights. Wallis and North (1986) estimate that by 1970 the transformation of property rights accounted for nearly one-half of gross national product in the United States, so this last extension is particularly important.

The Fun Factory metaphor is powerful because our intuition about production can be pushed to encompass transformation activities and levels of economic analysis far removed from the factory floor. When a worker with a welding rig attaches parts on a car, when a driver with a semitrailer truck moves the car, when a dealer with a showroom sells the car, when a banker with a computer prepares the loan, and when an agent with a tow truck repossesses the car, labor and capital are used to change the characteristics of the underlying goods. In this kind of analysis, the productive unit can range from a household to a firm or industry, even to the nation or the world as a whole.

The production function and the Fun Factory metaphor have been widely used in the neoclassical analysis of aggregate growth. Yet in this analysis the neoclassical model has been successful primarily at establishing a diagnosis by exclusion. Economic growth cannot be understood solely in terms of the accumulation of physical capital and labor—the fundamental concepts in the underlying metaphor. This insight, of course, was Solow's famous result (1957), and it stands to this day despite an enormous effort at refining the econometric techniques for measuring a growth-accounting residual and at extending the notion of effective labor to allow for accumulation of human capital. The formal growth-accounting evidence, historical accounts, and everyday experience all suggest that something extra, something like innovation, invention, technological change, or the discovery of new ideas, is needed to understand and

explain growth. Yet, having made this point, the Fun Factory metaphor offers no guidance about what an idea is, where ideas come from, and how the presence of ideas might matter for development strategy.

Other Toy Models

Another child's toy is a chemistry set. For this discussion, the set can be represented as a collection of N jars, each containing a different chemical element. From the child's point of view, the excitement of this toy comes from trying to find some combination of the underlying chemicals that, when mixed together and heated, does something more impressive than change colors (explode, for example). In a set with N jars, there are $2^N - 1$ different mixtures of K elements, where K varies between 1 and N . (There are many more mixtures if we take account of the proportions in which ingredients can be mixed and the different pressures and temperatures that can be used during mixing.)

As N grows, what computer scientists refer to as the curse of dimensionality sets in. The number of possible mixtures grows exponentially with N , the dimension of this system. For a modestly large chemistry set, the number of possible mixtures is far too large for the toy manufacturer to have directly verified that no mixture is explosive. If N is equal to 100, there are about 10^{30} different mixtures that an adventurous child could conceivably put in a test tube and hold over a flame. If every living person on earth (about 5 billion) had tried a different mixture each second since the universe began (no more than 20 billion years ago), we would still have tested less than 1 percent of all the possible combinations.

Within the metaphor of the chemistry set, it is obvious what one means by an idea. Any mixture can be recorded as a bit string, an ordered sequence of 0s and 1s of length 100. The bit at position j is set to 1 if element j is included in the mixture. In the crude representation used here, an idea is the increment in information that comes from sorting some of the bit strings into two broad categories: useful ones and useless ones. To represent this information, we can add two more bits on the end of each bit string describing a mixture. These are set at 00 if we know nothing about its properties, 10 if it is a useful mixture, and 01 if it is useless.

When a useful mixture is discovered and its trailing bits are changed from 00 to 10, the discovery makes possible the creation of economic value. It lets us combine raw materials of low intrinsic value into mixtures that are far more valuable. Once we have the idea, the process of mixing will require its own Fun Factory (specialized capital and labor). For example, the bit string representing nylon requires a chemical processing plant and skilled workers. Important as these tangible inputs are, it is still the idea itself that permits the resulting increase in value. In this fundamental sense, ideas make growth and development possible.

The potential for continued economic growth comes from the vast search space that we can explore. The curse of dimensionality is, for economic pur-

poses, a remarkable blessing. To appreciate the potential for discovery, one need only consider the possibility that an extremely small fraction of the large number of possible mixtures may be valuable.

There is a branch of physical chemistry that literally cooks up mixtures from the periodic table of elements. New mixtures to be evaluated are selected on the basis of theory, experience, and guesswork. Supporters call this "exploratory synthesis"; detractors call it "heat and beat" or "shake and bake" chemistry. A group of French chemists cooked up one of the 10^{30} possible mixtures, one consisting of lanthanum, barium, copper, and oxygen. More than a decade later, scientists at IBM decided to test the superconductivity properties of the resulting ceramic, even though the prevailing wisdom suggested that it violated several of the basic rules required of a candidate for a good superconductor. The IBM team won the Nobel Prize in physics for their discovery that this mixture became a superconductor at temperatures far exceeding those for all the known superconductors.

This "high-tech" example of a valuable mixture suggests only a small part of the enormous scope for making discoveries of economic importance. If a garment factory requires 52 distinct independent steps to assemble a shirt, there are $52! = 10^{68}$ different ways to order these steps in sequence. The number 52 is a useful reference point because it arises with another familiar toy, a deck of cards. The number of possible orderings for the 52 assembly operations is the same as the number of possible ways to arrange a shuffled deck of cards.

Even though it arises from a very simple toy, a number such as 10^{68} is very big, even in comparison with numbers like 10^{30} . (The total number of protons and electrons in the visible universe is estimated to be on the order of 10^{79} .) For any realistic garment assembly operation, almost all the possible sequences for the steps would be wildly impractical, but if even a very small fraction of sequences is useful, there will be many such sequences. It is therefore extremely unlikely that any actual sequence that humans have used for sewing a shirt is the best possible one.

Corporations that understand this point recognize that there will always be at least as much scope for improvement through large numbers of small changes in the way things are done in a manufacturing process as through laboratory research. Accordingly, they have promoted worker experimentation and systems for encouraging wide adoption of discoveries made on the line.

Conventional wisdom tends to suggest that all the important ideas come from research labs and that nothing remains to be discovered about mundane activities such as assembly line operation or garment assembly—but conventional wisdom completely missed the potential for "continuous improvement," as implemented in Japanese automobile assembly, just as it missed the potential for high-temperature superconductors made from ceramics. To understand growth, we need to understand not only how big ideas, such as high-temperature superconductors, are discovered and put to use but also how millions of little ideas, such as better ways to assemble shirts, are discovered and put to use. To under-

stand development, we need to understand how both kinds of ideas, but especially the millions of small ones, can be used and produced in a developing country.

Human Capital

It is possible to add human capital to both the Fun Factory and the chemistry set models, and it is instructive to do so. In the Fun Factory model there are only two basic kinds of inputs, human and nonhuman. Physical capital is an aggregate of many different durable nonhuman inputs. The concept of human capital lets us recognize and aggregate different kinds of human inputs. One person may be more productive than another, just as one machine may be more valuable than another. Some of the differences in productivity among workers are the result of investments and are durable, so the analogy with capital is close.

As powerful as this analogy is in labor economics, it sheds little light on the fundamental processes that generate growth. In some proximate sense it must be true that increases in human capital and physical capital explain the increases in value that we have experienced. But what does it really mean to say that the average worker today has several times the effective labor power of a worker in the last century? In a biological sense the population of workers today is virtually identical to the population that existed then.

To see how the same physical objects can be arranged in more valuable ways, consider first an example involving physical capital. The computer that I used to write this paper is about fifty times faster than the one I used just ten years ago, yet it is constructed from just about the same assortment of aluminum, copper, steel, plastic, silicon, and other raw materials. It is manufactured in about the same way and is sold for about the same price.

Now consider human capital. In my brain there are different physical connections between my neurons. These connections store the commands I need to use the new computer and new word-processing software. Just as my new computer is a more productive piece of physical capital, I have more valuable human capital than I did ten years ago.

From the point of view of the chemistry set model, the increase in the value of human and physical capital that is possible using fixed tangible inputs is just like the increase in value that arises when elements such as lanthanum, barium, oxygen, and copper are combined to make a high-temperature superconductor. The knowledge that one mixture or arrangement is more valuable than another is just like the knowledge that changing the layout in a microprocessor increases its processing power, that a different design for a hard disk will increase its storage capacity, or that storing the bit string representing the word-processing software on my hard disk will make the computer more useful. It is also just like the knowledge that reading a software manual rearranges connections in my brain and makes my human capital more valuable. (For later reference, note that on-the-job experience with the software is also extremely important for establishing and reinforcing these connections.)

In some accounting sense the combined increase in the value of human and physical capital explains the increased productivity in word-processing experienced in my office. But to explain productivity growth through these increases in human and physical capital begs the question of where the increased value of the capital originates. The increased value is created by new ideas. Whether it takes the form of a hardware design, software code, or an instruction manual, an idea is used to mix or arrange roughly the same physical ingredients in ways that are more valuable. And in each case, these ideas can be represented as pure pieces of information, as bit strings.

Ideas are therefore the critical input in the production of more valuable human and nonhuman capital. But human capital is also the most important input in the production of new ideas. Physical capital (a computer, for example) is sometimes used in an ancillary way, but a trained person is still the central input in the process of trial and error, experimentation, guessing, hypothesis formation, and articulation that ultimately generates a valuable new idea that can be communicated to and used by others.

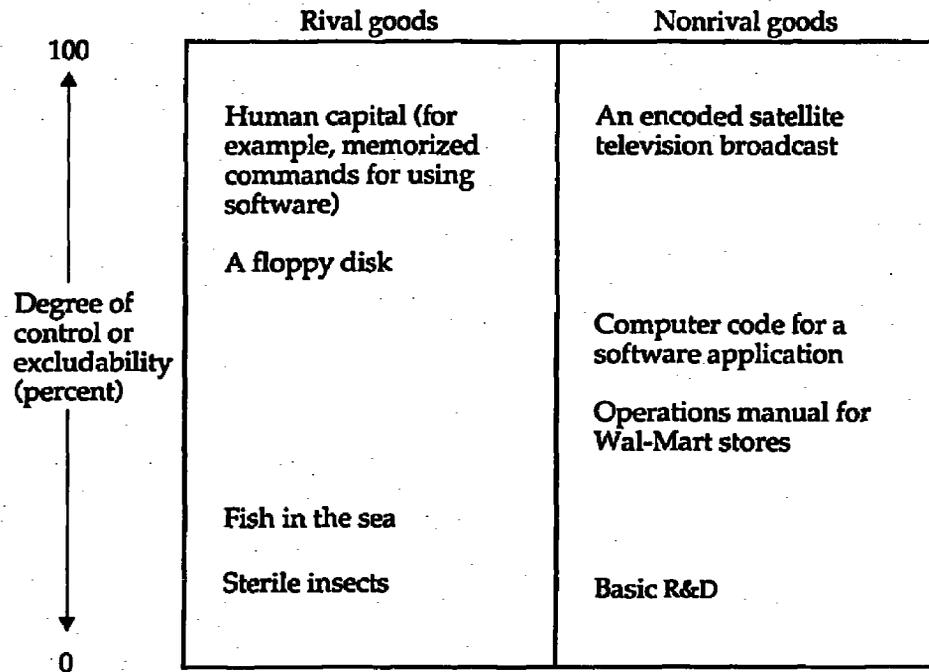
Because human capital and ideas are so closely related as inputs and outputs, it is tempting to aggregate them into a single type of good. After all, structures and equipment are different goods, and they both fit rather well in the category of physical capital. It is important, nevertheless, to distinguish ideas and human capital because they have different fundamental attributes as economic goods, with different implications for economic theory.

Attributes of Economic Goods

Figure 1 illustrates a two-way classification of different types of economic goods that has been useful in the economics of public finance. The vertical axis measures the degree of control or excludability (or appropriability) that is feasible for a good. The left-hand column lists rival goods—goods that are object-like in the sense that they have an opportunity cost. (The label “rival” reflects the fact that you and I are rivals for the use of one of these goods.) The right-hand column lists nonrival goods, which are like bit strings in the sense that everyone can use them at the same time.

As traditionally defined, a private good in public finance is one that lies in the upper-left-hand corner of the figure; it is both rival and fully excludable. According to this definition, human capital is as close to a perfect private good as one can get. There is no way for anyone to take advantage of my ability to remember commands for my word processor without getting my permission; therefore my ability is fully excludable or is subject to complete control by me. And because there is also no way for many people to make use of my ability at the same time, it is a rival good.

Many economists will find these assertions surprising. We often assert that there are spillovers from human capital, such as incomplete control or excludability. We also have an intuition that since I can teach others to do what I

Figure 1. *Economic Attributes of Selected Goods*

can do, what I know is nonrival. After I teach them, we will all be using the same computer commands at the same time.

These imprecise statements are the result of a failure to distinguish among human capital, the ideas that help produce human capital, and the ideas that human capital can produce. My human capital is literally a set of connections between neurons. Converting this rival good into a nonrival good—for example by putting what I know into comprehensible prose on paper—is a time-consuming activity. Once what I know is expressed in words, it can indeed be used by many people (neglecting for now the trivial cost of making a photocopy). As each person reads my words, the nonrival bit string represented by the text is converted back into the rival human capital. Human capital is therefore used to produce ideas, and ideas are used to produce human capital, but human capital and ideas are conceptually distinct goods. They occupy different places in figure 1 and have different implications for economic analysis.

The bit strings, or the how-to manual for using software, would be nonrival, not the connections in my brain. The bit string would also be difficult to control, as every software firm knows. For most of the successful computer programs, someone has written a book that is a substitute for the original manual and can be used by someone who has bootleg copies of the code for the program.

When I bought my software, I purchased a manual, several floppy disks, and the legal right to use the bit string representing the computer code. I used the manual to produce some human capital that is easy for me to control. The floppy disks themselves are somewhat harder to control because they could be stolen. For this reason they are listed farther down the column of rival goods.

The software code is even harder to control because it can be taken without the knowledge of the owner. This feature tends to make all nonrival goods less excludable and more difficult to control.

Because nonrivalry and limited control are correlated in practice, many informal discussions of ideas do not distinguish carefully between the two concepts, but for economic analysis the difference is important. Private firms survive in the business of writing and selling very costly computer code; therefore they must be able to assert some control over it. Software code is, however, the quintessential nonrival good. Most ideas of economic significance lie in the upper right portion of figure 1. They are nonrival and at least partially excludable.

Such goods are very different from the goods in the lower left side of the figure, the ones that are rival but impossible to control. Fish in the sea are the most familiar example for economists. Only one person can eat a fish. Sterile insects used to prevent the reproduction of agricultural pests are another example. A sterile male insect can mate with unsuspecting fertile insects in one valley or the other, but not in both at the same time. Because it is so difficult to establish property rights over these kinds of goods, market outcomes are inefficient. We overfish, and we underprovide sterile insects. Our understanding of the policy implications of this part of the figure is very clear. Property rights should be provided where possible—for example, tradable rights to catch fish. Where this is not possible, as in the case of the insects, the government must supply the good and use its tax power to cover the costs.

Although the sterile insects seem to be examples of public goods in the everyday use of the term (things that are provided by the government), the formal definition of a public good used by economists is one that lies in the lower right of figure 1; it is a good that is both nonexcludable and nonrival. Basic scientific research is perhaps the best example of a pure public good in this sense. Our understanding of how vaccines prevent disease and how drinking water mixed with electrolytes can prevent death from diarrhea both come from public support for basic research.

In the left-hand column (the rival goods), the policy implication is to move up the column—to create property rights whenever possible and to make them as strong as possible. For nonrival goods, this conclusion does not follow. Property rights to a nonrival good always imply a market price for the good that is higher than its opportunity cost because the opportunity cost is zero. As the usual analysis of patents suggests, strong property rights and the resulting monopoly profits are desirable because they create an incentive to discover new ideas, but they are undesirable because the difference between price and marginal cost creates distortions.

Depending on the type of good involved, we craft different solutions to this conflict between the two conditions required for efficiency. In some cases it is relatively clear which directions for new discoveries are worth pursuing, and the social cost of setting a price higher than marginal cost is very high. In these cases—for example in public health—the government pays for the research and

gives the results away. In other cases the government has no mechanism for deciding what to produce, and the social costs of prices higher than marginal cost seem trivial—for example, in the production of popular music (which these days is literally a bit string on a compact disc). In areas such as biotechnology and software it is unclear exactly how we should proceed. We do not know whether a gene fragment or a programming concept such as overlapping windows should be protected by law or, if so, whether the law of copyrights, patents, or trade secrets should be used. (For additional discussion of the policy dilemma presented by nonrival goods and of the different institutional arrangements that have emerged for dealing with it, see the paper by Paul David in this volume and the references cited therein.)

For thinking about growth and development, the important implications of figure 1 are as follows. First, the distinction between objects and ideas (between rival and nonrival goods) is far more important than the concept of excludability or control or the related concepts of spillovers and externalities. An approach to economic policymaking that neglects nonrival goods will miss most of the interesting issues.

Second, nonrival goods have the unique feature that their value depends on the size of the market in which they can be used. This is obvious for an idea such as the vaccine that is treated as a pure public good but is equally true for nonrival goods such as software or even ideas about how to run a discount store. If Bill Gates could have sold software only in Washington state, or if Sam Walton could have opened discount stores only in Arkansas, they would have been millionaires, at best, instead of the billionaires they became. This is the other side of the gains from trade that ideas can generate if they are widely used.

Third, any discussion of an economy in which nonrival goods such as software are privately provided must allow for departures from price-taking. The textbook description of perfect price-taking competition is logically inconsistent with the private provision of nonrival goods. Even in a case in which a bit string (for example, computer code) must be sold together with an object (for example, a floppy disk), monopoly pricing must still apply. The package consisting of the code and the disk will have a positive marginal cost equal to the cost of the disk, but the package must still sell for a price higher than marginal cost, or no one would be able to earn a return on effort devoted to writing software. Casual empiricism suggests that software sells for a price that is 10 to 100 times marginal cost.

Fourth, and finally, there is no hope that a decentralized equilibrium in which new ideas are discovered will be first-best Pareto optimal. The usual justification for the welfare theorems in terms of price-taking does not apply. No comparable justification on the basis of the Coase theorem will work unless it replicates the essence of a government—a decisionmaker with powers of coercion over everyone else in the economy.

This last point is clear only in a general equilibrium setting, which perhaps explains why it has not been adequately emphasized. In partial equilibrium

analysis, price-setting by a monopolist is consistent with Pareto optimality if control or the ability to contract is strong enough. For example, we know that a perfectly discriminating monopolist or a monopolist who charges a two-part tariff can produce the efficient level of output in a partial equilibrium model.

This partial equilibrium analysis simply does not extend to the economy as a whole. It violates a fundamental adding-up condition. Output for any particular activity can be written as $y = f(k, h; a)$, where h and k are lists of all the different human and nonhuman rival goods that are used in production and a is the idea or knowledge that makes this activity possible. For fixed a , we can think of the production of y in terms of the Fun Factory metaphor. By a standard replication argument, f is homogeneous of degree 1 in the rival goods that must be replicated to make a copy of the existing activity. But Euler's theorem tells us that h and k will exhaust the total value of output in this activity if competition is used to allocate these goods and they are paid their value marginal products. Nothing will be left to pay for a .

What is true for one activity is equally true for the economy as a whole. If aggregate output is written as $Y = F(H, K; A)$, F will be homogeneous of degree 1 in the human and nonhuman rival inputs. It is therefore impossible to use prices to allocate H and K between different activities and at the same time provide the incentives for discovering ideas, A . There is not enough income in the economy as a whole to go around. (See Romer 1990a for an elaboration of this point.)

II. TWO ISLANDS

The challenge now is to use the analytical framework outlined above to discuss the different economic activities taking place in Mauritius and in Taiwan (China). In both economies government policy is attuned to the role of ideas. In Mauritius policy changes in the 1970s and 1980s made it attractive for entrepreneurs from Hong Kong to put their ideas to use there. As a result, the citizens of Mauritius achieved large increases in income and employment. A more clear-cut case of the gains from trade could hardly be imagined. In Taiwan (China) ideas from the rest of the world are not merely put to work with domestic labor. To a much greater extent than in Mauritius, the government intervened in market exchange in its attempts to encourage the domestic production and exploitation of ideas that can earn a return on world markets.

Using Ideas: Mauritius

Table 1 reports some basic economic and social indicators for Mauritius over the interval from 1960 to 1985. For comparison, data are also presented for two economies with similar ethnic and economic backgrounds, India and Sri Lanka, and for Taiwan (China), which is discussed below. In Mauritius the share of investment in gross domestic product (GDP) is no better than that in the other

Table 1. Economic Performance, 1960-88
(percent, except as otherwise specified)

<i>Indicator</i>	<i>Mauritius</i>	<i>India</i>	<i>Sri Lanka</i>	<i>Taiwan (China)</i>
Rate of growth of income per capita, 1960-88	2.8	0.9	1.3	6.4
Rate of growth of population, 1960-88	1.7	2.2	1.9	2.2
Income per capita, 1960 (1985 U.S. dollars)	2,000	600	1,400	950
Share of investment in GDP, 1960-88	12	17	21	23
Literacy rate, 1960	60	28	75	54
Primary school enrollment rate, 1960	98	61	95	96
Average years of schooling, labor force, 1986	4.5	1.9	6.2	8.4

Sources: For income, investment, and population, Heston and Summers (1991); for literacy and schooling, Levine and Renelt (1992).

three economies and in the Heston-Summers data used here, it is substantially lower.¹ Investment in human capital, while better than in India, is not outstanding. Income per capita at the beginning of the sample period was higher than in the other three economies because of the high income earned by the wealthy owners of sugar plantations, but the distribution of income was very uneven.

When Mauritius was preparing for independence from Great Britain in the 1960s, prospects for development did not seem promising. The economy was dependent on sugar exports for 99 percent of its exports. James Meade, commissioned by the British government to comment on economic policy, entitled his 1961 report "Mauritius: A Case Study in Malthusian Economics" and devoted it to a discussion of how to cope with a real wage that would inevitably fall as population on the island grew. Young people with higher levels of education were encouraged to emigrate, but it was feared that not enough would leave. A local import-substituting manufacturing sector existed only because of protective tariffs.²

Since independence in 1968, the political situation has been fluid, almost to the point of instability. Up to the present, corruption, fraud, and drug trafficking have been a continuing source of government scandals. Political parties are organized along ethnic and religious (primarily Hindu and Muslim) divisions. There has been no majority government since independence. The emergence of a

1. According to the national income accounts data of the World Bank (see Levine and Renelt 1992), Mauritius, India, and Sri Lanka all invest about 20 percent of GDP. The Heston-Summers numbers adjust nominal investment spending to correct for variation in relative cost of capital goods in different countries. In principle, the real investment numbers from Heston and Summers are more relevant, but the downward adjustment for investment in Mauritius seems implausibly large.

2. The account that follows is drawn from World Bank data and from published studies by Gulhati and Nallari (1990) and Bowman (1991).

left-leaning party and the strikes that it organized in 1971 provoked an assassination attempt instigated by members of a rival party. This led to public riots and a state of emergency that lasted until 1976. Unions were suspended, and union and party leaders were arrested.

The average growth rate for per capita income of nearly 3 percent a year during 1960–88 was achieved despite the terms of trade shocks experienced by all primary goods exporters in the 1970s. Sugar prices soared in the early part of the decade, then fell dramatically in the second half just as oil prices went up. This led to predictable macroeconomic difficulties (budget deficits, inflation, and a balance of payments crisis), followed by a period of austerity and adjustment that was associated with zero growth during the five years from 1978 to 1982. Most of the growth that took place came in two brief spurts, one in the first part of the 1970s and the other in the latter part of the 1980s.

Despite all this, Mauritius stands out as a significant and surprising success story when compared, for example, with India and Sri Lanka. The only obvious candidate for explaining the success of Mauritius is the island's policy of supporting an export-processing zone (EPZ), which made investment attractive to foreigners. The EPZ was an administrative arrangement; it involved no geographic restrictions and no special investment in infrastructure. The main policies in this arrangement were unrestricted, tariff-free imports of machinery and materials, no restrictions on ownership or repatriation of profits, a ten-year income tax holiday for foreign investors, a policy of centralized government wage-setting, and an implicit assurance that labor unrest would be suppressed and wage increases would be moderate.

From the inception of the arrangement in 1971, employment in the EPZ grew to 17,000 workers by 1978, a significant number on an island with a total population of about 1 million. Over this same interval GDP per capita grew at 9 percent a year, fueled partly by the EPZ and partly by favorable sugar prices and harvests. During the macroeconomic difficulties between 1978 and 1982 EPZ employment growth slowed and almost ceased. Once the government had completed its adjustment process, growth resumed with a vengeance. Income and corporate tax rates were halved in 1983 (from about 70 to about 35 percent). Both domestic and foreign investment in the EPZ increased sharply. In 1982 the unemployment rate stood at 22 percent and total employment in the EPZ at about 20,000 workers. By 1988 the economy had essentially reached full employment through the addition of 70,000 jobs in the EPZ. In 1970 agriculture employed 60,000 people and the EPZ did not exist. In 1990 agriculture employed 46,000 workers and the EPZ employed 90,000, about one-third of all workers on the island. Jobs added in the EPZ accounted for two-thirds of the total increase in employment between 1970 and 1990.

Manufacturing in the EPZ is concentrated almost exclusively in garment production and was developed almost entirely because of the participation of entrepreneurs from Hong Kong who were drawn to Mauritius because of contacts with the small ethnic Chinese population on the island. These entrepreneurs

were motivated by the prospect of lower wages, a location free from the threat of expropriation by the mainland Chinese government, and a country of origin that would not be subject to quota limits set by the United States and the European Economic Community (EEC). Investors from Hong Kong now hold nearly three-quarters of the foreign investment in the EPZ.

According to one observer (quoted in Bowman 1991), "textile entrepreneurs flocked to the island bearing sewing machines" because of the EPZ arrangements. In the Fun Factory model, this inflow of capital must explain the increase in employment and output that subsequently took place. A little reflection suggests that the entrepreneurs brought much more than physical capital, for Mauritius had long enjoyed special trade status with the EEC, and sewing machines could always have been purchased on the open market. Nor were foreigners essential as a source of the savings needed to finance investment in physical capital; domestic savings ultimately accounted for a substantial fraction of total investment in the EPZ.

The entrepreneurs did bring a crucial array of ideas about the textile and garment business, including ideas on the specific kind of equipment to use, how to manage a small factory, how to manage relations with textile importers in the industrial countries, how to successfully exploit loopholes in quota limits, and hundreds of other ideas about running a modern garment assembly operation, such as knowledge of the sequence to use in sewing a shirt.

In a model with no ideas or in which ideas are already available throughout the world, it is difficult to explain the experience of Mauritius. If investment in physical capital is the cause of growth, one must understand why domestic investment did not take place prior to the arrival of the foreigners but did subsequently, and why it took place only in garment production. If a scarcity of human capital explains why people on Mauritius were poor, this surely could not have changed much in the interval between 1972 and 1979 or between 1982 and 1989. Moreover, if human capital were all that mattered, why did Mauritius do so much better than Sri Lanka, which had a much better record in education?

In a world in which ideas are under private control, events in Mauritius are easy to understand. Suppose that agricultural output is a constant returns-to-scale function of land, T , labor, L , and capital, K : $F(T, K, L)$ (F for farm). We can summarize garment output in the form $G(K, L; A)$, where A takes on the values 0 or 1 (G for garment). Output of garments is zero if $A = 0$ because no one knows how to run a garment operation. If an entrepreneur does know, $A = 1$ and garments are a constant returns-to-scale function of sewing machines, K , and labor, L .

When $A = 0$, wages are equal to the marginal product of labor in agriculture. For large L , wages can be very low. Sewing machines are freely available for sale at the price p_K , but none are imported because no one knows how to put them to use. Farm equipment is imported, funded either by domestic savings or by foreign investment. Investment in Mauritius earns the same rate of return as in

the rest of the world. Wages were low not because of any restrictions on flows of financial capital or inadequacy of domestic capital investment but because the idea, A , needed for garment production was not in use there.

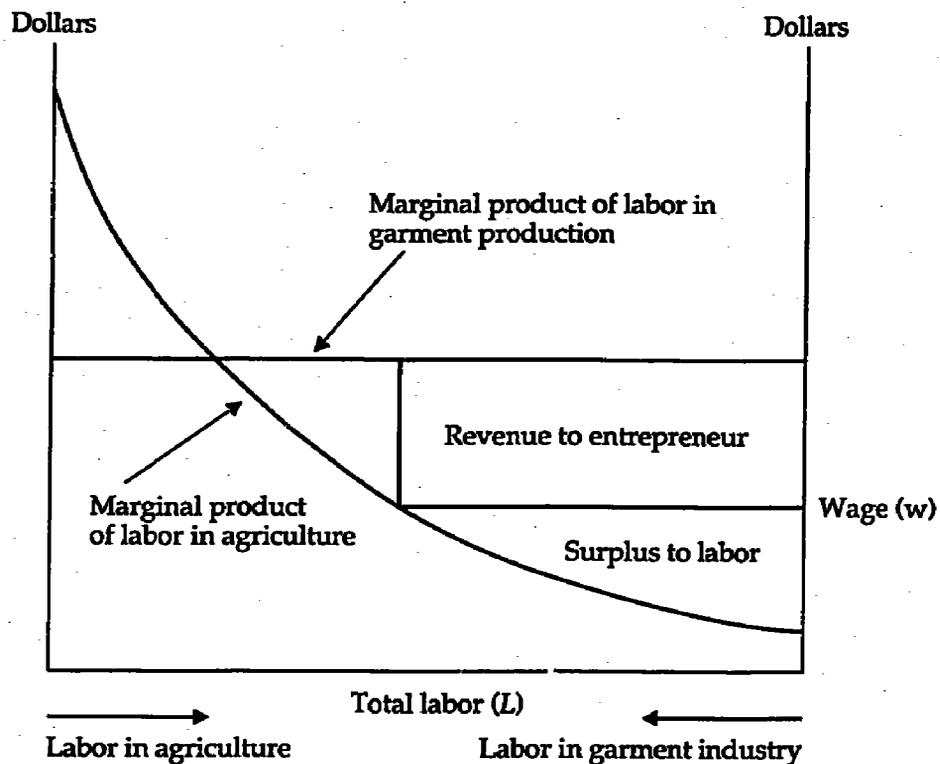
Now suppose that a single textile entrepreneur learns of the low wages on Mauritius, brings knowledge of A , and sets up shop. Sewing machines will now be imported. Income and employment will increase. An examination of many such cases could lead one to conclude that investment, particularly investment in machinery, causes rapid economic growth. (See DeLong and Summers 1991 for cross-country evidence of the correlation between investment in machinery and growth.) What detailed knowledge of the history of Mauritius shows is that investment was the proximate, but not the fundamental, cause of the growth that took place. It was the knowledge, A , brought by the foreigners that caused both the investment and growth.

To represent the private returns for putting an idea to use, let $\Pi(L, p_G, p_K)$ denote the restricted profit function for a single entrepreneur who employs L units of labor on Mauritius and faces a price for sewing machines p_K . Since $G(K, L; A)$ is homogeneous of degree one in K and L , the profit function Π defined by

$$\Pi(L, p_G, p_K) = \max_K p_G G(K, L; 1) - p_K K$$

will be linear in L . The entrepreneur is a monopsonist in the labor market in Mauritius, as depicted in figure 2. The excess supply curve for labor faced by the

Figure 2. Labor Market with a Monopsonist Entrepreneur



entrepreneur (which is the same as the marginal productivity curve in agriculture) is upward-sloping. The marginal product of labor for the entrepreneur is constant. The single entrepreneur will employ labor up to the point where wages are equal to w in the figure.

Workers in Mauritius are exploited in the sense that w is strictly less than the marginal product of labor. Yet they also receive a pure surplus or windfall gain from the entry of the entrepreneur (the area labeled "surplus to labor" in the figure). The entrepreneur collects the monopsony rectangle denoted "revenue to entrepreneur." Workers in agriculture also gain at the expense of landholders.

The usual analysis of government policy when faced with this kind of monopsonist recommends a minimum wage. By setting the minimum wage at close to the marginal product of labor in garment assembly, the government can drive the monopsony profits to zero, encourage the efficient level of employment, and capture the largest possible surplus for workers.

The problem with this analysis is that it neglects the possibility that the textile entrepreneur will not come if they are not offered a large enough profit. Coming to Mauritius is a costly activity, and these entrepreneurs have alternative uses for their time and energy. Welfare on the island is increasing when the minimum wage is raised from the monopsony level, but only up to the point at which the ex post monopoly revenue captured by the textile entrepreneur is just enough to offset the initial cost of coming there. Beyond this point, an increase in the minimum wage makes welfare on the island drop precipitously, back to the Malthusian equilibrium anticipated by Meade.

Extracting surplus from foreigners can therefore be a risky proposition, especially if policy decisions are made in the absence of knowledge about the potential entrants and their costs and alternative opportunities. The dangers of setting too high a value for the minimum wage apply equally to other costs imposed on foreigners—costs such as burdensome customs procedures, high required domestic equity participation, high explicit taxes, high implicit taxes collected as bribes, ex post extraction of rents by organized labor, or poor provision of services by government agencies. An authoritative central government that can centralize the decisionmaking process and lower the total rent extracted from potential investors can generate substantial benefits for the nation as a whole if it changes the EPZ from one that has not attracted any substantial foreign participation (and there are many of these in the world) to one that is attractive. (For a more detailed description of this problem of decentralized extraction of rents, see the analysis of corruption by Shleifer and Vishny 1992.)

The analysis suggests not only that attempts by the government to extract rents may be strongly counterproductive but also that the government must be strong enough to prevent this kind of extraction by other actors in the economy. The government might even be able to improve welfare by subsidizing the entry of the entrepreneur. If the fixed costs for the entrepreneur of coming to the island are greater than the monopsony profit but smaller than the sum of this profit plus the surplus to workers, the government could offer part of the surplus

to induce the entrepreneur to come. In the more realistic case in which a large number of firms, rather than a single monopolist, contemplate entry, the government might even facilitate collusion among the foreign firms to preserve their profits after entry.

Viewed in this way, the timing of the investment in Mauritius is easier to understand. The state of emergency, the banning of unions, and the imposition of centralized government wage-setting were crucial steps leading to the first wave of investment by foreigners. Resolution of government budget difficulties, a devaluation that reduced real wages for workers, and cuts in effective tax rates were the decisive actions in bringing about the second wave of investment. It is also relevant that while employment and income did increase substantially in Mauritius, the government made good on its pledge of wage moderation. Real wages have been allowed to increase very little in garment assembly. Now that full employment has been reached, there are signs that the government is moving away from the previous policy of wage moderation. How it will handle the next phase of development is not yet clear.

What is worrisome about a strategy of encouraging foreigners to employ low-skilled, low-cost labor is that the equilibrium wage for unskilled labor may be very low—even lower than the wages now earned on Mauritius. In the industrial countries during most of this century, A and H have been high and growing in relation to the small quantity of L present in these countries. Because of restrictions imposed by poor countries, labor in the rest of the world was segregated from labor in industrial economies. The restrictions were no doubt encouraged by domestic firms that wanted protection from foreign competition in output markets and in the labor market but may also have been fostered by a fear of exploitation. As a result, wages for L kept pace with growth in A and H in industrial countries.

Now many poor countries understand the advantages that come from removing these restrictions, and a very large quantity of labor from developing nations is on the verge of entering the worldwide market. For example, there are more than 125 EPZs in developing countries, with most of the increase in numbers coming in the last half of the 1980s. Mainland China and even India are now opening to the rest of the world. In addition, improvements in transport and communication have reduced the costs faced by an entrepreneur who wants to move production elsewhere in the world. The first countries to integrate their labor markets with markets in industrial countries experienced relatively large wage gains, but as more labor from other developing countries enters the market, wages will have to fall.

There is already evidence of this process in the pressure on wages for unskilled labor in industrial countries, especially the United States. Skill differentials are increasing as real wages for unskilled labor remain stagnant or fall. In industries where production in low-wage countries can most easily be undertaken (textiles, for example), there is already in place an extensive system of quotas designed to protect wages in industrial countries. Countries such as Mauritius can therefore

expect to be squeezed from two sides. There will be more competition from other countries with low labor costs and increased trade barriers in industrial countries.

In the very long run, the fall in wages will be partially offset by increases in worldwide stocks of K and H . Suppose that the ratio of these stocks to the world stock of labor L reaches the ratio that now prevails in industrial countries. A striking implication of the analysis of ideas as economic goods is that worldwide integration of markets will permanently increase wages for H in relation to wages for L . As emphasized in the discussion of ideas, the value of an idea increases with the size of the market. Because the production of ideas is human-capital-intensive, increased worldwide economic integration will drive up the returns to human capital in relation to returns to labor. There are large gains from trade that arise from worldwide integration. Unfortunately for poor people and poor countries, the gains will be captured disproportionately by the most highly skilled workers. (See Romer 1990b or Grossman and Helpman 1992 for formal models that illustrate this point.)

For a small economy, investing in schooling may not by itself be enough for it to become involved in the production of ideas, where the high returns to human capital lie. The production of ideas requires human capital, but it also requires access to existing ideas. A country like Sri Lanka that invests heavily in education but remains isolated from all the economically important ideas that are in use in industrial countries has no hope of ever becoming a player in the global production of ideas. Recall that before Mauritius opened the EPZ, its policy was to encourage migration of educated youths, for whom there were no prospects on the island.

A more worrisome possibility is that the relatively open strategy toward the rest of the world pursued by Mauritius, or even a strategy of totally free trade, may not be sufficient to bring local human capital into use in the production of ideas. It is easier to use ideas in a small country than it is to begin producing ideas. It takes a relatively narrow range of ideas to open up a particular activity such as a garment factory. Because of constant returns to scale, the market incentives will then be to increase production of the associated good without incurring the cost of starting another activity. In Mauritius only one idea has been put to work because almost all of the EPZ output is in garments. Prospects for developing new products—that is, for producing ideas—are therefore quite weak.

Taiwan (China)

Authorities in Taiwan (China) used a wide variety of approaches to encourage the use of ideas there, with the explicit intention of shifting to the domestic control and production of ideas. Generally speaking, the government moved from a period of import substitution during the 1950s and 1960s toward an export-oriented strategy in the 1970s and 1980s. More recently it has placed increased emphasis on human capital and research subsidies similar to those

used in industrial economies, but in applications with specific commercial goals. The pattern of intervention in several different industries illustrates the eclectic and flexible approach used by the government to achieve its aims.³

In the early 1950s the government gave special attention to the textile industry, which got its start with mainlanders who came to Taiwan (China) with their knowledge and looms. Early policy supports included tariffs and quantitative limits on imports of yarn and finished products, restrictions on entry for new firms, and controlled access to raw materials. In the early years the government supplied raw cotton to spinning mills and bought all the finished yarn, which it then supplied to firms with looms. Later the government used cheap credit to encourage firms to operate at larger scales and to integrate vertically, relying on the local office of a U.S. engineering firm to evaluate individual requests for credit. Exports grew rapidly from 1952 to 1958, the year when the island became a net exporter of textiles. After the exchange rate regime was changed in 1958 to encourage exports, textile exports to the United States grew so rapidly that quota limits were imposed in 1961.

To diversify the textile industry away from cotton, the government encouraged a move into synthetic fibers, acting primarily as an intermediary between domestic firms and foreign firms with advanced technologies. By 1954 the domestic chemical industry could produce most of the intermediate inputs needed to make rayon. With the help of advisers from the United States, the government brought together a U.S. firm and several local textile firms and supervised the negotiations leading to the creation of a joint venture that began production in 1957. The U.S. firm provided the planning, the equipment, and worker training. In 1962 this joint venture, in collaboration with a state financing agency, created another joint venture to make nylon, this time relying on technology from a Japanese firm. As private firms began to enter the synthetic fibers industry in the late 1960s and 1970s, they relied increasingly on licensing instead of joint venture agreements. The government continued to assist in finding foreign partners with technology to share and in negotiating the terms of the technology agreements, over which it retained authority for final approval. By 1981 Taiwan (China) was the fourth biggest producer of synthetic fibers in the world.

The electronics industry started like the textile industry, building from a domestic base in radio assembly that was protected by restrictions on imports. As in synthetic fibers, radio assembly was aided in the early years by a technology agreement with a foreign firm. Then, in the 1960s, the government set up an EPZ designed to encourage electronics assembly by foreign firms. By 1965 twenty-four U.S. firms had made arrangements for production in Taiwan (China). The industry developed around a few large foreign assemblers and many small domestic suppliers of components. A government-supported electronics working group assisted in marketing, training of personnel, and product

3. This account is drawn from Wade (1990).

expositions. In 1973 the government opened the Industrial Technology Research Institute (ITRI), which supported advanced training programs for engineers. The Electronics Research and Service Organization (ERSO), which operates under ITRI, supplied the first basic input-output system (BIOS) used in Taiwanese clones of IBM personal computers. More recently ERSO supported a move into semiconductor design and fabrication, opening a model shop for wafer fabrication and negotiating a technology transfer agreement with a U.S. firm. The government has begun to aggressively court Taiwanese nationals who had been trained and were employed in electronics and other high-technology fields in the United States. By 1968 electronics was second only to textiles in total exports, and in 1984 it became number one.

Over time the government in Taiwan (China) has increasingly emphasized exports and has moved away from the traditional import-substituting model of development. It is largely for this reason that Taiwan is regarded as an open, or at least an outward-oriented economy. It has nevertheless continued to employ many restrictions designed to protect and develop domestic industry. Proposals for foreign investment outside of an EPZ must be approved by the government and are often subject to conditions that limit sales in the domestic market, mandate local-content requirements, or set export targets.

Three cases are indicative of the general pattern. When the Singer Sewing Machine Company asked permission to open a manufacturing plant in the 1960s, the government required that within one year the company purchase more than 80 percent of its parts domestically. Singer did facilitate technology transfer and helped upgrade the domestic components industry but was not able to meet its goal by the end of the first year. The agreement with the government was then renegotiated. When the National Distiller and Chemical Corporation from the United States proposed the construction of a polyethylene plant, it was given a five-year tax holiday, permitted to sell in a domestic market that was protected from imports for three years, and allowed unlimited repatriation of profits. In return, the government insisted that after five years the firm would convert to a joint venture, with half of the equity held by residents of Taiwan (China). When Proctor and Gamble opened a plant in the 1970s, it was required to export 50 percent of its output.

Export requirements were initially imposed to generate foreign exchange but were increasingly used to guarantee that a foreign company would bring to Taiwan (China) a technology sophisticated enough to compete in world markets. It is anticipated that through the company's purchases from suppliers, through the experience of workers, and sometimes through mandated sales of equity, aspects of this technology will diffuse to Taiwanese citizens. In general, negotiations with foreign firms are characterized by a wide latitude for discretion and by agreements that are customized to fit the circumstances (and bargaining strength) of each foreign investor. Aggressive targets are set but are subject to renegotiation on the basis of new information.

Explicit tariff barriers have fallen, but nontariff barriers continue to be used as indirect subsidies for domestic firms. Imports of foreign equipment and intermediate inputs are subject to a complicated system of discretionary administrative control designed to support domestic suppliers of a good. One Taiwanese study cited by Wade estimates that in 1984 about half of all imports by value were subject to some form of nontariff barrier. In addition, government officials monitor detailed reports of imported inputs used by foreign firms, watching for cases in which they can arrange for a domestic supplier to provide the imported goods.

A key characteristic of government intervention in Taiwan (China) is the freedom and authority with which government officials can act. Three examples tell the story. The chief economic planner for the government once ordered the public destruction of 20,000 low-quality domestic light bulbs and threatened to liberalize imports if quality did not improve. The government approved the arrangement with Singer Sewing Machines despite the opposition of domestic sewing machine manufacturers, who ultimately benefited from the improved quality of the parts industry. In 1982 the government granted a two-year import ban on videocassette recorders (VCRs) to protect two domestic manufacturers from Japanese competitors. After one year the government gave a public warning that it would bring in a foreign firm in a joint venture if the prices and technology of the domestic firms did not achieve world standards by the end of the two-year ban. The protected firms did not live up to this standard, and eighteen months after the initial ban the government approved a joint VCR production venture between Sony and a new Taiwanese firm.

Some economists see the success of Taiwan (China) as a vindication of *laissez-faire*. Others attribute it to an explicit industrial policy that steered the economy into a sequence of important activities. A third interpretation is that the particular industrial activities undertaken in Taiwan were determined primarily by market forces and followed the general pattern observed in other countries at similar stages of development (Pack 1992). According to this view, what mattered was not the government's steering but its use of the accelerator. Taiwan moved very rapidly through the stages of industrial development, in large part because of its success in gaining access to and control of foreign technology. This is reflected in a high rate of measured productivity growth, which stands in contrast to the low productivity growth recorded in rapidly developing economies where growth is driven by extremely high rates of capital accumulation. (See Young 1992 for a discussion of Singapore and Pack 1988 or in this volume for a general description of this phenomenon.)

The description in the previous section of the gains from using ideas suggested reasons why government intervention might be useful. Ideas that are privately controlled create more economic value when they are introduced into an economy than the holder of the idea can extract. A description of the opportunities for producing ideas reinforces this point.

The discussion of production on Mauritius focused on a single manufacturing activity and did not consider the production of ideas at all. Suppose that there are many different manufacturing activities which can be indexed by j , and suppose that each activity requires its own idea A_j :

$$Y = F(T, K, L) + \sum_j G_j(K_j, H_j, L_j; A_j).$$

Suppose as well that search, the production of new ideas, depends on human capital, H_A , used exclusively in search, and on the entire list (A_1, A_2, \dots) of ideas that are in use within a specified geographic area. Suppose also that success in search also depends on the amount of human capital that is used in production in each of the manufacturing activities, (H_1, H_2, \dots) through a process of "discovering" by doing:

$$\dot{A} = S[H_A, (A_1, A_2, \dots), (H_1, H_2, \dots)].$$

Finally, recall the example of learning how to use computer software by using it. To capture this, we can write human capital acquisition as a function of the use of specialized human capital on the job in a conventional learning-by-doing specification:

$$\dot{H}_j = \mu H_j.$$

Note that this description of the accumulation of new ideas and new human capital relies on two different kinds of joint product assumptions. Someone with human capital of type j who is employed in activity j produces manufactured good j , produces more human capital of type j , and (occasionally) makes new discoveries of the "better ways to sew a shirt" variety.⁴

In this complete model it is clear that when a new idea comes into an economy, it helps domestic citizens in three ways. First, as noted above, the new idea creates a surplus for unskilled labor by making possible a new productive activity in which it can be employed. Second, it creates opportunities and production of specialized human capital through on-the-job training. In effect, it creates a new opportunity for investing in human capital, and the returns from this investment may be very high. Finally, it increases the productivity of research and discovery. If enough such ideas are present, this may tip the economy from a no-

4. See Romer (1990b) for a discussion of a basic model with a manufacturing sector and a separate research sector in which search builds on existing ideas. Grossman and Helpman (1992, ch. 8) cover the case assumed here in which only ideas that are available locally can be used in research. They derive the result that government intervention in support of research may be necessary to move a geographic region out of a corner equilibrium with no research. Glaeser and others (1991) describe the theoretical motivation for allowing a diverse set of ideas to enter as inputs in the search process for new ideas and offer evidence from different cities in the United States that supports the importance of this effect. Empirical support for the idea that human capital is acquired on the job in developing countries can be found in the survey of experience in EPZs by Rhee, Katterbach, and White (1990). For example, they find that many managers of domestic firms operating in the EPZ in the Dominican Republic were once employees of foreign-owned firms in the EPZ.

discovery equilibrium to one in which many new products and processes are discovered.

It is possible that the firm that brings an idea to an economy can capture part of these additional gains. For example, it may be able to get educated employees to work at a lower wage in exchange for the opportunity to acquire human capital on the job. Even if this is true, the firm is likely to capture less than 100 percent of the gains—perhaps substantially less. And economists widely agree that a firm captures very little of the gains that its ideas create for others engaged in research.

For all these reasons, the social return to having an idea like the technology for polyethylene in use in an economy may be substantially greater than the private gain that foreign holders can capture. Thus, subsidies from the government to attract these ideas may be essential to get them to come. (Recall the three-year protection from competing imports offered to the polyethylene plant.) The total gain for domestic citizens, net of the cost of the subsidy, may still be very large.

III. POLICY AND POLITICS

In the United States the Congress gives research grants directly to some colleges and universities. Success in attracting these grants depends on the effectiveness of the lobbying firm employed by the school and on the seniority and committee assignments of local senators and representatives. Because neither the quality of previous research nor the quality of the proposed research plays any important role in allocating these grants, observers have concluded that they encourage rent-seeking but not good science.

In many countries direct and indirect subsidies are granted to firms through measures such as tariff and nontariff restrictions on competing imports. Success in attracting these subsidies depends on the political power of the beneficiaries. Because neither success in introducing new ideas into an economy nor success in reaching worldwide standards for price and quality play any role in the allocation of these subsidies, most observers have concluded that these interventions encourage rent-seeking but not economic development.

One might conclude from the U.S. experience that the government should never give grants to support research. One might even follow the fashion in economics and construct a model to show that the market provides the optimal level of research. Neither conclusion is warranted. We know that there is a valid economic justification for supporting basic scientific research at universities. We also know that some institutional arrangements for allocating grants (mandates from Congress) do not achieve the intended goal but that other arrangements (peer review) do a reasonably good job.

Despite frequent protestations to the contrary, the economic case for intervention to encourage the use of ideas in developing economies is at least as strong as the economic case for supporting basic research in industrial economies. The problematic assertion is that it is possible to create institutions analogous to peer

review that can undertake beneficial economic intervention. The required institutions may not exist, and it may not be possible to create them in a given country. Laissez-faire may be a second-best solution, but we must recognize that deciding whether this is the case depends at least as much on political and sociological analysis as on economic analysis.

Together with the economic analysis outlined above, my amateur political and sociological analysis leads me to the following general conclusions. First, there is much evidence suggesting that the specific arrangements used in Taiwan (China) cannot achieve their goals in a modern democracy or in most open political systems. (For a good summary of the evidence from political science showing that a Taiwanese system of discretionary bureaucratic decisionmaking is not feasible in the United States, see Wilson 1989.) To succeed, these arrangements must be part of a larger political system that can support wide latitude for discretion on the part of a strong, authoritarian government that is willing and able to override parochial interests. In addition, these arrangements require a configuration of bureaucratic competence and ruthless dedication to national economic success that is relatively rare and may be impossible to sustain.

Contemporary evidence suggests that interventionist institutions may not continue to function well even in economies such as Japan, Taiwan (China), Singapore, and the Republic of Korea, where they seem until now to have been a success. In these countries the experience with extensive intervention extends only through the working career of one—perhaps uniquely dedicated—generation of bureaucrats working in an unusual political environment. In Korea the powers of the state, traditionally used to support the large corporations, were turned against the Hyundai group when its head became a candidate for the presidency. In Japan concern with corruption is growing at the same time as the system for allocating political power shows signs of being more openly contested. Neither development bodes well for the long-run viability of a system based on honest, independent bureaucrats with extensive discretionary power.

Until other politically viable institutions for fostering development can be discovered, the one safe piece of advice to offer developing countries is that integration with world markets offers large potential gains. The gains from using someone else's ideas come from a source that is different from the classical gains from trade. The division of the gains may not correspond to intuitive notions of fairness, but they can be very large and very important nonetheless. This is absolutely clear for a small country such as Mauritius, but it is equally clear for the very large and rapidly growing economy on mainland China. The gains it receives from interaction with Hong Kong and Taiwan (China) far outweigh the small and risky gains that might be achieved through a more tightly controlled industrial policy.

The other safe counsel is to increase savings and schooling, but both of these activities require a reduction in current consumption that may be very costly for the poorest countries. Formal education also works with a long lag. In contrast,

openness to investments by foreigners bearing ideas costs nothing, except perhaps a bit of national pride. On-the-job training can in many cases be even more effective than classroom education in developing human capital. Once gains in income from direct foreign investment are forthcoming, high savings and large investments in schooling are easier to finance and can be used to lay the groundwork for further gains in income.

Beyond this, cautious attempts to encourage the development of local expertise may be valuable. For example, government-financed setting of standards and support for advanced training for people in the private sector may be appropriate. These activities are politically safer than direct subsidies for private firms, which inevitably carry a much greater risk of capture and political manipulation and are difficult to make contingent on the desired actions by firms.

In the event that a government does undertake some form of subsidy for firms, the market in the rest of the world must always be used as the benchmark by which success is judged. Protection from foreign competitors is therefore the worst possible way to offer a subsidy for undertaking some activity. Attempts to imitate Taiwan (China) can all too easily end by yielding the closed markets and stagnation of India and Brazil.

Finally, having issued all of the cautions about the risks of intervention and the limits imposed by political and institutional constraints, one must not lose sight of the endogenous nature of political and institutional constraints. Just as in a child's chemistry set, there is far more scope for discovering new institutional arrangements than we can possibly understand. In the United States, if we had naively applied the theory of rent-seeking to the analysis of research grants, we would have concluded that government support for research can never be effective. We would not have invented peer review after World War II.

As the world becomes more and more closely integrated, the feature that will increasingly differentiate one geographic area (city or country) from another will be the quality of public institutions. The most successful areas will be the ones with the most competent and effective mechanisms for supporting collective interests, especially in the production of new ideas.⁵

The challenge for economic analysis is therefore somewhat delicate. We must take seriously the economic opportunities presented by the potential for producing new ideas and for diffusing existing ideas to the widest possible extent. In so doing, we must recognize that ideas are economic goods which are unlike conventional private goods and that markets are inherently less successful at producing and transmitting ideas than they are with private goods. We must be willing to learn from cases where collective action has been socially productive. The experience of Taiwan (China) can teach us something about what is feasible from an economic or technological point of view, even if that island's politics and institutions could not and should not be replicated elsewhere. We must be

5. For a discussion of the role of infrastructure and development and suggestive elaboration of the parallel between development in poor countries and in U.S. cities, see Rauch (1986, 1992).

open to the possibility of the discovery of new kinds of institutions for supporting the production and use of commercially relevant ideas.

Yet at the same time we must send the correct signals to developing countries about what is possible given existing political constraints. Here, the experience in Mauritius is likely to be a better guide, especially in the early stages of development. There is much that can be gained merely by using ideas produced elsewhere. There is great risk in adopting interventions, especially protectionism.

REFERENCES

- Bowman, Larry W. 1991. *Mauritius: Democracy and Development in the Indian Ocean*. Boulder, Colo.: Westview.
- DeLong, J. Bradford, and Lawrence H. Summers. 1991. "Equipment Investment and Economic Growth." *Quarterly Journal of Economics* 106 (May):445-502.
- Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer. 1991. "Growth in Cities." NBER Working Paper 3787. National Bureau of Economic Research, Cambridge, Mass.
- Grossman, Gene M., and Elhanan Helpman. 1992. *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press.
- Gulhati, Ravi, and Raj Nallari. 1990. *Successful Stabilization and Recovery in Mauritius*. EDI Development Policy Case Series, Analytical Case Study 5. Washington, D.C.: World Bank.
- Heston, Alan, and Robert Summers. 1991. "The Penn World Table (Mark5): An Expanded Set of International Comparisons, 1950-1988." *Quarterly Journal of Economics* 106 (May):327-68.
- Levine, Ross, and David Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review* 82 (September):942-63.
- Meade, James. 1961. "Mauritius: A Case Study in Malthusian Economics." *Economic Journal* 71 (September):521-35.
- Pack, Howard. 1988. "Industrialization and Trade." In Hollis Chenery and T. N. Srinivasan, eds., *The Handbook of Development Economics*. New York: North-Holland.
- . 1992. "New Perspectives on Industrial Growth in Taiwan." In Gustav Ranis, ed., *Taiwan: From Developing to Mature Economy*. Boulder, Colo.: Westview.
- Rauch, James. 1986. "The Transfer of Production from Rich to Poor Countries." *Journal of Development Economics* 23:41-53.
- . 1992. "Does History Matter Only When It Matters Little? The Case of City-Industry Location." Working Paper, Department of Economics, University of California, San Diego.
- Rhee, Young Whee, Katharina Katterbach, and Janette White. 1990. "Free Trade Zones in Export Strategies." Industry Series Paper 36. World Bank, Industry and Energy Department, Washington, D.C.
- Romer, Paul M. 1990a. "Are Nonconvexities Important for Understanding Growth?" *American Economic Review* 80 (May):97-103.
- . 1990b. "Endogenous Technological Change." *Journal of Political Economy* 98 (October):S71-S102.

- Shleifer, Andrei, and Robert Vishny. 1992. "Corruption and Economic Growth." Working Paper. Harvard University, Cambridge, Mass.
- Solow, Robert M. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39:312-20.
- Wade, Robert. 1990. *Governing the Market*. Princeton, N.J.: Princeton University Press.
- Wallis, J. J., and Douglass C. North. 1986. "Measuring Transactions Costs in the American Economy, 1870-1970." In Stanley Engerman and Robert C. Gallman, eds., *Long Term Factors in American Economic Growth*. NBER Studies in Income and Wealth, vol. 51. Chicago, Ill.: University of Chicago Press.
- Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.
- Young, Alwyn. 1992. "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore." *NBER Macroeconomics Annual 1992*. Cambridge, Mass.: MIT Press.

COMMENT ON "TWO STRATEGIES FOR ECONOMIC DEVELOPMENT:
USING IDEAS AND PRODUCING IDEAS," BY ROMER

Kaushik Basu

Before reading this paper for the conference, I thought it would be a good idea to familiarize myself with some earlier studies on growth. After working through the mathematics and equations of some of the literature on growth economics, I turned to Paul Romer's paper and soon came across his observation that discussions of growth and development "might be more fruitful if we spent less time working out solutions to systems of equations and more time defining precisely what the words we use mean." Although this left me wondering about the judiciousness of my intellectual investment, I was pleased that the paper does live up to this observation about the role of mathematics on the one hand and of words on the other.

I welcome Romer's methodological position on the use of equations and mathematics. While I believe that mathematics is an important instrument for an economist, I am also aware that many economists see mathematical complication as an end in itself. That is why, when an economist says "I have tried to make the model as simple as possible," often the truth is that he or she has made the model as complicated as his or her skills permit. This is unnecessary because economics requires a kind of logical intuitive skill that is very much its own and offers immense intellectual challenge.

Romer's paper illustrates this very nicely. Much of the paper proceeds with the aid of metaphors. He begins by explaining the standard production function, using the metaphor of "the Play-Doh Fun Factory." I must confess that I had never heard of the Play-Doh Fun Factory. But now I have a fairly clear idea of what it is because I used Romer's metaphor in reverse. Because I happened to know what a production function is (to the extent that this is possible), I used his metaphor to deduce what the Fun Factory must be. Now I'll have to see a real Fun Factory to determine how useful this method of deduction is.

I. IDEAS, BELIEFS, AND ECONOMIC PROGRESS

The central theme of Romer's paper is to explain economic growth by giving pride of place to the creation and utilization of "ideas." What he means by an

Kaushik Basu is professor of economics at the Delhi School of Economics.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

"idea" is novel. It is close to but distinct from Lucas's (1987) "idea," which is a kind of human capital. Romer's idea is a "nonrival" good—one agent's using it does not preclude other agents from using it. It is like a candle flame that can light other candles costlessly. Unlike a standard public good, however, which is also nonrival, an idea is an excludable good. An agent can prevent another from using an idea. This, as we shall see in a moment, gives it a very special place in the economic development of nations.

The model analyzes the experiences of two economies—Mauritius and Taiwan (China). This provides a brief and suggestive exercise. More detailed empirical analyses will be necessary to show conclusively how the model works. The paper goes on to deduce policy prescriptions. But prescriptions cannot be deduced from purely positive analysis. David Hume knew this, and Romer makes no mistakes. Recognizing the role of politics and value judgments in policymaking, he lays down a few prescriptions with warnings, ifs, and buts. I like the ifs and buts. The world is too interesting for easy, polar solutions to be of much use. I believe that when a person confidently claims that he knows the exact solution to the problem of underdevelopment, the only safe conclusion is that he does not know what he does not know.

Romer argues that savings and education are important to boost growth. But he notes that these require cuts in consumption and that poor countries may not be able to take consumption cuts. So the first step should be to "import" ideas by attracting foreign investment and then, after growth picks up a little, to push for more savings and education. The paper recognizes the role of markets and governments, as does Krugman's paper in this volume. Although it does not recommend a Taiwanese-style government intervention, which worked in Taiwan (China) but may not work elsewhere, it notes that collective action can play a significant role.

I do not wish to dwell on the prescriptions of the paper but instead to examine some of its analytical constructs. The essence of the role of ideas is captured by the following aggregate production function:

$$(1) \quad Y = F(K, L, H, A)$$

where Y is output, K capital, L and H unskilled and skilled labor, and A the stock of ideas. Let the developing country and the industrial nation have the same production function.

Assuming that

- (a) F exhibits constant returns to scale in K , L , and H ,
- (b) capital moves to another country if productivity is sufficiently high there, and
- (c) the marginal product (and therefore salaries) of both unskilled and skilled workers is lower in the developing country,

the author concludes that developing countries have a smaller stock of ideas. An idea is an excludable good, despite being nonrival, and it can remain in the

industrial country without automatically flowing to the developing country. It follows from this that a developing country stands to gain if it can attract ideas. And since ideas normally come embodied in investment, developing countries need investment from industrial countries.

If we start from (1), Romer's deduction is fine. But is it necessary to start from (1)? Note first that (1) is clearly a frontier production function—if it were not, assumption (a) would be difficult to justify. Hence some important matters may be lodged in the distinction between the frontier production function and the "actual" one. I return to this issue in the next section.

The second way to modify (1) is to allow for other variables, similar to A , in the function. For instance, are there other variables that, like ideas, are nonrival but nevertheless do not flow easily across nations; and differences that explain how assumptions (a), (b), and (c) can be valid? I believe the answer is yes. Although several variables may be candidates, I focus here on "beliefs"—a variable that my familiarity with the Indian economy suggests is extremely important in thwarting or abetting growth. All people have beliefs—beliefs about, for example, how an economy works (I call these positive beliefs) and what is fair and unfair (normative beliefs). I follow the *method* that Romer uses but throw this new variable into the brew. Note that beliefs are nonrival because my belief can be costlessly taken up by others. And although beliefs are not excludable in the same way as ideas, they do not flow easily across nations, groups, and communities.

I maintain that beliefs, quite irrespective of whether they are right or wrong, can play a major role in aiding or hindering economic progress. In other words, they can be economically advantageous or economically disadvantageous. A belief is economically advantageous if the fact that citizens hold the belief leads to greater economic efficiency and progress. I am also arguing that it is not necessarily the case that the right beliefs are economically advantageous, and vice versa.

Consider positive beliefs (that is, beliefs about the way the world is). As India struggles to restructure its economy, the power of these beliefs is evident everywhere. A huge amount of government money goes into subsidizing the price of fertilizer in India. Attempts to remove such a subsidy have turned out to be politically impossible. And the majority of those opposing the removal do so not because they are themselves adversely affected by it (as public choice theory suggests) but because they believe that removing the subsidy will be bad for the economy. Indeed, we have seen the power of such beliefs with respect to devaluation, direct foreign investment, and a host of other matters.

This is a problem peculiar to economics because people, for some reason, have views on economic matters, unlike many other subjects. Engineers, for instance, do not confront this problem because people do not typically hold views on what is a good aircraft wing design. If they did hold strong views on this, and engineers had to take them into account, my hunch is that most airplanes would not fly.

More surprising, even normative beliefs can have consequences. As Romer recognizes, people often reject deals that are advantageous to them individually if they feel that the deal is unfair because it gives more to others than they deserve. This is well illustrated by experiments with the "ultimatum game" (see Guth, Schmittberger, and Schwarze 1982; Frank, Gilovich, and Regan 1992). Economists often advise governments to allow for wage differentials between two workers doing the same job if their efficiency differs. This misses the point that this feature of the economy may be quite beyond the reach of the government. If people believe that it is unfair to pay different wages to people doing the same job, there may not be much the government can do (because the power of popular opposition can restrict the government's options). Note that such a belief cannot be described as right or wrong. Indeed, normative beliefs typically cannot. Nevertheless, such beliefs can be economically advantageous or disadvantageous. For instance, the belief that paying different wages to people in the same profession is unfair is likely to be bad for the economy.

Let B be the set of beliefs held by the people. What I am suggesting is that the production function (1) should instead be thought of as: $Y = F(K, L, H, A, B)$. Clearly (a), (b), and (c) can be explained not just by intercountry differences in A , but also by differences in B .

A technical problem can be raised at this point. What matters with beliefs, unlike with Romer's ideas, is not the quantity, but the quality. There may be a way of getting around this. Call the following the monotonicity axiom: for all sets of beliefs, B and B' , either B results in higher output for all K, L, H , and A , or B' results in higher output for all K, L, H , and A . If this axiom were true, we could always represent every B by a number such that a B with a larger numerical representation raises output.

The monotonicity axiom is a strong one, but it is not special to beliefs. A similar problem can arise with ideas as well. Observe that in practice there can be good ideas and bad ideas (for instance, how to grow food and how to make bombs). And since ideas usually come in clusters, packaged with, for example, investment, the simple notion of more ideas being better is not self-evident. To make such a claim, even with ideas, we need to make use of something like the monotonicity axiom.

Note that ideas can indeed be critically important in sustaining economic growth, but it is not just ideas that matter. Other invariables that share the two critically important characteristics of ideas—nonrivalry and excludability—can be important catalysts of economic progress.

II. INFORMATION, STRATEGY, AND DISAGGREGATION

I want to make two observations on the subject of good and bad ideas. If it is true that ideas can be both good and bad, then Romer's thesis pertains to *good* ideas. But in importing ideas, it may not always be easy to ensure that the good ones are acquired because the market for ideas may be subject to serious problems of asymmetric information. That is, it may be difficult to know what you

are buying until you have bought it—because to know what idea you are buying is to know the idea.

Second, whether a particular idea is good or not may not be an inherent feature of the idea but could depend on the structure of the market in which it is used. And in certain strategic environments what appears to be a good idea may turn out to be detrimental for the agents involved. Thus, for instance, as we know from Vickers (1985), Fershtman and Judd (1987), and Sklivas (1987), if one firm in a quantity-setting duopoly gets the idea of hiring a manager and giving him an objective function distinct from that of profit, it will be good for the firm. But if both firms get this idea, both will be worse off. Hence, the same idea can be good or bad depending on the market structure, on how widely the idea is available, and on which segment of society one is considering.

Finally, let us examine the nature and role of the aggregate production function in Romer's model. As noted earlier, (1) is the frontier production function. Hence, when Romer considers the case of K , L , and H doubling with A constant, he assumes that new users of capital and skilled labor have access to the existing stock of ideas, A . But this may not be the case. In particular, since ideas are excludable, ideas may not flow freely among regions, sectors, and firms, even within a country. Thus the actual relation among K , L , and H , on the one hand, and output, on the other, may not be one of constant returns. Thus assumption (c) need not in practice imply a differential access to ideas between developing and industrial nations, since (a) may fail to be true for the actual production function because of intracountry obstacles to the flow of ideas. This observation does not pertain to Romer's theory but draws attention to the fact that the assumption of free flow of ideas within a nation, implicit in Romer's formulation, may be rather restrictive. To understand reality, this assumption will need to be relaxed.

III. CONCLUSION

Romer's paper shifts our attention from the nuts and bolts of capital and labor to more unusual goods, like ideas, in the study of development. But once this shift has been made, we cannot stop at ideas. There are other unusual goods that begin to catch our attention. Beliefs are only one; institutional organization is another. The relation between organization and growth has been discussed in the literature in recent times (see, for example, Drazen and Eckstein 1988; Basu 1990), but it should be possible to do an analysis in which institutional organization enters the picture in the same way that ideas do.

In Romer's work ideas are treated as always desirable, and more ideas always lead to greater productivity. It may be more realistic to take a disaggregated view and explicitly recognize that ideas can be both good and bad. Once this is done, it becomes clear that whether ideas are good or bad may depend on the strategic structure of the market. This structure and the transmission of ideas within countries may be useful directions for future research.

REFERENCES

- Basu, Kaushik. 1990. *Agrarian Structure and Economic Underdevelopment*. New York: Harwood Academic.
- Drazen, Allan, and Zvi Eckstein. 1988. "On the Organization of Rural Markets and the Process of Economic Development." *American Economic Review* 78:431-43.
- Fershtman, Chaim, and Kenneth L. Judd. 1987. "Equilibrium Incentives in Oligopoly." *American Economic Review* 7:927-40.
- Frank, Robert H., T. Gilovich, and D. T. Regan. 1992. "Do Economists Make Bad Citizens?" Cornell University, Department of Economics, Ithaca, N.Y.
- Guth, W., R. Schmittberger, and B. Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3:367-88.
- Lucas, Robert E. 1987. "On the Mechanics of Economic Development." Nancy L. Schwartz Memorial Lecture, J. L. Kellogg Graduate School of Management, Northwestern University, Evanston, Ill.
- Sklivas, S. D. 1987. "The Strategic Choice of Management Incentives." *Rand Journal of Economics* 18:452-58.
- Vickers, John. 1985. "Delegation and the Theory of the Firm." *Economic Journal* 95 (supplement):138-47.

COMMENT ON "TWO STRATEGIES FOR ECONOMIC DEVELOPMENT:
USING IDEAS AND PRODUCING IDEAS," BY ROMER

Marcelo Selowsky

I have organized my discussion around two questions. First, how does this paper—and the associated literature—change or contribute to what I already know? Second, in what way does it help me better identify or fine-tune policy interventions in this area?

I was trained in the late 1960s in the tradition of Harry Johnson, T. W. Schultz, Zvi Griliches, Arnold Harberger, and George Stigler. From that tradition I developed a strong belief in the importance of open trade and investment, the role of human capital, the importance of investment in research and development and in know-how, the distinction between the frontier of technology and the rate of adaptation, information as an investment, and so on. The importance of property rights and patents and the removal of constraints hindering investment in new ideas was also a part of that tradition.

These notions were embedded in a very useful organizing principle: growth was a generalized process stemming from the accumulation of physical and human capital and the rate of adaptation to a changing stock of ideas and know-how. All this could also be seen as a process of investment subject to rates of return. Possible gaps between social and private rates of return to those investments provided the welfare basis for government intervention through subsidization or taxation. This principle was rich in theoretical insights and testable microeconomic hypotheses, as evidenced in Griliches's (1958) pioneering work on investment in the development of hybrid seed varieties (which was followed by Ph.D. dissertations in several developing countries); Stigler's (1961) work on information; and Schultz's (1964, 1975) work on the adoption of technologies by small-scale farmers, which generated a massive amount of research in developing countries with important policy implications.

This overall tradition has permeated my operational activities, particularly in Latin America. Open trade and investment, the entry of the private sector into areas previously reserved for the state, and aggressive policies of investment in human resources are part of the policy agenda in Latin America today. And I believe it is necessary to subsidize some types of human capital formation, as

Marcelo Selowsky is chief economist for Europe and Central Asia at the World Bank.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

well as some nondiscretionary investment in research and development. This is what I learned in the late 1960s, and this has shaped my thinking today.

Let me return to Professor Romer's paper. I have gone through twenty-five pages of taxonomies describing different degrees of appropriability and non-rivalry of goods, of ways in which ideas may enter the production function, and of ways in which ideas may be produced—all at the aggregate level. And what are the author's conclusions? That ideas have some element of a public good; that the production of ideas is probably a function of competitive pressures and, therefore, of the trade regime; that the production and acquisition of ideas are probably associated with the international mobility of people and investment; and, given that the production of ideas is intensive in human capital, that the rate of return on that capital may be associated with the openness of the economy. The paper ends by warning that trade protection and industrial policies which select individual sectors or firms for special treatment are dangerous routes for supporting the generation of know-how and ideas.

I am pleased that these conclusions emerge from the so-called new theories of growth. But what has Professor Romer said that is new? Isn't this the same message the old tradition taught? Where is the value added of this literature, at least as far as policy implications are concerned?

I am willing to go even further: in emphasizing aggregative analysis and looking for steady-state growth properties, this new literature has directed attention away from micro empirical analysis to aggregative mathematical dynamics. Where are the new testable hypotheses that would help fine-tune policy interventions? Consider the variability of behavior that affects the formation of human capital across cultures; the enormous difference in children's study habits is a case in point. (The significant differences in the time and effort devoted to supporting a child's learning by some families in the United States have been recently documented.) These differences are perhaps important elements in understanding the capacity to generate and adapt ideas. These are the areas on which we should be focusing. But the aggregative nature of the new literature does not help in this process.

Professor Romer's paper provides the same answers as mine: there may be arguments in favor of subsidizing the process of producing ideas, but not through protection or any other intervention that promotes discretionality or rent-seeking. He must agree that these conclusions are well known and that they are too general to guide any alternative subsidy policy. Should we have a general subsidy or matching funds to support research and development across all firms? This will obviously maximize neutrality. Or should the government contribute matching funds for research at universities or technological institutes, provided the private sector does the same? Should we subsidize on-the-job training? What channels should we subsidize; that is, where are the key externalities not being internalized? This is where new research should focus; Professor Romer's paper does not help in this search.

REFERENCES

- Griliches, Zvi. 1958. "Research Costs and Social Returns: Hybrid Corn and Related Innovations." *Journal of Political Economy* 66 (October).
- Schultz, T. W. 1964. *Transforming Traditional Agriculture*. New Haven, Conn.: Yale University Press.
- . 1975. "The Value of the Ability to Deal with Disequilibria." *Journal of Economic Literature* 13(3):827-46.
- Stigler, George J. 1961. "The Economics of Information." *Journal of Political Economy* 49(3):213-25.

COMMENT ON "TWO STRATEGIES FOR ECONOMIC DEVELOPMENT:
USING IDEAS AND PRODUCING IDEAS," BY ROMER

T. N. Srinivasan

Paul Romer has provided us with a provocative paper on the distinction between knowledge, on the one hand, and human and physical capital, on the other, in the growth process.

It is difficult to exaggerate the importance of his statements that the rest of the world must always be used as a benchmark by which the success of industrial policy is judged and that protection from foreign competitors is the worst way to encourage domestic firms to undertake any activity, even research and development. He is correct that integrating a developing economy with world markets offers large potential gains compared with insulating it from such markets through protection. Openness to investments by foreigners bearing ideas is an inexpensive way of promoting growth, and such openness need not preclude the development of local expertise where warranted. I share Romer's skepticism about industrial policy for industrial as well as developing countries. I agree that Japanese or Korean industrial policy institutions are difficult to replicate successfully in other countries and that even in these countries their long-run survival is not entirely assured. Indeed, as he points out, interventions similar to those in the Republic of Korea or Japan have been tried in other countries, such as Brazil and India, without success.

Perhaps the main reason for the lack of success was that the world market benchmark was not applied. I vividly recall my experience in Delhi in the early 1970s when my institution, the Indian Statistical Institute, installed a domestically produced photocopying machine. It was half the size of an office room and involved several steps to produce one smudged page every five minutes! There was a market for it only because it was the only machine in town—the notorious indigenous clearance angle of the Indian import licensing system had made sure of that.

Romer draws attention to the importance of the export processing zone (EPZ) in explaining the success of Mauritius. The only industry in the EPZ is textile manufacturing, established by Chinese entrepreneurs from Hong Kong who were attracted to the EPZ not only by the economic incentives but, more impor-

T. N. Srinivasan is Samuel C. Park, Jr. Professor of Economics at Yale University.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

tantly, because the notorious Multifibre Arrangement did not yet apply to Mauritius. In contrast, India is an example of failed EPZs and the systematic emasculation of a once-thriving and internationally competitive textile industry by misguided government policy. In the 1960s India established an EPZ at Khandla, on the west coast. It failed, first, because Khandla was neither well connected with the hinterland nor endowed with the necessary infrastructure and human capital, and, second, because the government was ambivalent about permitting foreign investment (and other economic activities that were too different from those it permitted in the rest of India). It is well known that expansion and modernization of integrated textile mills in India were not permitted, to protect the handwoven cloth sector and domestic textile machinery makers.

Romer points out that although Taiwan (China) also established EPZs and bonded factories, it went further than Mauritius in encouraging domestic production and exporting ideas for sale in world markets. It did protect domestic industry through discretionary administrative controls.

India also established several national laboratories under the aegis of the Council of Scientific and Industrial Research, as well as several institutes of technology and polytechnics, agricultural research institutes, and universities. And India's discretionary administrative controls for granting domestic industry made-to-measure protection indefinitely are well known. Except in agriculture—where India's plant breeders successfully developed several high-yielding varieties of wheat and rice—the rest of the research and development effort must be deemed cost-ineffective, if not altogether a failure. The contrast between India and Taiwan (China) appears to be that in the latter the research institutions were given explicit commercial objectives and the government, in granting (mostly temporary) protection, never lost sight of the goal of international competitiveness. In India protection once granted became a permanent entitlement.

I now turn to Romer's model of growth. He refers to the work of DeLong and Summers (1991), in which variations in the rate of investment in equipment explain a significant part of the variation in growth in a cross-section of countries. One strand of this argument was developed by Mahalanobis (1955), the father of India's heavy-industry development strategy. He argued for keeping the cost of capital goods as low as possible, particularly equipment that is at the farthest end from final consumption goods. If the government had complete control over this industry, it would be able to control prices and shape the pattern of industrialization. And by regulating imports, it could also influence investment in the private sector. As it turned out, however, domestic production of equipment by the public sector, price controls, and the restriction of competition from imports raised rather than lowered the cost of this equipment.

In Romer's model the sector that produces knowledge or ideas plays a role analogous to that of the capital goods sector in the two-sector model of Mahalanobis (1955) and Feldman (1928, as described in Domar 1957), a model well known at least among economists of my generation. With no nonproduced factor of production such as labor, holding the marginal product of capital in

each sector constant, this model generates endogenous growth without having to appeal to indefinite scale economies or externalities of any kind. Because capital, once installed in a sector, cannot be shifted, the share of investment devoted to accumulation of capital in the capital goods sector determines *endogenously* the long-run rate of growth of the economy.

An expanded version of the model (see the Appendix) is helpful in clarifying the circumstances under which the opening of the economy to world trade can augment growth. Assume that there are two consumer goods as well as two investment goods (instead of one each) and that the marginal product of capital is constant in the production of each good. The utility function and the function that transforms the output of the two investment goods into aggregate investment are Cobb-Douglas. Capital stock in the consumer (investment) goods sector is mobile within the sector producing the two consumer (investment) goods. There is no intersectoral mobility of capital.

Assume, for simplicity, that the share of investment devoted to the accumulation of capital stock in the capital goods sector is exogenously fixed rather than endogenously determined through intertemporal welfare maximization. Under autarky all four goods are produced in positive amounts. Suppose now that this economy is opened to free trade in consumer goods, with the relative price of the two consumer goods being fixed in world markets. Then the economy will specialize in producing one of the two consumer goods in which it has comparative advantage, and the welfare of the economy will rise in relation to autarky. But as long as the share of investment devoted to the capital goods sector is unchanged and that sector is closed to foreign trade, the long-run growth rate of the economy will be unchanged. In contrast, if the capital goods sector is open to free trade (again at fixed world relative prices) while the consumer goods sector is closed, there will be a positive long-run growth effect and a positive welfare effect in relation to autarky. The implication is that, from a growth perspective, keeping the growth-inducing sector (say, the capital goods sector or the knowledge sector) closed to international competition is costlier than closing the consumer goods sector. But of course, keeping neither closed would be even better because the model contains no avenue for strategic trade policy or dynamic externalities as a result of learning effects. I make no apologies for this. I do not believe that the factors that lead to rewarding strategic interventions in foreign trade in models of New Trade Theory are present in developing countries. Nor do I think that dynamic learning effects are automatic. For example, in the Indian automobile industry there was very little evidence of learning, in spite of more than four decades of sheltered existence.

In Romer's model the foreign monopsonist textile producer earns a rent representing the difference between the wage bill and the output produced (at a constant marginal product) at the optimal number of workers employed. By setting a minimum wage, the government can ensure that part of this rent accrues to the workers. The maximum amount that can be so transferred is determined by the requirement that the monopsonist's share of the rent be

enough to compensate for the fixed cost of entry into Mauritius. But transferring rent through minimum wage is a nonoptimal distortionary policy in this context because there are two objectives: to induce the monopolist to enter and to induce the choice of the competitive employment level. A policy that stipulates the employment level or, equivalently, a minimum wage equal to the marginal product of labor and that induces entry through a lump-sum subsidy is superior to the minimum-wage-only policy. This result is the analogue of Oi's (1971) well-known two-part tariff.

I would like to conclude with a very simple model of innovation and imitation. Suppose a research budget of $\$R$ can be spent encouraging independent innovation efforts at a cost of $\$x$ each and imitation efforts at a cost of $\$y$ each. To set up innovation, a fixed cost of $\$F$ has to be incurred regardless of the number of efforts. Suppose the probability of success of an innovation effort is π and the probability of success of an imitation effort is $\bar{\pi}$. Then the probability S that at least one effort will succeed is

$$S = 1 - [1 - \alpha\pi(x)]^n [1 - \bar{\pi}(y)]^m$$

where $\alpha = 0$ if no innovation efforts are encouraged, $\alpha = 1$ if $n \geq 1$ such efforts are encouraged, and m is the number of imitation efforts. The budget constraint is

$$\alpha F + nx + my = R.$$

Let us assume that

$$\pi(0) = \bar{\pi}(0) = 0, \pi'(x) > 0, \bar{\pi}'(y) > 0, \pi''(x) < 0, \bar{\pi}''(y) < 0$$

and that both π and $\bar{\pi}$ approach 1 asymptotically as x and y tend to infinity. Now if $R \leq F$, obviously it is not feasible to undertake any innovation effort. Whether or not more than one imitation effort is undertaken would in general depend on the tradeoff (as a result of the budget constraint) between increasing the number of efforts m and the expenditures per effort y ; that is, any *increase* in m can only come about with a *decrease* in y and vice versa. Since S increases with m for fixed y and with y for fixed m , that tradeoff involves two opposing effects on S . If, for example, $\bar{\pi}(y) = 1 - e^{-\beta y}$, $\beta > 0$, these two effects just offset each other, and the probability of success is the *same* whether one or more imitation efforts are undertaken. Thus $S = 1 - e^{-\beta R}$. If $R > F$, undertaking innovation effort is feasible. Once again, if $\pi(x) = 1 - e^{-\gamma x}$ and $\bar{\pi}(y) = 1 - e^{-\beta y}$, then $S = 1 - e^{-\beta R}$ if no innovation is undertaken, and $S = 1 - e^{-\beta R + \beta F + (\beta - \gamma)nx}$ if $n \geq 1$ innovation efforts are undertaken. Clearly, if $\beta \geq \gamma$ so that imitation is at least as productive as innovation, no innovation effort should be undertaken. Even if $\beta < \gamma$, if $\beta F + (\beta - \gamma)(R - F) > 0$ (that is, if the larger productivity of innovation is not enough to offset the fixed cost F when all available resources are devoted to innovation), once again innovation would not be worthwhile. Only if $\beta F + (\beta - \gamma)(R - F) \leq 0$ is innovation worthwhile. Of course, with other functional forms for π and $\bar{\pi}$, the two effects do not just offset each other

and hence the choice of n and m has to be considered. An implication of this simple model is that the Mauritian approach of imitation and the Taiwanese approach of innovation as well as imitation could both be optimal for their circumstances.

APPENDIX

A. Consumption goods sector

Output of good $i = Q_i^c$ ($i = 1, 2$); domestic use of good $i = C_i^c$ ($i = 1, 2$); stock of capital = K^c .

$$(A.1) \quad \text{Production frontier} \quad \beta_1 Q_1^c + \beta_2 Q_2^c = K^c$$

$$(A.2) \quad \text{Utility function} \quad U = (C_1^c)^\alpha (C_2^c)^{1-\alpha}$$

B. Investment goods sector

Output of good $i = Q_i^I$; domestic use of good $i = A_i^I$; stock of capital = K^I .

$$(A.3) \quad \text{Production frontier} \quad \gamma_1 Q_1^I + \gamma_2 Q_2^I = K^I$$

$$(A.4) \quad \text{Aggregate investment} \quad I = (A_1^I)^\delta (A_2^I)^{1-\delta}$$

C. Capital accumulation

Let λ be the share of investment devoted to accumulation of K^I . Then

$$(A.5) \quad \dot{K}^I \equiv \frac{dK^I}{dt} = \lambda I$$

$$(A.6) \quad \dot{K}^c \equiv \frac{dK^c}{dt} = (1 - \lambda)I.$$

D. Autarky

$$C_i^c = Q_i^c \quad \text{and} \quad A_i^I = Q_i^I \quad \text{for} \quad i = 1, 2.$$

Maximization of (A.2) subject to (A.1) leads to:

$$(A.7) \quad C_1^c = \alpha K^c / \beta_1, \quad C_2^c = (1 - \alpha) K^c / \beta_2 \quad \text{and} \quad U = \left(\frac{\alpha}{\beta_1} \right)^\alpha \left(\frac{1 - \alpha}{\beta_2} \right)^{1-\alpha} K^c.$$

Maximization of (A.4) subject to (A.3) leads to:

$$(A.8) \quad A_1^I = \frac{\delta K^I}{\gamma_1}, \quad A_2^I = \frac{(1 - \delta) K^I}{\gamma_2}.$$

Substituting (A.8) in (A.4), one gets:

$$(A.9) \quad I = \left(\frac{\delta}{\gamma_1}\right)^\delta \left(\frac{1-\delta}{\gamma_2}\right)^{1-\delta} K^I$$

$$(A.9') \quad \equiv \eta K^I \quad \text{where} \quad \eta = \left(\frac{\delta}{\gamma_1}\right)^\delta \left(\frac{1-\delta}{\gamma_2}\right)^{1-\delta}$$

Using (A.9') in (A.5) and (A.6) and solving:

$$(A.10) \quad K^I = K_0^I e^{\lambda \eta t}$$

$$(A.11) \quad K^c = K_0^c + \left(\frac{1-\lambda}{\lambda}\right) K_0^I (e^{\lambda \eta t} - 1).$$

Thus the long-run growth rates of K^I , K^c , I , and U are the same and equal $\lambda \eta$.

E. Free trade in consumer goods at a relative price π^c of good 2 in terms of good 1 and autarky in investment goods

Without loss of generality, assume $\pi^c > \beta_2/\beta_1$. Then it is optimal to produce only good 2, export part of the output, and import good 1. It is easy to show that $Q_1^c = 0$, $Q_2^c = K^c/\beta_2$, $C_1^c = \alpha \pi^c K^c/\beta_2$, $C_2^c = [(1-\alpha)K^c]/\beta_2$, and $U = \alpha^\alpha (1-\alpha)^{1-\alpha} (\pi^c)^\alpha K^c/\beta^2$. It is easy to verify that given $\pi^c > \beta_2/\beta_1$, welfare U under free trade is higher than under autarky for any K^c . Since the investment goods sector is closed to trade, the dynamics of the system are unaffected so that the paths of K^c and K^I continue to be given by (A.10) and (A.11). Hence the long-run growth rates of K^c , K^I , I , and U are still $\lambda \eta$, although the level of U at each t is higher than under autarky.

F. Free trade in investment goods at a relative price of π^I of good 2 in terms of good 1 and autarky in consumption goods.

Without loss of generality assume $\pi^I > \gamma_2/\gamma_1$. Then it is optimal to produce only good 2, export part of the output, and import good 1. It is easy to show that $Q_1^I = 0$, $Q_2^I = K^I/\gamma_2$, $A_1^I = \delta \pi^I K^I/\gamma_2$, $A_2^I = [(1-\delta)K^I]/\gamma_2$, and $I = \delta^\delta (1-\delta)^{1-\delta} (\pi^I)^\delta K^I/\gamma_2$. It is easy to verify that given $\pi^I > \gamma_2/\gamma_1$, investment I under free trade is higher than under autarky for any K^I . Now using (A.5) and (A.6) and solving, one gets:

$$K^I = K_0^I e^{\lambda \mu t} \quad \text{where} \quad \mu = \{\delta^\delta (1-\delta)^{1-\delta} (\pi^I)^\delta\} / \gamma_2$$

and

$$K^c = K_0^c + \left(\frac{1-\lambda}{\lambda}\right) K_0^I (e^{\lambda \mu t} - 1).$$

Now given $\pi^I > \gamma_2/\gamma_1$, it follows that $\mu > \eta$. Hence the values of K^I and K^c at each time t under free trade in investment goods are higher than their corresponding values under autarky, and the long-run growth rates of K^c , K^I , and I are the same at $\lambda \mu$, which is also higher than its value $\lambda \eta$ under autarky.

Since the consumption goods sector is under autarky given K^c , $Q_1 = C_1 = \alpha K^c / \beta_1$, $Q_2 = C_2 = [(1 - \alpha)K^c] / \beta_2$, and $U = (\alpha / \beta_1)^\alpha [(1 - \alpha) / \beta_2]^{1-\alpha} K^c$. Since K^c is higher at each t under free trade in investment goods than under autarky, U is higher as well. Since K^c grows faster, U grows faster as well.

REFERENCES

- DeLong, J. Bradford, and Lawrence H. Summers. 1991. "Equipment Investment and Economic Growth." *Quarterly Journal of Economics* 106 (May):445-502.
- Domar, Evsey. 1957. *Essays in the Theory of Economic Growth*. London: Oxford University Press.
- Mahalanobis, P. C. 1955. "The Approach of Operational Research to Planning in India." *Sankhya: The Indian Journal of Statistics* 16, pts. 1-2:3-130.
- Oi, Walter. 1971. "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly." *Quarterly Journal of Economics* 85(1):77-96.

FLOOR DISCUSSION OF THE ROMER PAPER

A participant from the World Bank asked Romer to explain how ideas that are nonrival but excludable fit with the club theory, which deals with the issue of goods (such as swimming pools) that are nonrival but excludable. Romer explained that club theory assumes that the element of nonrivalry essentially goes away at finite capacity; with partially nonrival goods, congestion effects set in. Ideas are fundamentally different because there are no congestion effects. That might seem like a small degree of difference, but the implications are profound.

A participant from the Brookings Institution asked Romer what kind of intellectual property law would best disseminate ideas and encourage production. Romer responded that this question was too complex for the present discussion but that basically the only reason firms invest in generating ideas is to have a monopoly on their results for a certain length of time: ideas were used to reap monopoly rents. The outcome is an imperfect situation wherein ideas are not freely available, which leads to interventions. This is not market failure in the usual sense of the word; it happens because ideas, which are incremental in nature, are appropriable. Most of the discussion on this issue, he added, has been based, unknowingly, on the object side of the four-way figure in his paper. The belief that stronger property rights are unambiguously welfare-improving applies only to that objects column. The subtleties of what would constitute good intellectual property laws are immense because of the important efficiency and distributional questions.

A participant from the University of Buenos Aires asked Romer to comment on the phenomenon of incomplete specification. Suppose that a firm with a license is given an engineering manual that is incomplete. One would have to generate knowledge to understand how to use the imported technology, he said. What kind of knowledge would be needed to use that plant? Should the government subsidize the acquisition of that knowledge?

Codified specifications are an issue in the transfer of technology, said Romer, but that should not stop the flow of ideas. People, as well as manuals, might have to be imported, but it could be done.

A World Bank participant asked what enabled developing countries to draw on the stock of global knowledge? To what extent should there be information

This session was chaired by Kemal Dervis, director, Country Department II, Europe and Central Asia Regional Office, the World Bank.

services and research institutes? What kinds of things must be done to maximize the externalities available from foreign investment? Continuing this line of thinking, another speaker said that since knowledge had some aspects of a public good, one might want to subsidize its production. Could Romer say which countries might have a comparative advantage in producing knowledge, and which might be better off importing it?

Romer said that support for the importance of ideas ultimately comes from people who look narrowly at the microlevel disaggregated evidence. Their work had influenced him greatly, and he realized that his macroanalysis was useless if it did not get down to a micro level. What he had tried to do was to tie things up in a little package the way economists did with the Solow model when they did growth accounting.

Making policy claims in this area, said Romer, took more insight into politics and institutions than he had, but if someone said that he had to reach some policy conclusions, he would rely on the evidence on the ground. He would learn a lot more about what works and what does not work—looking at cases like Taiwan (China), Mauritius, and Brazil—rather than spend time with a pencil and paper trying to create a model and derive some conclusions from it. He would encourage schooling, training, and other investments in human capital, which have high payoffs, and he would encourage investments in machinery. But Mauritius illustrates that even a low-income country that cannot afford to provide the schooling and develop the machinery it may desire has the option of bringing in qualified personnel and equipment.

A participant asked Romer and the other discussants which they thought was preferable: some kind of direct, transparent allocation for subsidizing research, or indirect subsidies through tariffs and other forms of protection?

Marcelo Selowsky (discussant) asked Romer if he thought there was something between the two extremes: picking one firm and completely subsidizing its research and development, or telling all firms that for every ten dollars they invested in research and development (R&D) the government would give them one dollar. Romer said he thought that was really two questions. First, how would you design the institutions? Second (and harder to figure out), what fraction of GDP should go there? There was a big gap between social and private returns that could be exploited, even in the United States, if we could figure out how to do it, said Romer.

One participant suggested an auctioning of subsidies that involved cofinancing, to maximize the mobilization of research funds. The government would say, we're going to devote \$100 million to research, and whoever is willing to put up the most cofinancing per subsidy dollar will get the subsidy funds. So, if company A bid five dollars of cofinancing per subsidy dollar and company B bid ten dollars per subsidy dollar, company B would win the subsidy.

Romer said that the auction had the flavor of market competition, which he thought was better than the government's trying to collect revenues (thus requiring coercion), but that he would probably favor using subsidies to train engi-

neers and letting the market allocate where the engineers would go. The political economy there was clear-cut, he thought. He offered the example of the U.S. Department of Defense, which is implementing an industrial policy for parallel supercomputers. The department has picked two winners, with two different architectures. Romer said he thought this mechanism was wrong—first, because the U.S. Congress would take it over and eventually misuse it, and, second, because he would not want to rely on the judgment of a few people. He thought it would be better to tax everybody who buys a supercomputer now—at, for example, 2 to 5 percent of the price—and say that the tax must be devoted to research on making supercomputers better, but let the buyers decide to whom it would go. In other words, the government would say that x percent of all spending has to go to this activity, but the users of the new technology would decide where that research money would go.

Another participant said that the interesting question was, what proportion of resources should go to basic research and what proportion to directed research—and directed in what way? He pointed out that Romer had not yet responded to Marcelo Selowsky's question, in his comments, about testable propositions.

What critical experiment would support or disprove what I am talking about? asked Romer. Take a case like this: a country in which GNP increases rapidly, although there have been no imports of goods or capital—the only thing that has happened is that some people came in who had ideas, as Hong Kong entrepreneurs had come to Mauritius. This country gets no machinery from outside, it is not allowed to change the level of schooling, and the labor force is fixed. But by virtue of people coming in with ideas, there is a big increase in GNP or in output per worker. Romer said he would take that as strong evidence in favor of the model he was describing. Mauritius was not the perfect example, but it came close: it had large domestic savings, and although it was not manufacturing its own sewing machines, it could do so. Romer said that he had not made his full case yet, but he thought he could. He said his notion of success had been that in the end people would say, we knew this all along. The question is not whether it is true but why there is a disparity between what is true and what the models predict. If we knew this proposition was true all along, why didn't we spend time tracking down models that captured it?

A speaker on a later panel brought the discussion back to the empirical question of how to encourage activities and ideas for technical change. He had two comments. First, governments must be careful about centrally governed or tax-subsidized R&D because R&D is typically done by larger firms, and certain activities might be better handled by smaller firms. Second, in the industrial market economies, differences in the capacity to generate and exploit ideas are determined by the technological activities financed by firms themselves. If the firms are good capitalists, they will know what is good for them and will fund it. On the whole, he thought, subsidies do not work in advanced countries.

The interesting question, said the speaker, is why, over very long periods, some countries are willing to put money into technology-generating activities and others are not. We do not know why the Japanese spend more than the Americans, or why over eighty years the Germans spent more than the British. The British went through imperialism, welfare-statism, and Thatcherism, and it did not make a bit of difference. We do not know why technological activity has exploded in firms in the Republic of Korea and Taiwan (China) in the past five to ten years but not in Latin America. Firms differ enormously in dynamic efficiencies across countries. Those are the facts; explaining them is the challenge. Inventing clever subsidy schemes without understanding these differences could be dangerous, he concluded.

A participant from Ankara University in Turkey said that the two economies Romer had chosen as examples were small islands with few trade barriers. He asked if the internal dynamics of ideas were distinctly different in small island economies than in large countries such as India and Brazil or other Latin American countries.

Romer said that limits set by the industrial world would pose an increasingly serious problem for development. The Multifibre Arrangement was just the beginning of what might become a series of protective actions. Wages had already fallen in the United States and would continue to fall in most of the industrial world, and procedures would continue to be implemented to solve income redistribution problems. If we do not come up with better approaches than the Multifibre Arrangement, said Romer, the developing countries will suffer. Still, there was something to the notion of learning from the rest of the world, even if it was done for the sake of domestic consumers rather than for export. Suppose the West simply behaved badly in this area? What strategy could be articulated for developing countries? They might form large trading blocs that could learn from the industrial countries by allowing technology imports. They might allow Honda to come in and build an automobile plant in Turkey, if necessary, and then if Turkey couldn't sell its Hondas in Europe or the United States, it could sell them to other developing countries. One country could sell Hondas, one could sell pharmaceuticals, one could sell computers, and so forth. If something works, use it over and over again—but that was going to become harder to do.

A participant from the World Bank wondered if Romer thought that moving toward globalization necessarily meant mass production. What if a country wanted to resist that momentum and preserve its culture and handicrafts instead of having McDonald's here, there, and everywhere? Should such a goal be subsidized or supported, even if it meant loss of production? Romer said he agreed absolutely with Kaushik Basu (discussant) on the importance of beliefs and culture and that some beliefs would not facilitate growth. Similarly, some values would not support growth—and by definition, when two sets of values conflicted, those who held one set would seem wrong to those who held the other. There would be great stress, he said, as countries—for better or worse—

abandoned value systems that were not conducive to growth and adopted new ones. Whether that was good or not was philosophically almost impossible to decide, but values and beliefs would continue to change. Taking ideas seriously, he said, leads to the conclusion that the arena in which one operates is inherently a second-best kind of world; that is when logic provides some space for intervention. But the best interventions take advantage of opportunities for the collective good, not for individual decentralized action. Romer noted that once his view of productive opportunity was taken seriously, it was apparent why values and norms were not a matter of indifference. Later in the discussion, Basu observed that people get locked into a value system. It might be in a society's interest to do away with a certain set of values, and everyone might be aware of that, but it might not be in an individual's interest to abandon those values if others were adhering to them.

T. N. Srinivasan (discussant) said that values, beliefs, or ideologies—whatever one called them—almost all have a functional origin. Belief systems that prohibited eating pork and beef were more cost-effective than having an army of inspectors examining pork production. When beliefs become dysfunctional because of changes in technology, he said, they may last a while, but they won't last indefinitely. It is too easy, he said, to treat beliefs as exogenous and as a given that is unresponsive to market forces. Romer agreed with him completely but said that subtle issues were involved. Suppose certain people had beliefs that were not conducive to growth, and suppose one could come in and get them to change their beliefs in a way that produced faster growth. Would that make them better off? Another participant pointed out the philosophical difficulty of evaluating such a change of preferences. If someone changes my preferences, he said, my welfare with changed preferences can be observed, but it cannot be compared with my welfare with unchanged preferences.

Basu found Srinivasan's idea that beliefs had a functional origin too simplistic. A belief might have a functional origin, or it might not. Basu preferred an evolutionary explanation of changing values and belief systems: societies and clusters that adhere to dysfunctional values for long periods will do worse and will vanish from view through natural selection. He said that Srinivasan's distinction between exogenous and endogenous was also simplistic. In a sense, he said, everything on earth is endogenous; "exogenous" was really an analyst's construct and blurred reality, depending on one's viewpoint.

Basu said he did not agree that it was only a technical point that the assumption of constant returns to scale was troublesome. Ideas were being treated as excludable in terms of cross-border movements, he said, but not as excludable in terms of movement within a country, which was why he saw the justification for constant returns to scale. Allowing for the fact that an idea might be excludable even after it came into the country, doubling all other factors might not result in a doubling of output if those factors did not accrue where the ideas were. If one was concerned about the distribution of these factors, one must be somewhat concerned about the distribution of ideas.

Labor Market Institutions and Policies: Help or Hindrance to Economic Development?

Richard B. Freeman

This paper examines the wide disagreement about the value of institutional interventions in developing country labor markets between (World Bank) economists who see government regulation of wages, mandated contributions to social funds, job security, and collective bargaining as "distortions" in an otherwise ideal world and International Labour Organisation (ILO) economists who stress the potential benefits of interventions, hold that regulated markets adjust better than unregulated markets, and endorse tripartite consultations and collective bargaining as the best way to determine labor outcomes. It presents a scorecard of evidence to judge which view is closer to the truth on particular issues.

The paper finds little support for the notion that interventions are major impediments to resource allocation, structural adjustment, or stabilization programs, although in some cases they have sizable costs. At the same time, it finds little evidence on the value of social pacts and related consultative modes of adjustment favored by the ILO. The paper proposes a different perspective on labor market policies and institutions—as factors in the political economy of economic reform—and develops a model designed to capture the role of interventions and institutions in buttressing support for economic reforms.

Labor-market policies—minimum wages, job security regulations, and social security—are usually intended to raise welfare or reduce exploitation. But they actually work to raise the cost of labor in the formal sector and reduce labor demand . . . [increase] the supply of labor to the rural and urban informal sectors, and thus [depress] labor incomes where most of the poor are found. (World Bank 1990b, p. 63)

Minimum wages have an important role to play in protecting low income groups. . . . structural adjustment also calls for a sound industrial relations system and a commitment to tripartite dialogue. . . . Over the long run suppression of free industrial relations jeopardizes prospects for economic development. (ILO 1991b, p. 5) The 1980s highlighted . . . the need to re-regulate the labor market. (ILO 1991a, p. 65)

Richard B. Freeman is professor of economics at Harvard University, senior research fellow at the Centre for Economic Performance of the London School of Economics, and director of labor studies, National Bureau of Economic Research, Cambridge, Mass.

As the preceding quotations indicate, there is considerable disagreement about the value of institutional interventions in developing country labor markets. On one side are economists who see unregulated labor markets as neo-classical bourses in which government regulation of wages, mandated contributions to social funds, job security, and collective bargaining create "distortions" in an otherwise ideal world. These economists view their task as the technical one of measuring the adverse effects of these policies. The view that interventions are first and foremost distortions pervades most World Bank analyses of labor issues, and I shall accordingly call this the Bank Distortion View, although some Bank analysts reject it and many economists outside the Bank endorse it.

On the opposite side are institutionally oriented economists who believe that the social aspects of labor markets create such large divergences from the competitive ideal as to make that model a poor measuring rod for policy. These analysts stress the potential benefits of interventions, hold that regulated markets adjust better than unregulated markets to shocks, and endorse tripartite consultations and collective bargaining as the best way to determine labor outcomes. When efficiency conflicts with the social protection of labor, they place greater weight on the latter. The view that interventions are first and foremost socially beneficial dominates ILO analyses of the labor market, and I will accordingly call it the ILO Institutional View. *Towards Social Adjustment* (Standing and Tokman 1991) forcefully presents this perspective, superimposing the word "social" over the word "structural" in its title.

The differing views of how labor market interventions affect social well-being have contradictory policy implications. If you believe that interventions reduce growth and hamper adjustment, you will recommend that countries eliminate them and will make elimination a condition for adjustment loans under the slogan of deregulating labor markets. If you believe that interventions improve well-being, you will advise governments to encourage unionism and collective bargaining under ILO conventions,¹ and to regulate market outcomes and adhere to labor standards.

Is there compelling evidence for either the World Bank or the ILO point of view? Does experience in developing country labor markets and labor market research provide greater support for one or the other?

In this paper I examine these questions. There has been little open debate between ILO and World Bank researchers, and their analyses often coincide on particular issues. I play *agent provocateur* here because I believe confrontation will increase our understanding more than a continuation of the status quo, with

1. The key ILO conventions are No. 87, Freedom of Association and Protection of the Right to Organise, and No. 98, Right to Organise and Collective Bargaining, with article 4, "the right to negotiate wages and conditions freely without outside interference is a fundamental aspect of the freedom of association." Excluding certain matters, making collective agreements subject to prior approval or enabling them to be declared void because they run counter to government economic policy is incompatible with article 4. In addition, convention 131 on minimum wage fixing and convention 154 on collective bargaining are important endorsements of the validity of interventions.

each school of thought proceeding along its own path, implicitly dismissing rather than seriously addressing the claims of the other.

The paper has four sections. Section I introduces the combatants. Section II provides a scorecard of evidence to judge which view is closer to the truth on particular issues. Section III offers a different perspective on labor market policies and institutions—as factors in the political economy of economic reform—that has surfaced in both World Bank and ILO analyses. Section IV summarizes conclusions and offers suggestions for future research.

I. THE COMBATANTS: BANK DISTORTIONISTS VERSUS ILO INSTITUTIONALISTS

In the blue corner, from Washington, D.C., we have . . .

The Distortionist Case: Interventions Are Bad

The distortionist case hinges on four claims about interventions: they misallocate labor, waste resources through rent-seeking, impair adjustments to economic shocks, and deter investment, thereby reducing rates of growth.

Since the claim that interventions misallocate resources follows from basic price theory, all economists know the basic arguments and the provisos about market failures that qualify those arguments. The major premise is that absent interventions, labor markets set wages at opportunity cost levels and determine Pareto-efficient levels of employment, work rules, training, and so on. Since the unfettered market meets optimality conditions, interventions can only make matters worse. The resultant allocative distortions are typically measured by Harberger welfare triangles.

The possibility of transferring income from one group to another through interventions suggests that interest groups will devote resources to rent-seeking instead of to activities that raise national output (Krueger 1974). Irrespective of the success of such rent-seeking, those resources are deadweight losses to society as a whole. The economics of pressure groups also suggests, however, limits to the resultant distortions because taxpayers will organize against policies with especially large deadweight losses (Becker 1985). While no political regime, least of all a democracy, can eliminate rent-seeking, a state committed to few interventions presumably will see less such activity than a state in which interventions are common. Rent-seeking distortions are measured by the resources spent to alter or preserve ownership rights and, depending on the way rights are established, may come to equal the value of the rents (Krueger 1974).

Distortionists also believe that interventions reduce wage or employment flexibility and thus adjustment to economic shocks. Collective bargaining or wage indexation schemes that maintain real wages when national output falls or that preserve relative wages when changes are needed to induce labor to move across sectors to meet balance of trade problems reduce the ability of the economy to respond to new situations. Because economic theory has less to say about dynamic adjustments than about comparative statics, the a priori case for adjustment distortions is weaker than the case for allocative distortions.

A final potential adverse effect is to reduce rates of return to investment. Administratively set or collectively bargained wages that redistribute economic rent from capital to labor are bad because they reduce the profitability of investments. Taxes or fiscal deficits that fund public employment are bad because they crowd out funds for more productive private investment. Many small interventions and rent-seeking activity may interact to lower overall economic efficiency and returns to investment (Olson 1982). Since growth rates cumulate to massive differences in per capita output over time, the existence of intervention-induced growth distortions is potentially the most important argument in the distortionist armory.

Claims that labor market interventions have an adverse effect do not follow mechanically, it should be noted, from "pure theory." Distortionist analysts make selective use of economic theory. For example, those who believe that social security payroll taxes adversely affect savings and investment reject Ricardian equivalence; those who use nonwage costs to measure interventionist distortions reject the fungibility of modes of compensation; those who argue that employment protection laws have efficiency costs ignore Coase's theorem that property rights do not affect efficiency. Even distortionist criticisms of minimum wages involve more than applying optimizing calculus. A small country whose modern sector capital stock is foreign owned can benefit from labor market interventions that "soak" foreign capital just as it can benefit from an optimal tariff.² Distortionist arguments are not the final word of economic theory.

To validate claims that interventions have major allocative, rent-seeking, adjustment, or growth costs requires empirical evidence that interventions are effective in producing differentials in pay or conditions of work that would not otherwise arise in unfettered markets and that they have sufficiently large adverse effects on resource allocation to affect the overall economy. The issue of magnitude is critical, for the general finding that welfare triangle losses are relatively small has led many economists who accept the basic tenet of distortionism to regard the costs of labor market interventions as of second-order importance compared with macroeconomic distortions. Distortionist arguments should weigh heavily in policy only if estimates of static welfare losses are misleadingly low, if adjustment distortions seriously impair stabilization programs, or if distortions deter investment enough to reduce growth.

And in the red corner, from Geneva, Switzerland . . .

The Case That Interventions Are Good

Because there is no general institutionalist theory, the case for interventions is more diffuse and less analytically grounded than the case against them. A major

2. Consider the extreme case in which all modern sector capital is foreign owned and a minimum wage and hiring/firing law presents foreign capital with an all-or-nothing labor market choice. Then, imposing this legislation benefits domestic workers. In a dynamic setting the gains from redistributing quasi-rents to local workers must be balanced against potential loss of future investments.

strand of institutionalist thinking simply rejects the relevance of neoclassical analysis: "The choice between various wage policy options must not be based on the conclusions drawn from an ideal economy" (ILO 1990, p. 38). One problem with the theory is its "overbearing focus on prices" (Standing 1991b, p. 25) and neglect of other adjustments that may offer more socially desirable directions for competition (Sengenberger 1991, p. 237). When actual labor markets operate differently from the ideal, institutional modes of influencing outcomes, such as collective bargaining, tripartite negotiations, and government-mandated wages or labor standards, can be Pareto improvements. In the institutionalist view, they usually are.

Institutionalists invariably stress the benefits of interventions in the form of, say, insurance from adverse market outcomes or redistribution to low-wage workers:

The quest for labour security is a legitimate objective . . . market mechanisms may need to be circumvented in the interest of social values. . . . Social progress and labour standards should not be sacrificed in the name of efficiency. (Standing and Tokman 1991, p. 1)

The market system is a powerful tool for economic management [but] . . . there may be some side effects on economic security and equity that . . . give rise to a need for state intervention . . . [to] moderate income inequality and provide some minimum economic security. (World Bank 1991, pp. 41-42)

Standard economic analysis allows us to measure the posited benefits of interventions. One can derive demands for unemployment insurance, social security, and the like from individual maximizing behavior under risk aversion; calculate the consumer surplus attributable to programs that meet those demands; and compare the surplus with the inefficiency costs that result from labor supply responses to the program. (Risk aversion parameters may be harder to estimate than elasticities of demand or supply, but difficulty of estimation has never stopped an econometrician.) Hansen and Imrohoroglu's (1992, p. 118) simulation of the benefits and costs of unemployment insurance suggests that for the United States "replacement rates as high as .65 are optimal and the welfare benefits of unemployment insurance are quite large," although moral hazard and nonoptimal replacement rates can produce costs in excess of benefits. Whatever the result, it is important to measure rather than assert the posited benefits of interventions.

ILO support of collective bargaining is based on a moral imperative:

Countries which are members of the ILO are presumed to accept the value judgment that free collective bargaining between employers and autonomous pluralistic trade unions is the best method of determining terms and conditions of employment. Access to such mechanisms is regarded as a basic human right. Therefore, governments are expected to introduce legis-

lative provisions to encourage the development of trade unions and free collective bargaining. (ILO 1990, p. 39)

But support for collective bargaining or other institutional modes of wage-setting (such as extension of contracts by ministers of labor to nonunion workers, as in Western Europe) can also be grounded in theories of bargaining. Analyses of prisoners' dilemma games, for instance, show that bargaining partners with long horizons can reach cooperative solutions. This in turn underlies the standard argument that "neocorporatist" centralized bargaining resolves the prisoners' dilemma of wage-wage inflation at lower costs of unemployment than decentralized arrangements.

There are two arguments for labor standards or legally mandated benefits. The first asserts that standards force employers to "overcome the misguided preoccupation with cost-cutting (via lower wages), and [redirect] attention to the strengthening of productive power (via training, technical innovation, etc.)" (Sengenberger 1991, p. 249). This claim asserts but does not demonstrate that managers have a bias toward cheap labor solutions rather than toward equally—or even more—productive high-wage modes of competing. The second defense for mandated standards is that they are solutions to moral hazard or selectivity issues that make it unprofitable for firms to offer socially desirable benefits or contracts and thus are akin to lump sum users' taxes (Summers 1988). This argument resonates with standard theory of market failure.

In short, economic theory is rich (weak) enough to provide arguments for interventions as well as against them. The more the world is filled with prisoner's dilemma games, certain types of moral hazard problems, and the like, the greater is the institutionalist case. The closer the world is to the competitive ideal, the less compelling is that case. The game theory finding that modest differences in the rules of games (that is, institutions) can substantially affect outcomes implies that one cannot dismiss institutional claims as atheoretic, although the claims may be wrong. All of which means (no surprise) that we must look at evidence to decide who is closer to the truth.

II. THE SCORECARD

In this section I assess World Bank, ILO, and other research on the validity of the distortionist and institutionalist views in several areas: sectoral wage differentials, nonwage labor costs, minimum wages, wage adjustments, employment security regulations, and collective bargaining. Because each developing country has its own experiences with labor market interventions, in principle I should draw on hundreds of cases for my scorecard. Instead of essaying this herculean task, I have limited my review of studies to Bank and ILO research in the 1980s and to the smattering of countries on which I have first-hand knowledge. I recognize that cases or studies which I missed might lead to a somewhat different scoring of the debate.

Are Sectoral Wage Differentials an Indication of Distortion?

In the 1960s and 1970s there was general agreement that massive public-private, formal-informal, or urban-rural wage differentials in developing countries, particularly in Africa, proved that something was wrong with wage determination that affected the overall pattern of development. Intervention in favor of formal sector workers was an obvious culprit. But much changed in the 1980s. As Colclough (1991, table 5, appendix tables 1 and 2) and Lindauer, Meesook, and Suebsaeng (1988, table 3) document, for many countries in Sub-Saharan Africa public sector pay fell markedly in relation to gross domestic product (GDP) per capita, particularly for higher-wage public employees. With government accounting for much of modern sector employment, traditional urban-rural differentials also dropped, in some cases sharply (Jamal and Weeks 1992). In Latin America, government wages also fell significantly, falling 40 percent in real terms from 1985 through 1989 in Venezuela, for example.

The drop in public and urban salary premiums in the 1980s has several implications for the distortionist-versus-institutionalist debate. It refutes fears that institutional rigidities make pay-setting inflexible in the modern sector (of which more later) and obsolesces much distortionist concern over sectoral wage differentials and urban bias in labor market outcomes. Most important, it raises new questions about the effect of low pay for government workers on the operation of the public sector (Lindauer, Meesook, and Suebsaeng 1988). In many countries the issue for the 1990s is not how to reduce excessive government pay but rather how to build a productive public sector capable of operating social safety nets for those who lose from adjustment and stabilization programs and capable of managing infrastructure and education systems for long-term development.

Microstudies of wages in several developing countries have revealed sizable pay differentials among comparable workers that cannot be explained by state or union interventions, which casts further doubt on the distortionist interpretation of wage-setting. Summarizing studies of the Bombay labor market, Mazumdar (1989a, p. 11) concluded that "the popular (Bank) view exaggerates the role of institutional interventions in creating and maintaining this wage gap" because "long before the era of trade union or government intervention, wages in large textile factories were high in comparison with alternative earnings," and that "large wage differences are found in urban labor markets in countries (Indonesia) where the institutional apparatus for wage determination is at a rudimentary level" (p. 2).³ These findings resonate with U.S. research that shows sizable wage differentials among similar workers absent institutional interventions, in contrast to small differentials in interventionist Sweden.

3. Mazumdar (1989b, pp. 10-11) further notes the wide variation in wages among villages in India and the uniform daily wage rates for laborers with differing skills, absent unionism or government pay regulations.

To see how the dispersion of earnings across sectors varies among countries with differing levels of development and interventions in the labor market, I calculated standard deviations of log earnings among manufacturing industries using data from the ILO *Yearbook of Labour Statistics* (1991c).⁴ The results, in table 1, show that differentials are greater in low-income and lower-middle-income countries than in upper-middle-income developing countries and higher-income countries; are greater in less interventionist than in more interventionist high-income countries; and are least in formerly Communist countries.⁵ This pattern suggests that interventions reduce rather than increase differentials in the industrial sector, which certainly rules out the use of high-wage dispersion across industries as an indicator of interventionist distortions and suggests, if anything, the opposite: that low-wage dispersion can be taken as an indicator of interventionist distortion. The evidence supports the use of low dispersion as an indicator of interventionist wage policies in the industrial sector (Freeman 1988), but not necessarily as a measure of distortion. The reason is that the sizable variation in sectoral wages in relatively unfettered labor markets can be interpreted as the "failure" of those markets to establish equal pay for equal work in accord with the neoclassical model rather than as the "right" wage structure. Studies in the United States that have sought to explain wage variations in industry have generally concluded that a sizable proportion of the variation is rent-sharing because firms in sectors that do well pay above-market wages even though there is no union—or other—intervention.

Are Nonwage Costs an Indication of Distortion?

Another often-mentioned potential labor market distortion is institutionally induced nonwage costs of labor (ranging from payroll taxes and unemployment compensation to other fringe benefits). To the extent that these costs add to the competitive market cost of employment, they will reduce the number of workers in affected sectors. Exploring this possibility, Riveros (1989) found "no clear time trend in relative labor costs which would suggest that enforcement of these costs introduces (rising) distortions" (p. 19); "that in most less developed countries the existence of nonwage costs does not necessarily constitute a distortionary factor" (p. 22); and that "a certain level of nonwage costs associated with deferred payments or with an insurance system may not be all that distortionary" (p. 20). Whether the failure of nonwage costs to proxy distortions arises from the fungibility of labor costs, the quality of the data, the short time-series to which it was applied, or a more fundamental problem with the distortionist

4. These data are based on establishment surveys and are exceedingly crude because of differences in the size of units reporting in the different countries.

5. Note that I follow World Bank practice in placing China among the lower-income countries rather than the Communist countries.

Table 1. Standard Deviations of Log Earnings across Industries in Manufacturing, by Level of Economic Development

<i>Category of economy</i>	<i>Year</i>	<i>Standard deviation</i>
<i>Low-income</i>		
Burundi	1985	0.222
China	1990	0.172
Gambia, The	1987	0.473
India	1986	0.234
Kenya	1989	0.419
Malawi	1988	0.483
Myanmar	1989	0.126
Average	n.a.	0.304
<i>Lower-middle-income</i>		
Chile	1990	0.345
Colombia	1986	0.089
Dominican Republic	1985	0.460
Egypt	1987	0.330
Fiji	1987	0.351
Guatemala	1985	0.363
Mauritius	1990	0.254
Mexico	1990	0.189
Netherlands Antilles	1986	0.351
Panama	1985	0.312
Peru	1988	0.383
Philippines	1987	0.453
Turkey	1988	0.274
Zimbabwe	1987	0.313
Average	n.a.	0.319
<i>Upper-middle-income</i>		
Argentina	1989	0.179
Cyprus	1989	0.219
Greece	1988	0.143
Korea, Rep. of	1990	0.278
Portugal	1987	0.290
Puerto Rico	1990	0.211
South Africa	1988	0.353
Uruguay	1989	0.116
Yugoslavia	1989	0.212
Average	n.a.	0.222
<i>High-income, less interventionist</i>		
Canada	1990	0.245
Hong Kong	1990	0.210
Ireland	1989	0.216
Japan	1989	0.251
Spain	1989	0.243
United Kingdom	1990	0.176
United States	1990	0.267
Average	n.a.	0.230
<i>High-income, more interventionist</i>		
Austria	1990	0.228
Belgium	1989	0.175
Denmark	1990	0.114
Finland	1990	0.153
France	1987	0.140
Germany	1990	0.160

(Table continues on the following page.)

Table 1 (continued)

<i>Category of economy</i>	<i>Year</i>	<i>Standard deviation</i>
Israel	1986	0.252
Luxembourg	1989	0.196
Netherlands	1989	0.144
New Zealand	1990	0.167
Singapore	1990	0.214
Sweden	1990	0.104
Average	n.a.	0.171
<i>Centrally planned</i>		
Bulgaria	1986	0.133
Cuba	1988	0.101
Czechoslovakia	1989	0.109
Hungary	1990	0.216
Poland	1989	0.141
Romania	1989	0.108
Ukraine	1989	0.095
U.S.S.R.	1989	0.100
Average	n.a.	0.125

n.a. Not applicable.

Source: Calculated from ILO (1991c).

view is debatable. What is clear is that the exercise did not yield a serviceable measure of distortions for cross-country comparisons.⁶

The inference I draw from evidence on interindustry wage dispersion and nonwage costs is that there is no easy way to measure interventionist distortions using market price data. What is needed instead is to look at specific government programs that intervene in wages or employment, to which I turn next.

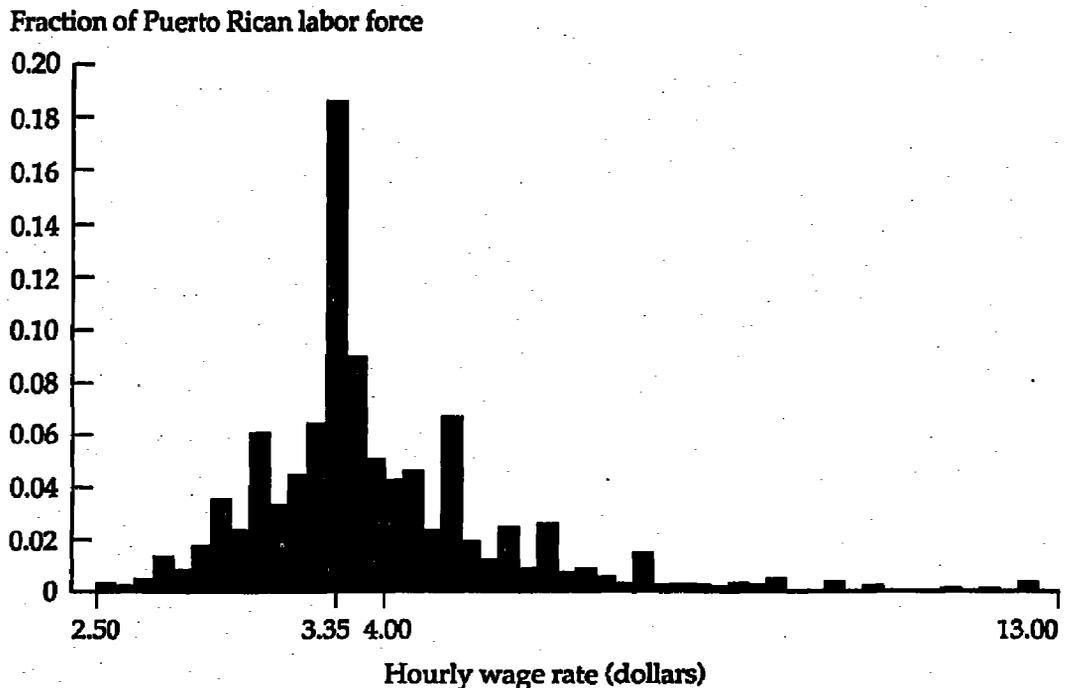
Minimum Wages

The minimum wage is a *bête noire* to distortionists because it is the textbook case of an intervention that misallocates resources: an effective minimum wage reduces employment. The major question is whether the induced increase is worth the loss of jobs. If it does raise the wages of the most poverty-stricken at little cost to employment, many would find this an appealing way to redistribute income. If, by contrast, the cost is many jobs, and only a few highly paid formal sector workers benefit at the expense of lower-paid informal or rural sector workers, few would favor minimum wage policies. What does the evidence show?

There is evidence that an enforced minimum wage substantially reduces employment. Consider, for example, the application of the U.S. minimum wage to Puerto Rico, where productivity and earnings are considerably below mainland levels. As figure 1 shows, the U.S. minimum of \$3.35 an hour essentially dominated the distribution of earnings in Puerto Rico in 1983, creating a

6. Even so, nonwage costs turn up as an indicator of distortions in other studies (López and Riveros 1990), presumably for lack of anything better, and give weak results comparable to those in Riveros (1989).

Figure 1. Effect of U.S.-Level Minimum Wage on the Puerto Rican Labor Market: Distribution of Hourly Earnings, 1983



Note: U.S. minimum wage = \$3.35 an hour.
Source: Castillo-Freeman and Freeman (1991).

remarkable spike at that rate of pay. Analyses by Castillo-Freeman and Freeman (1991) show that the imposition of the minimum wage raised average earnings on the island, lowered the aggregate employment-population ratio by a significant amount, and shifted employment away from low-wage sectors (which had to raise pay substantially to meet the minimum).

But such minimum wage intervention is far from the norm in the developing world. Many countries set minimum wages too low or are too lax in enforcing the law for the regulation to have much effect.⁷ In an assessment of Mexico's minimum wage law, Gregory (1986, pp. 260-61) concluded that "the relationship of legal minimum wages to market-determined wages has evinced frequent and substantial changes at different times . . . and increases in the former were not a necessary precondition for raising real wage levels of those employed toward the lower end of the urban wage distribution." Similarly, Fallon's (1987, pp. 7-8) study of labor regulation in India rejected the importance of minimum wages: "unskilled wages were substantially above minimum rates in large establishments . . . [implying] that the latter are also ineffective. In smaller establishments . . . most firms paid at or within 20 percent of the minimum . . . con-

7. Noncompliance rates in Mexico are 25 percent; in Costa Rica 20 percent; in Jakarta, Indonesia, 70 percent (ILO 1990, p. 27). See also Starr's (1981, pp. 138-41) study of minimum wage fixing.

sistent with the view that advisory boards use going wage rates in small establishments as the basis for setting minimum rates in the first place.”

By contrast, Fallon and Lucas (1991, p. 397) regard the large increases in minimum wages in Zimbabwe after independence as substantially affecting the wage structure. Paldam and Riveros's (1987) review of minimum wages in Latin America reports mixed effects of the minimum on wages and a “lack of causal connections between minimum-wage changes and aggregate employment” in Chile, the one country for which they analyzed the relation between the minimum and employment (p. 26). They conclude that the “existence of minimum wage causes aggregate effects only when it is used aggressively as a policy tool” (p. 1). In a less structured regression analysis, ILO researchers report statistically insignificant relations between changes in real minimum wages and changes in real average wages in the 1970s and 1980s in fourteen African and Latin American countries, leading them to reject any impact from the minimum in these cases (ILO 1990, table 21).⁸ Additional work by López and Riveros (1989) on the effect of minimum wages on skilled and unskilled workers in Latin America raises doubts about the value of World Bank or ILO exercises based on limited time-series: the regressions suggest that minimum wages raised the wages of skilled workers in Argentina; reduced their wages in Chile, Colombia, and Uruguay; and had weak effects on the wages of unskilled workers in all cases (López and Riveros 1989, tables 1 and 2)—a pattern of results that does not make much economic sense.

What negates concern that minimum wages have in fact been highly distortionary, however, is not the weak results from multivariate regressions but rather the evidence that when push came to shove in the 1980s, real minimums fell precipitously in many countries (see table 2). The minimum floor proved to be sawdust—not hardwood, as distortionists feared.

Does evidence that the minimum wage (and possibly other labor market interventions) rarely distorts the labor market seriously also mean that these policies fail to accomplish their institutionalist goal of providing protection for workers? In part, it does point out their limited effect. But there is another way to interpret the evidence that I think is more useful. This is to view the interventions as endogenous to economic conditions, and thus sensitive to their costs and benefits, rather than as exogenously given. From this perspective, countries will rarely set minimum wages at levels that cut seriously into employment. If extensive unemployment results, the minimum will often be unenforceable because both workers and employers will have incentives to collude to avoid the law and save jobs. I hypothesize that in many cases countries follow a strategy of “optimal selective enforcement” of minimum wage (and other) regulations; that is, they effectively implement these laws to

8. The ILO and Bank studies overlap for four countries. The ILO regressions show no effect in Argentina, Brazil, Chile, and Mexico. The Bank study finds little effect in Argentina but effects in the other countries. The number of years covered and the mode of statistical analyses vary.

protect workers when the cost is modest in terms of lost jobs (demand is high in the labor market; employers have low elasticity of demand for labor), but they enforce the laws weakly when the cost of employment is sizable (demand is low; elasticity is high). Such a story is consistent with observed experience and points to the possibility of a more formal "political economy" model of optimal minimum wage-setting and enforcement.

Job Security and Other Employment Regulations

Job security regulations require firms to gain the approval of government or other institutions (such as works council in much of Western Europe) for layoffs and in some cases mandate high severance pay. Such regulations raise the costs of reductions in staff and potentially affect the speed of adjustment and total employment. Evidence on the effects of these provisions for developed countries, where they are strictly enforced, is mixed. Houseman (1991) found that strong job guarantees in continental European countries resulted in smaller job losses in the declining steel industry than in the laissez-faire United Kingdom. Lazear (1990) found that mandatory severance pay reduced employment across Organization for Economic Cooperation and Development (OECD) countries, although the results were "not especially robust to specification" (p. 725) nor to before-and-after comparisons for the same country. For developing countries, Fallon and Lucas (1991) estimated wage and employment adjustment equations before and after passage of job security laws in India and Zimbabwe and found little evidence that the laws affected wages or speeds of adjustment but considerable evidence that they reduced total employment in relation to output—an odd finding, since job security provisions that do not affect wages or the speed of adjustment carry no extra cost that would deter employment. On the other side, Standing (1989, pp. 46–48) reports that almost all firms in a recent ILO survey stated that a comparable Malaysian job security law had no impact on employment.

While the results of these studies are equivocal, Spain's experience with job security regulations provides a strong case in which relaxation of regulations spurred job growth. In 1980 the government introduced a fixed-term employment contract as an alternative to permanent contracts (that dated back to Franco's dictatorship), and in 1984 it enlarged the fixed-term contract option. The result was a growth of aggregate employment, consisting almost exclusively of persons on fixed-term contracts, beyond what was likely on the basis of past productivity trends and output expansion. But employment growth is not the full story, for, as one would expect from human capital theory, workers under fixed-term contracts appear to get less training than permanent employees, which bodes poorly for their future (Alba-Ramirez 1991). Marshall (1991) also found that labor laws affect the kind (although not necessarily the volume) of employment and concluded that temporary and part-time work was more common in Lima than in Buenos Aires when Peru encouraged temporary contracts to reduce unemployment while Argentina did not.

Table 2. Indices of Real Minimum Wages and Real Average Earnings in Selected Developing Countries

<i>Country</i>	<i>Real minimum wage</i> (1980 = 100)	<i>Real manufacturing wage</i> (1980 = 100)	<i>Real nonagricultural wage</i> (1979 = 100)
	1989	1978-88	1984
<i>Latin America</i>			
Argentina	68.4	94.0	140
Bolivia	—	—	48
Brazil	68.6	106.8	108
Chile	63.5	99.3	120
Colombia	105.0	119.3	94
Costa Rica	110.2	85.0	—
Dominican Republic	87.4 ^a	—	—
Ecuador	43.7	74.7 ^b	96
El Salvador	35.6	68.1 ^b	—
Guatemala	78.5 ^a	77.3	136
Haiti	97.6	—	—
Honduras	74.0	92.0	—
Mexico	46.9	64.7	75
Panama	99.8	—	—
Paraguay	136.5	107.9	94
Peru	23.2	89.2	106
Uruguay	78.6	117.3	75
Venezuela	77.1	—	77
	1985-86	1980-86	1984
<i>Africa</i>			
Algeria	85.0 ^b	—	—
Botswana	92.1	—	—
Burkina Faso	88.9	—	—
Burundi	115.0	122.7	117
Cameroon	108.0	—	—
Central African Republic	57.5	—	—
Congo	62.2	—	—
Côte d'Ivoire	78.6	—	—
Ethiopia	77.0	—	—
Gabon	86.8	—	—
Gambia, The	65.0	—	—
Ghana	154.6	140.6	44
Guinea	63.6	—	—
Kenya	58.8	77.2	85
Liberia	83.0	—	—
Madagascar	64.0	—	—
Malawi	120.1	61.9	73
Mali	91.4 ^b	—	—
Mauritius	76.0	103.7	—
Morocco	117.9 ^b	55.0 ^b	—
Niger	77.0	—	—
Nigeria	79.0	89.0 ^a	—
Rwanda	73.4	—	—
Senegal	74.3	—	—
Sierra Leone	—	29.8	46
Somalia	16.1	—	—
Sudan	44.2	—	—

Table 2 (continued)

Country	Real minimum wage (1980 = 100)	Real manufacturing wage (1980 = 100)	Real nonagricultural wage (1979 = 100)
	1985-86	1980-86	1984
Tanzania	32.8	74.0 ^b	57
Togo	76.7	—	—
Tunisia	110.4	—	—
Zaire	112.0	—	—
Zambia	80.5	58.9	79
Zimbabwe	—	97.6	101

— Not available.

Note: These data are highly suspect because of the size of firms covered and the price indices.

a. 1988.

b. Other years.

Source: ILO (1990); Standing and Tokman (1991, p. 213).

Government Employment

High and increasing government employment in some developing countries in the 1970s raised concerns that a large public sector might itself be a major distortion in the labor market. These concerns are forcefully exhibited in Gelb, Knight, and Sabot's (1991) nightmare scenario of the public sector as a sinkhole of waste. Their simulations show that under some conditions a bloated government sector can choke off productive employment and economic growth and suggest that distortionists should concentrate more on the number of public employees and their job activities than on pay differentials. This warning recalls some African experiences, where reductions in government budgets generally took the form of lower pay rather than lower employment, with disastrous effects on public sector competence (Lindauer, Meesook, and Suebsaeng 1988).

There is, however, no economic law that public employees are nonproductive. Malaysia's experience of rapidly growing public sector employment in a period of economic expansion serves as a fruitful counterexample. More generally, Kormendi and Meguire (1985) report that growth of public sector spending in relation to output across countries is uncorrelated with growth of per capita income. I interpret this as indicating that while government employment beyond some level may prove disastrous, few countries let things get that out of hand. Here, as with minimum wages, there are presumably political checks and balances that limit the distortionist nightmare from becoming reality.

Wage Adjustments

In the 1980s the sluggish world economy and the debt crisis were major tests for the labor markets of developing economies. Did institutions obstruct stabilization and adjustment programs in accord with distortionist dogma? Did countries that relied more on consensual modes of adjustment fare better than others in accord with institutionalist dogma?

Table 3. A Numerical Model of Declining Support for a Beneficial Economic Reform Program for Five Periods of Time after the Reform

<i>Period</i>	<i>Winners</i>	<i>Losers</i>	<i>Expected gain</i>	<i>Percentage in favor of reform</i>
0	—	—	0.10	100
1	0.25	-0.75	-0.50	25
2	0.44	-0.56	-0.12	44
3	0.58	-0.42	0.16	58
4	0.69	-0.31	0.38	69
5	0.77	-0.23	0.54	77

Note: The reform gives winners a gain of one unit in each period and costs losers one unit in each period. Of the population of losers, 25 percent advance to the winners' group in each period. The discount rate is 0.9.

In period 0 the present value of gain is $-0.5 - 0.11 + 0.13 + 0.28 + 0.35 = 0.10$. Support is 100 percent.

After period 1 the discounted gain for losers is $-0.5 - 0.11 + 0.13 + 0.28 = -0.20$. Support is 25 percent (winners only).

After period 2 the discounted gain for losers is $-0.5 - 0.11 + 0.13 = -0.48$. Support is 44 percent (winners only).

After period 3 the discounted gain for losers is $-0.5 - 0.11 = -0.61$. Support is 58 percent (winners only).

The answer to the first question is no. At a crude level, the sharp drops in real wages shown in table 3 refute distortionist fears that labor market institutions or interventions produce wage rigidity when declines are necessary.⁹ Detailed studies of labor markets in twelve countries (Horton, Kanbur, and Mazumdar 1991, p. 17) confirm this reading of the data. In Costa Rica wage indexation rules contributed to rapid downward adjustment of real wages during inflation. In Bolivia the elimination of much labor legislation did not produce economic recovery. Analysis of microsurvey data in Côte d'Ivoire shows further that aggregate wages can significantly understate real wage flexibility by failing to allow for compositional changes associated with reduced employment (Levy and Newman 1989). Between 1979 and 1984 aggregate real wages in Côte d'Ivoire rose 17.5 percent, apparently contributing to the loss of modern sector employment that accompanied structural adjustment. But research based on microdata that adjusts for changes in the skill composition of the work force shows that real wages corrected for changes in skill composition fell 8 percent! The misleading aggregate data failed to take into account the fact that the least skilled were more likely to lose their jobs, biasing upward the average wage. In fact, the disaggregated data show considerable downward flexibility of wages among new hires. Fallon and Riveros (1989, p. 23), who studied sixteen countries, including Latin American and African countries whose labor institutions are often severely criticized, concluded that "there is little prima facie evi-

9. I recognize that the real wage data used in this table suffer from potential problems with the universe surveyed and the price indexes. In some cases the underlying data show sharp fluctuations in real wages in short periods that must be at least in part due to data problems. Still, the overall pattern shown for so many countries is unlikely to be the result of flaky data.

dence on downward real wage rigidity as often suggested by opponents of exchange rate devaluation."

To what extent ought the finding of real wage flexibility be modified by changes in unemployment in developing countries in the 1980s? Institutional interventions may not have prevented real wages from falling during economic declines, but perhaps they produced a suboptimal rate of reduction with accompanying open unemployment. There is evidence that open unemployment (admittedly poorly measured) rose in several developing countries in the 1980s (see, for instance, Vandemoortele 1991) and that employment in the informal sector grew substantially. Nevertheless, to argue that the solution is even greater real wage reductions than those shown in table 2 for many countries seems excessive, because it puts the entire burden of adjustment to macroeconomic distress on wages and the labor market. When unemployment rises and real wages fall in industrial countries with little institutional intervention, as in the United States, no one calls for real wages to fall more rapidly; the response is to seek ways of expanding the economy or augmenting the skills of workers. When the reduction in real wages necessary to eliminate open unemployment exceeds the huge reductions observed in many developing countries, I would look beyond the labor market for the root cause of the economic disaster.

Turning to the tripartite forums or social pacts favored by institutionalists, I am unable to judge whether or not they improve the economic or social face of adjustment and stabilization. Such arrangements played a substantial role in reducing real wages during the 1980s in Belgium and Australia, among other OECD countries, but I am unfamiliar with studies assessing the role of these arrangements in developing countries. It would be valuable to see how social pacts such as the Moncloa Pact and ensuing social accords in Spain or the Pacto Solidaridad Económica in Mexico actually work and whether they contribute in an important way to the adjustment process. My limited knowledge of tripartite forums in the marketizing economies of the East makes me suspicious of claims that these institutions are all that important, but the situation could be different elsewhere.

Collective Bargaining

The success in the 1980s of the East Asian economies that suppressed or severely restricted unions (Korea, Singapore, Malaysia, and Taiwan, China) and the mid-1980s success of Chile (after a long period of economic failure under military rule) raises the nasty question of whether suppressing unions contributes to economic growth. No study has dealt head-on with the question for these countries, though Lindauer and others' (1991) analysis of the labor market in Korea shows that suppression of labor was associated with high accident rates and produced a remarkably disgruntled work force despite large gains in real wages. The experience of advanced and developing countries in general does not sustain any generalization that less unionism means more growth but rather shows that unions are no impediment to rapid economic

development. Japan and Germany, in particular, have had outstanding growth records with labor institutions that are a far cry from laissez-faire ideals. The poor performance of the U.S. economy in the 1980s, when the private sector was largely nonunion compared with the 1950s and 1960s, also shows that low levels of unionism are no guarantee of economic success. Studies on the relation between unions and adjustment in Horton, Kanbur, and Mazumdar (1991) find that union responses to adjustment programs range from militant opposition to active cooperation and that the strength of unions need not bear any simple relation to the prospects for recovery (p. i). In particular, "weakening the unions (as in Bolivia) does not seem to be sufficient to ensure recovery" (p. 55). Their conclusion? A warning that more detailed examination of the role of unions and other labor market institutions is needed "before launching into a wholesale advocacy of dismantling such institutions" (p. 57).

The ILO's World Employment Programme has undertaken enterprise-level surveys on the microeffects of unions on wages, mobility, flexibility, training, and productivity. The results from 3,000 establishments in Malaysia (Standing 1989, 1991a) show that unionism is associated with wage and nonwage outcomes similar to those found in industrial countries (Freeman and Medoff 1984): higher wages and reduced employment growth (the standard neoclassical monopoly effect); smaller wage differentials within enterprises, lower turnover, greater fringe benefits, higher productivity (standard "voice" effects); and more job training. The analysis also reveals that industrial unions have greater effects on some outcomes and smaller effects on others than weaker "house" or company unions. The overall effect of unions is positive, despite the welfare triangle losses from higher wages and lower employment.

In sum, extant studies reject the proposition that unions are a general impediment to macroadjustments or to enterprise performance in developing countries, although they may be so in particular cases, such as in Peron's Argentina.

III. LABOR INSTITUTIONS AND THE POLITICAL ECONOMY

The design of adjustment programs should take into account the political support necessary to sustain the program . . . Compensatory measures, such as severance pay and job retraining, should encourage exit from groups that oppose reforms and entry into groups that benefit from (and will support) the program. (World Bank 1990a, p. 8)

A considerable body of experience points to the crucial importance of political and institutional factors in determining the success or failure of structural adjustment programs . . . Where this (a sound labor relations system and a commitment to tripartite dialogue) is not the case . . . the consequences have been popular protests . . . governments have abandoned the adjustment programme, or they have lost power. (ILO 1991b, p. 2)

There is a growing awareness among World Bank and ILO analysts that labor market institutions and policies play a more complex political role than recog-

nized in the distortionist-institutionalist debate. Consonant with this position, I offer below a model of how labor market interventions influence attitudes toward reform programs and modes of expressing those attitudes. The discussion, which is based on Freeman (1992), is more speculative than that in the preceding sections.

Time Pattern of Benefits and Costs

Consider an economic reform that pays off in the future but that costs workers in the present. For simplicity, assume that workers initially receive numeraire wage 0 and that the program creates two classes: winners, who earn $W (> 0)$ after they attain that status; and losers, who earn $-L (< 0)$. Assume further a transition probability of p per period for moving from the losing to the winning group. Under these conditions the value of the reforms in year t will be:

$$(1) \quad pW \sum (1-p)^i - L(1-p)^t = W - (W+L)(1-p)^t$$

where the summation is from $i = 0$ to $t - 1$. In continuous time, we have

$$(1)' \quad -L \exp^{-rt} + W(1 - \exp^{-rt}) = W - (W+L) \exp^{-rt}$$

which is negative at low values of t ($= -L$ in year 0) but approaches W as t rises. The present value of the change from 0 to ∞ at discount rate r is:

$$(2) \quad W \int (\exp^{-rt} - (W+L)) \exp^{-rt} = [(pW - rL)/r(r+p)]$$

which must be positive for the program to be worthwhile. I assume that t goes to ∞ for algebraic simplicity and vary r to reflect different lengths of working time.

The present value model provides a framework for considering the pattern of support for reforms among workers and over time. Older workers have few years to reap benefits, so r will be high for them, implying that they will be less supportive of reforms than younger workers. More interesting, equation 2 shows that workers may prefer a program that generates more inequality of earnings ($W - L$) to one that generates less inequality. They will prefer greater inequality when their chance of becoming a winner exceeds their discount rate ($p > r$), since they then benefit more from high future W than from lower current L . This is a variant of Hirschman's (1973) "tunnel effect," according to which losers in the early phase of growth tolerate rising inequality because they view the gains of others as a sign of future gains for them.

Consider next how support for reforms changes over time in a fixed homogeneous population. Initially everyone favors the program ($2) > 0$. In period one there are p winners and $1 - p$ losers. Winners continue to favor the program, but the present value of benefits falls among losers because they have fewer years to reap the rewards (in the infinite horizon model, r rises). In period two there are $p + p(1 - p)$ winners and $(1 - p)^2$ losers, whose present value of benefits drops further. At some period T the present value turns negative for losers, potentially producing massive opposition (see the example in table 3), after

which support rises as p percent of the remaining losers move into the winning group. The critical period for the reform program occurs when support bottoms out. If more than 50 percent of the population turns against the program, a democratic government might back away from reforms—even though the program has, by assumption, a positive payoff.

The potential for erroneously rejecting reforms will be enhanced when personal experiences influence an individual's expected transition probability. If each person updates his expected p along Bayesian lines, losers will continually revise downward their present value assessment of the program. If people have different unknown transition probabilities, losers with high ps may mistake bad luck for low ps and erroneously place themselves in the low- p group. Similarly, random shocks create the danger that some will misinterpret a bad draw (world economic slowdown; changed terms of trade) for a bad program and reject reforms. All of which strengthens the point of table 3—that support for reforms will follow a U-shaped curve.

What happens if we extend the analysis to a changing labor force, with new cohorts favorable to reforms entering the labor force and older cohorts leaving the labor force to become pensioners in each period? The influx of new workers has the potential for counterbalancing the loss of support among existing workers, modifying the U-curve of support. If all pensioners (including those who gained from reforms) oppose the reforms because they reduce the real value of pensions, however, this may offset the rising support of new workers, so that the relative sizes of the two groups will affect the analysis. But perhaps some pensioners support reforms because their children are likely to benefit or because they foresee increased pensions with successful reforms. To deal sensibly with these and other possible problems (for instance, likely declines in support for reforms in a given cohort when winning is not an absorbing Markov state) the model must be made more complicated. In principle, one can derive separate U-curves of support for various age cohorts under differing assumptions and then sum them to get an aggregate curve of support for the population. As with other issues of aggregation, the weights on the groups will help determine the overall shape of the support curve.

Rather than expanding the model (see Freeman 1992), however, I turn to the more salient issue of how labor institutions and policies can influence the attitudes that underlie the support curve and the actions that those attitudes may precipitate.

Interventions and the Benefits and Costs of Reforms

The most straightforward way for interventions to affect attitudes is through side-payments to losers that alter the benefit-cost calculation. The U-shaped curve of support suggests that the timing of payments may be critical. Interventions will be most valuable when support bottoms out and may be least effective in preserving reforms when they are spread over time (or, what may be worse, if they decline over time as the fiscal costs of interventions become

clearer). With respect to specific interventions, job training and active labor market programs that increase employability are undoubtedly preferable to straight "bribes" or subsidies that keep alive unprofitable enterprises, but the latter may still be worthwhile if they buy additional time for painful reforms. In the United States trade adjustment assistance to workers who lost their jobs because of trade did little to promote employability and may have reduced labor mobility, but this was a small price to pay for additional free trade. Consistent with the notion that transfers may be a price for certain economic policies, Bates, Brock, and Tiefenthaler (1991) present crude data that countries with larger per capita transfer payments programs have more open economies.

Going beyond government programs, an alternative way to attract support is to give losers institutional power to defend their interests in the postreform world—for instance, collective bargaining rights for workers whose market pay falls but who may be able to negotiate a "share" of gains through union activity. Profit-sharing or distribution of stocks or national bonds to workers in firms undergoing privatization can also offer losers options to benefit from the future gains of reform even if they are likely to do poorly in the postreform competitive market. Since side-payments must be paid from taxes (inflation), they will lower the benefits to winners (and expected winners) and extract a deadweight loss from society as a whole. This means that buying support for programs through social or labor market interventions has a clear danger: the payments may build up distortionist inefficiencies that abort the reforms. The benefits of interventions in the form of higher tolerance for the costs of reforms must be weighed against the distortionist costs of the interventions. By normal diminishing-productivity arguments, the issue is not one of whether to intervene but rather of how much and in what ways to intervene to give losers some possibility of making gains and thus buttressing support for the reforms.

Labor relations institutions can also influence expectations of gains from reforms. In a world with heterogeneous labor, Hirschman's tunnel effect will work only if losers see persons like themselves benefiting from reforms. This suggests the virtue of unions that include private and public sector workers, not, as in East bloc marketizing economies, unions concentrated in (largely losing) state enterprises. Similarly, policy (collective bargaining) might spur profitable enterprises to share economic rents with workers during the initial phase of reform so that there is a clear example of workers' benefiting from the gains. This thrust is consistent with recent World Bank efforts to encourage governments to package reforms to produce identifiable benefits and create public support for the broader reform effort.

Influencing Reforms through Protest and Voice

When workers decide, rightly or wrongly, that reforms are undesirable, there is a danger that they will protest and attempt to overturn the program. In my model successful protests at the bottom point of the support curve risk a self-

fulfilling prophecy of failure: if people had greater tolerance for the costs of transition, the program would work as planned, but if losers have sufficient power to protest, the program fails, possibly producing a Latin American-style populist policy cycle (Sachs 1990). One way to reduce the likelihood of such a scenario is through labor policies that limit freedom of association or the ability to stage a broad strike—for instance, through laws that encourage enterprise-level unions and discourage broader union groupings. A more extreme possibility is to suppress unions for some period. If Korean-style suppression of labor could guarantee 6 to 8 percent annual growth in real wages for two decades, many developing economies would sign on. Although there are no studies linking the suppression of unions to growth, most recent work shows that dictatorships (which usually suppress unions) have lower or no higher per capita growth or success in adjusting than democracies (which invariably permit free unions). (See Scully 1988; Kormendi and Meguire 1985; Renner 1986; Weede 1983; and Haggard and Kaufman 1990.)

The polar opposite to weak or suppressed unionism is an all-encompassing union organization that negotiates “tripartite pacts” or neocorporatist centralized wage-setting arrangements with business and government. All-encompassing unions presumably internalize distortionary costs in favor of a broad national economic perspective (Olson 1982). Empirical analyses suggest that they worked well in certain time periods in industrial countries (Bruno and Sachs 1985; Calmfors and Driffil 1988; Freeman 1988). Such systems are not easy to institute or maintain, however, as the ongoing breakdown of cooperative centralized arrangements in Sweden shows. They require a strong labor movement, with leaders able to assess the economic scene and convince workers to accept current consumption losses for future gains; a business community that accepts labor as a social partner; and a government willing to share some prerogatives with its social partners.

Finally, labor institutions can contribute to a reform program by providing social feedback on program outcomes. Even with the best intentions, governments following World Bank and International Monetary Fund (IMF) advice may blunder in the specifics of stabilization and adjustment programs. Inflation costs may be greater than expected. Unemployment and output losses may be bigger. Workers, pensioners, or children may suffer more than anticipated in the short run. If technocrats and politicians are more attuned to the world financial community than to local realities, they may be slow to realize that things are not working and thus to make adjustments. The greater the uncertainty about the success of reforms, and the more removed government officials are from the lives of the citizenry, the greater is the need for independent groups to provide feedback about the real effects of programs and to pressure politicians to make changes. The same unions and business groups that from a rent-seeking perspective endanger reforms can, from a social perspective, contribute to the program's success. More abstractly, “winner's curse” considerations, whereby more optimistic assessments of the outcome of reforms (rather than gloomier assessments)

take precedence in political debate, suggest that nearly all reforms will have greater short-run costs than expected and make the feedback from labor and management critical in correcting errors.¹⁰

Lacking detailed studies of specific reform programs to test the validity of these ideas, I can only offer some examples where a political economy analysis seems relevant. One example is Venezuela, whose adoption of standard IMF and World Bank policy reforms in the 1990s sufficiently unnerved the population to produce major riots and an attempted military coup in 1992. Prior to its reform program, Venezuela's economic policies fit an interventionists' nightmare; the government wasted the bonanza of high oil prices, interfered in the economy in all sorts of ways, and brought the country to near economic ruin. But the short-run costs of the standard prescriptions were greater than anticipated, in part because of the sluggish response of the private business sector to the new economic environment and the inability of an ineffective and underpaid bureaucracy to implement social interventions to buffer the costs of the reforms. In addition, the president and the reformist technicians were unable to enlist the support of the population for the reforms. This set the stage for loss of support as time proceeded, and for some requisite backtracking.

Zambia's failure to stick with its 1985 IMF stabilization package has been attributed in a Bank report to the "unrealistic assumption that the majority of middle- and lower-income urban Zambians would tolerate pauperization" (Colclough 1989). In other words, the government failed to give adequate consideration to political factors. An ILO paper argues that the case of Zambia shows how faulty assumptions about the labor market led to the failure of adjustment (Vandemoortele 1991, p. 84). In the marketizing economies of Eastern and Central Europe and the former U.S.S.R., a failure to alleviate social costs or to develop appropriate labor market institutions may be prove to be the Achilles' heel of economic reforms.

IV. CONCLUSIONS

There were three surprises to me in preparing this paper.

The first was that studies designed to support the distortionist view of labor markets in developing countries failed to make a stronger empirical case than they did. Part of the problem is the lack of adequate measures of distortions, and part is the excessive attention given to limited time-series data as opposed to detailed studies of worst-case situations. More can be learned, in my opinion, about which interventions are excessive or disastrously implemented from detailed case studies than from cross-country time-series regressions with weak data. If the Uruguayan social security retirement system is the economic disaster

10. The expression "winner's curse" refers to the problem in auctions, in which the person with the most optimistic view of the value of the good will win the auction. As long as the average view of the value is correct, the winner necessarily pays more than the good is worth. Similarly, in political discourse, the reformer who has the most optimistic view will make the biggest promises—"reform will bring nirvana in two years"—and potentially win the policy debate.

that some claim, the distortionist lesson to be drawn is not that social security systems are bad, but rather that systems should not be developed along Uruguayan lines. This research problem aside, the principal reason for the weak distortionist case has to be that declines in the 1980s in real wages—and changes in relative wages—in many developing countries showed that many distortionist interventions were paper tigers at crunch time. It is ironic that distortionists, who generally revere unfettered markets, understated the power and flexibility of labor markets to overcome potentially inefficient interventions.

In hindsight, I should not have been all that surprised at the weak empirical case for the distortionist view nor at flexible real and relative wages in developing country labor markets. Research on labor markets in industrial countries shows that labor markets work tolerably well and that real wages are flexible downward under diverse institutions, ranging from decentralized U.S. labor markets to centralized Swedish or Australian wage-setting. From a distortionist perspective, German labor relations, with strong unions and government extension of contracts, mandated works councils with veto power over some enterprise decisions, worker representatives on boards of directors, and so on, ought to make that country one of capitalism's basket cases. Similarly, imagine what a full distortionist critique would say about the Japanese labor market, with its idiosyncratic institutions and practices. But the German and Japanese economies work quite well. No labor market works exactly according to simple neoclassical models, but most respond reasonably well to shifts in market conditions.

The second surprise is the paucity of studies on two major claims of the disputants: the distortionist claim that labor market interventions impair investment or growth, and the institutionalist claim that consultative modes of decisionmaking or collective bargaining are superior to less structured labor market modes of adjustment. Some Bank researchers have begun to buttress the distortionist model in ways that address the former issue. López (1991b) and Gelb, Knight, and Sabot (1991) show under what assumptions the public sector can kill growth. But much more is needed, both conceptually and empirically. Olson (1982) and Kendix and Olson (1990) offer some evidence linking indicators of institutional rigidities to unemployment rates and growth of per capita income in industrial countries, but more is needed for developing countries. Given the importance of infrastructure and investments in education on growth, we need to examine how these interventions fare in the new economic environment of stabilization and adjustment programs (Birdsall and James 1990).

On the other side, I found little information for assessing the possible role of tripartite decisionmaking bodies and social pacts in adjustment in developing countries and thus had to rely on industrial countries to see such arrangements at work. Because many of the marketizing economies of Eastern and Central Europe, including Russia, have instituted such organizations, it is important to determine whether they can serve useful functions in the absence of a strong private sector. Studies of Spanish and Mexican social pacts and of tripartite and consultative decisionmaking are needed to assess the value of these institutional

interventions or to suggest other institutional mechanisms that could enlist private bodies into the reform process.

The third surprise was the most pleasant: finding convergent World Bank and ILO interest in the political role of labor interventions in economic reforms. To the extent that this interest reflects problems observed in implementing reforms in the 1980s, it is the best possible empirical support for this paper, though not necessarily for my specific arguments or model. There is an exciting practical research agenda here: determining the conditions under which the political economy advantages of interventions outweigh potential interventionist costs; determining which interventions are most effective in buttressing support, and which are likely to lead to more economic troubles; and determining which reforms outside the labor market are worth interventions in that market. The presumption underlying my analysis is that the benefits from more open economies, convertible currency, stabilization, and the like dwarf the costs of labor market interventions. Is this correct? Looking at labor market interventions as part of the political economy of reform suggests a very different research agenda than that reviewed in my scorecard.

Finally, I was struck by the extent to which views of labor market interventions seem grounded not so much on models or econometric evidence but on observation of specific country experiences. I think it is no accident that the institutionalist perspective comes from Western Europe, where Germany, Austria, Scandinavia, and others provide examples of reasonably successful institutional interventions, whereas the distortionist perspective comes from the Americas, where analysts contrast the largely unfettered American economy with state interventions in Latin America. If I am correct that first-hand experiences or specific cases have greater salience than econometric modeling, research on labor policies and institutions would benefit from more detailed investigations of how specific interventions and institutions work in particular countries as opposed to aggregate statistical analyses. Certainly specific studies are a necessary first step toward making valid generalizations that take account of idiosyncracies that allow some interventions and institutions to work in some places but not in others, and thus to draw lessons across country lines.

REFERENCES

- Alba-Ramirez, Alfonso. 1991. "Fixed Term Employment Contracts in Spain: Labor Market Flexibility or Segmentation?" Paper delivered at EALE Conference, Spain, September 1991.
- Bates, Robert, Philip Brock, and Jill Tiefenthaler. 1991. "Risk and Trade Regimes: Another Exploration." *International Organisation* 45(1):1-18.
- Becker, Gary. 1985. "Public Policies, Pressure Groups, and Dead Weight Costs." *Journal of Public Economics* 28(3):329-47.
- Birdsall, Nancy, and Estelle James. 1990. "Efficiency and Equity in Social Spending: How and Why Governments Misbehave." Policy Research Working Paper 274. World Bank, Latin America and the Caribbean Regional Office, Country Department I, Population and Human Resources Operations Division, Washington, D.C.

- Bruno, Michael, and Jeffrey Sachs. 1985. *Economics of Worldwide Stagflation*. Cambridge, Mass.: Harvard University Press.
- Calmfors, Lars, and John Driffil. 1988. "Bargaining Structure, Corporatism, and Macroeconomic Performance." *Economic Policy* 6:13-61.
- Castillo-Freeman, Alida, and Richard Freeman. 1991. "When the Minimum Wage Really Bites: The Effect of the United States-Level Minimum on Puerto Rico." *Proceedings of the Industrial and Labor Relations Association*. Madison, Wis.: Industrial Relations Research Association.
- Colclough, Christopher. 1989. "The Labor Market and Economic Stabilization in Zambia." Policy Research Working Paper 222. World Bank, Country Economics Department, Macroeconomic Adjustment and Growth Division, Washington, D.C.
- . 1991. "Wage Flexibility in Sub-Saharan Africa." In Guy Standing and Victor Tokman, eds., *Towards Social Adjustment*. Geneva: International Labour Office.
- Fallon, Peter. 1987. "The Effects of Labor Regulation upon Industrial Employment in India." Discussion Paper 287. World Bank, Development Research Department, Washington, D.C.
- Fallon, Peter R., and Robert Lucas. 1991. "The Impact of Changes in Job Security Regulations in India and Zimbabwe." *World Bank Economic Review* 5(3):395-413.
- Fallon, Peter R., and Luis A. Riveros. 1989. "Adjustment and the Labor Market." Policy Research Working Paper 214. World Bank, Country Economics Department, Macroeconomic Adjustment and Growth Division, Washington, D.C.
- Freeman, Richard. 1988. "Labour Market Institutions and Economic Performance." *Economic Policy* 6:63-80.
- . 1992. "What Direction for Labor Market Institutions in Eastern and Central Europe?" NBER conference on "Transition in Eastern Europe," Cambridge, Mass., February 26-29.
- Freeman, Richard, and James Medoff. 1984. *What Do Unions Do?* New York: Basic Books.
- Gelb, Alan, John Knight, and Richard Sabot. 1991. "Public Sector Employment, Rent Seeking, and Economic Growth." *Economic Journal* 101(408):1186-99.
- Gregory, Peter. 1986. *The Myth of Market Failure: Employment and the Labor Market in Mexico*. Baltimore, Md.: Johns Hopkins University Press.
- Haggard, Stephan, and Robert Kaufman. 1990. "The Political Economy of Inflation and Stabilization in Middle-Income Countries." Policy Research Working Paper 444. World Bank, Country Economics Department, Macroeconomic Adjustment and Growth Division, Washington, D.C.
- Hansen, Gary, and Ayse Imrohorglu. 1992. "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard." *Journal of Political Economy* 100(1):118-42.
- Hirschman, Albert. 1973. "The Changing Tolerance for Economic Inequality." *Quarterly Journal of Economics* 87(4):544-66.
- Horton, Susan, Ravi Kanbur, and Dipak Mazumdar. 1991. "Labor Markets in an Era of Adjustment: An Overview." Policy Research Working Paper 694. World Bank, Economic Development Institute, Washington, D.C.
- Houseman, Susan. 1991. *Industrial Restructuring with Job Security*. Cambridge, Mass.: Harvard University Press.
- ILO (International Labour Organisation). 1990. "Wages, Labour Costs and Their Impact on Adjustment, Employment, and Growth." Governing Body Committee on Employment, GB.248/CE/2/1. Geneva.

- . 1991a. "Employment Policies in the Economic Restructuring of Latin America and the Caribbean." Tripartite Symposium on Structural Adjustment and Employment in Latin America and the Caribbean. WEP 1-4-07 (Doc. 2). August. Geneva.
- . 1991b. "Social Protection, Safety Nets and Structural Adjustment." Governing Body Committee on Employment. GB.2511/CE/4/5. November. Geneva.
- . 1991c. *Yearbook of Labour Statistics*. Geneva.
- Jamal, Vali, and James Weeks. 1992. "Africa Misunderstood, or Whatever Happened to the Rural-Urban Gap?" International Labour Office, Geneva.
- Kendix, Michael, and Mancur Olson. 1990. "Changing Unemployment Rates in Europe and the USA: Institutional Structure and Regional Variation." In Renato Brunetta and Carlo Dell'Aringa, eds., *Labour Relations and Economic Performance*. Basingstoke, U.K.: Macmillan.
- Kormendi, Roger, and Philip Meguire. 1985. "Macroeconomic Determinants of Growth: Cross-Country Evidence." *Journal of Monetary Economics* 16(2):141-63.
- Krueger, Anne. 1974. "The Political Economy of the Rent-Seeking Society." *American Economic Review* 64(3):291-303.
- Lazear, Edward. 1990. "Job Security Provisions and Employment." *Quarterly Journal of Economics* 105(3):699-726.
- Levy, Victor, and John L. Newman. 1989. "Wage Rigidity: Micro and Macro Evidence on Labor Market Adjustment in the Modern Sector." *World Bank Economic Review* 3(1):97-117.
- Lindauer, David L., Oey Astra Meesook, and Parita Suebsaeng. 1988. "Government Wage Policy in Africa: Some Findings and Policy Issues." *World Bank Research Observer* 3(1):1-25.
- Lindauer, David L., Jong-Gie Kim, Joung-Woo Lee, Hy-Sop Lim, Jae-Young Son, and Ezra K. Vogel. 1991. "Korea: The Strains of Economic Growth." Harvard Institute for International Development and Korea Development Institute, Cambridge, Mass.
- López, Ramón. 1991a. "Microeconomic Distortions: Static Losses and Their Effect on the Efficiency of Investment." Policy Research Working Paper 665. World Bank, Country Economics Department, Trade Policy Division, Washington D.C.
- . 1991b. "How Trade and Macroeconomic Policies Affect Economic Growth and Capital Accumulation in Developing Countries." Policy Research Working Paper 625. World Bank, Country Economics Department, Trade Policy Division, Washington, D.C.
- López, Ramón, and Luis A. Riveros. 1989. "Macroeconomic Adjustment and the Labor Market in Four Latin American Countries." Policy Research Working Paper 335. World Bank, Country Economics Department, Trade Policy Division and Macroeconomic Adjustment and Growth Division, Washington, D.C.
- . 1990. "Do Labor Market Distortions Cause Overvaluation and Rigidity of the Real Exchange Rate?" Policy Research Working Paper 485. World Bank, Country Economics Department, Trade Policy Division and Macroeconomic Adjustment and Growth Division, Washington, D.C.
- Marshall, Adriana. 1991. "The Impact of Labour Law on Employment Practices: Temporary and Part-Time Employment in Argentina and Peru." Labour Market Programme DP/38. International Labour Office, Geneva.
- Mazumdar, Dipak. 1989a. "Government Intervention and Urban Labor Markets in Developing Countries." EDI Working Paper. World Bank, Economic Development Institute, Washington, D.C.
- . 1989b. *Microeconomic Issues of Labor Markets in Developing Countries: Analysis and Policy Implications*. EDI Seminar Paper 40. Washington, D.C.: World Bank.

- Olson, Mancur. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation and Social Rigidities*. New Haven, Conn.: Yale University Press.
- Paldam, Martin, and Luis A. Riveros. 1987. "The Causal Role of Minimum Wages in Six Latin American Labor Markets." Discussion Paper 270. World Bank, Development Research Department, Washington, D.C.
- Remmer, Karen. 1986. "The Politics of Economic Stabilization: IMF Standby Programs in Latin America, 1954-1984." *Comparative Politics* 19(1):1-24.
- Riveros, Luis A. 1989. "International Differences in Wage and Nonwage Labor Costs." Policy Research Working Paper 188. World Bank, Country Economics Department, Macroeconomic Adjustment and Growth Division, Washington, D.C.
- Sachs, Jeffrey. 1990. "Social Conflict and Populist Policies in Latin America." In Renato Brunetta and Carlo Dell'Aringa, eds., *Labour Relations and Economic Performance*. Basingstoke, U.K.: Macmillan.
- Scully, Gerald. 1988. "The Institutional Framework and Economic Development." *Journal of Political Economy* 96(3):652-62.
- Sengenberger, Werner. 1991. "The Role of Labour Market Regulation in Industrial Restructuring." In Guy Standing and Victor Tokman, eds., *Towards Social Adjustment*. Geneva: International Labour Office.
- Standing, Guy. 1989. "The Growth of External Labour Flexibility in a Nascent NIC: A Malaysian Labour Flexibility Survey." World Employment Program Working Paper 35. International Labour Office, Geneva.
- . 1991a. "Do Unions Impede or Accelerate Structural Adjustment? Industrial versus Company Unions in an Industrialising Labour Market." World Employment Program Working Paper 47. International Labour Office, Geneva.
- . 1991b. "Structural Adjustment and Labour Market Policies: Toward Social Adjustment?" In Guy Standing and Victor Tokman, eds., *Towards Social Adjustment*. Geneva: International Labour Office.
- Standing, Guy, and Victor Tokman, eds., 1991. *Towards Social Adjustment*. Geneva: International Labour Office.
- Starr, Gerald. 1981. *Minimum Wage Fixing: An International Review of Practices of Problems*. Geneva: International Labour Office.
- Summers, Lawrence H. 1988. "Some Simple Economics of Mandated Benefits." *American Economic Review* (May):177-83.
- Vandemoortele, Jan. 1991. "Labour Market Informalisation in Sub-Saharan Africa." In Guy Standing and Victor Tokman, eds., *Towards Social Adjustment*. Geneva: International Labour Office.
- Weede, Erich. 1983. "The Impact of Democracy on Economic Growth: Some Evidence from Cross-National Analysis." *Kyklos* 36(1):21-39.
- World Bank. 1990a. *Adjustment Lending Policies for Sustainable Growth*. Country Economics Department. Washington, D.C.
- . 1990b. *World Development Report 1990*. New York: Oxford University Press.
- . 1991. *Czechoslovakia: Transition to a Market Economy*. Washington, D.C.

COMMENT ON "LABOR MARKET INSTITUTIONS AND POLICIES:
HELP OR HINDRANCE TO ECONOMIC DEVELOPMENT?" BY FREEMAN

Victor E. Tokman

I would like to comment on three aspects of Freeman's paper: the effects of interventions on wage determination, the labor market consequences of labor regulations, and government versus private employment. In addition, I would like to suggest that future research efforts explore the linkages between labor market interventions and policy sustainability, incorporating collective—not only individual—reactions.

I. WAGE POLICIES

Three main premises related to wage policies justify action through a reform package: the potentially negative effect of minimum wage policies on employment; the downward rigidity of real wages, which slows the process of adjustment (and in turn is linked to collective bargaining); and wage differentials, which should be wide enough to encourage labor mobility.

Minimum Wage Policies

According to Freeman, legal minimum wages are too low (or the standard is too weakly enforced) to constitute a distortion, and endogenous economic factors work to prevent the introduction of distortions that could come about if minimum wages are set too high. The author is right in arguing that real minimum wages deteriorated in the 1980s and that they are presently too low in the developing countries. In Latin America, for example, the average real minimum wage in 1991 was \$88 a month—35 percent below the 1980 level. Minimum wages during the period 1980–91 fell 40 percent in Venezuela, 60 percent in Mexico, and 85 percent in Peru. It cannot be argued that these wages are high in terms of international competitiveness, since the hourly labor cost in manufacturing in Latin America is ten times less than in the United States and is below the present level in Hong Kong, the Republic of Korea, and Singapore. The

Victor E. Tokman is director of the International Labour Organisation's Regional Employment Program for Latin America and the Caribbean (ILO/PREALC), Santiago, Chile. The views expressed in this comment do not necessarily represent those of the ILO.

reasons for the deterioration are threefold: accelerated inflation, weaker union bargaining power, and deliberate policies to abandon the minimum wage for the sake of stabilization and adjustment prescriptions.

I would argue that in spite of earlier experience, minimum wage policies constitute an important policy instrument and that the risk of their introducing distortions is, as Freeman argues, not high. They are effective in setting a floor for the wage structure below which it is socially unacceptable to hire labor. They have also been effective in raising the incomes of the poor (provided that the macroeconomic situation was favorable and could absorb the potential demand and cost effects). This was clearly the case in Colombia in 1983–84 and in Chile after 1989.

The potential distortions in the labor market are not significant. The risk of increasing already high wages and hence of reducing employment in modern sectors is not great; the evidence consistently shows that the entry level in those sectors is usually more than twice the minimum wage and is determined through collective bargaining. The endogenous economic conditions that keep this instrument under control are also important and have been reinforced as a result of economic policy reform. The traditional regulatory mechanism of macroeconomic imbalances that doomed populist governments is today reinforced by the requirements of more open economies. Because there is no possibility of passing on wage readjustments by raising prices without affecting international competitiveness, real wage increases tend to follow productivity changes more closely.

Wage Adjustment and Collective Bargaining

Freeman argues that real wages, on average, have shown downward flexibility, particularly during the 1980s. But there might be tradeoffs between wages and employment, depending on prevailing macroeconomic conditions. Recent Latin American experience suggests that if output decreases, as in 1981–83, both wages and employment fall but that the distributional outcomes are influenced by the flexibility shown by each variable. The same pattern occurs during economic recovery, when employment and wages usually expand but at variable rates (depending on their respective elasticities). Elasticities are in the short run not only a technical coefficient but also the result of economic policies. In Mexico, for instance, unemployment did not increase during adjustment, but real wages contracted significantly. Uruguay's case illustrates the opposite behavior; real wages were not significantly affected, but unemployment expanded. In Chile after 1990, employment and real wages increased in tandem with a sustained economic recovery.

Freeman refers to the need for more analysis to determine whether tripartite agreements are important for efficient wage-setting. This is a welcome suggestion, but I would add that the analysis should go beyond wage issues and examine other effects. In particular, tripartite agreements have been demonstrated to be unique instruments in times of transition from authoritarian to democratic governments. This has been the case in Chile after Pinochet, in

Uruguay after 1984, and in Spain after Franco. Their importance lies in ensuring that the rules of the game are legitimized by the collective actors. This reduces uncertainty, thus encouraging investment and, hence, employment. It also indicates social acceptance. In some cases (Chile is an example) labor legislation may be required to restore a degree of labor protection that had been eliminated. In general, tripartite agreements require action at the national level because signals must be widely transmitted to be effective.

The paper also notes the influence of trade unions and levels of collective bargaining on adjustment outcomes. Nelson (1991) concludes that the strength of the labor movement is an important element to be considered. An aggressive stance on wages tends to be associated with moderately strong unions, generally at the sector or branch levels. Enterprise-based unions tend to be more moderate because they are in a better position to perceive the effects of their claims on the firm, but surprisingly, the largest and most powerful unions also tend to be moderate because they are close to macroeconomic decisions and results. In developing countries the situation can be more complicated because of the militancy of the few large unions and the importance of public sector unions. The economic cycle is another important factor. Depressions reduce militancy—although whether the political regime is authoritarian or democratic does not seem to be crucial; the nature of the political parties and the ways unions are connected to them are more important.

Wage Differentials

Freeman examines the alleged effect of labor market interventions on wage differentials and finds no evidence that differentials increase as a result of interventions. The prevailing argument, however, holds that wage differentials are not large enough to provide an incentive for labor to move into more productive sectors. That was clearly the case in Eastern Europe, where wage differentials were very small, but the argument is not valid everywhere. Wage differentials tend to be associated with the degree of homogeneity of the economic structure, independent of the intensity of institutional intervention. It is, for instance, difficult to allege that wage differentials in Latin America need to be widened, since they are already the largest in the world. In addition, higher wages are associated with higher productivity (or with protected sectors). Under these circumstances, allocative distortions are possible, but they are becoming less important because economies are more open. Such distortions thus are more a matter of trade policy than of institutional labor measures.

I share the author's view that the flexibility of labor markets in the 1980s has been underestimated. In Latin America during the past decade a *de facto* flexibility occurred without institutional changes or government guidance. Only toward the end of the decade were labor reforms enacted in a number of countries.

This *de facto* flexibility took the form of increased unemployment (on average, up to 10 percent) during the depression; an increased share of informal

employment in urban centers (from 33 million to 62 million people between 1980 and 1990); and an increased share of part-time and subcontracted jobs, which conforms to Freeman's statement that he finds no evidence of distortions being introduced as a result of labor regulations. If anything, judging from the effect of labor reform in Spain, the increased flexibility in labor hiring and firing regulations produces a substitution effect as permanent workers are replaced with temporaries. At the same time, however, some overprotected groups—both in labor and in business—were able to slow adjustment in key sectors, particularly ports, railways, mining, and other basic industries. The result of unfettered market operation has been a double distortion: a new class of unprotected labor has emerged that needs attention, while the necessary changes in some of the overprotected sectors are still pending.

The paper argues that public employment performance during the 1980s does not support the interpretation that government intervention in the labor market choked off productive employment. The data for Latin America confirm this finding: real wages in the public sector fell even faster than average wages, and the share of public sector employment in nonagricultural employment remained constant from 1980 to 1990. By the second half of the decade an increasing number of Latin American countries even registered a declining share of public employment, a development that was driven by the need to reduce fiscal deficits and privatize government enterprises under adjustment policy packages. In addition, this decline took place during an inflationary period in which purchasing power was eroded.

An important question for the future is who will be responsible for creating jobs. Clearly, the government will not be as actively involved as in the past. That places the main responsibility on the private sector. In a move that has policy implications, many governments have, in practice, abandoned full employment as a priority objective. Because a favorable economic environment is necessary to motivate entrepreneurs to create employment, this requirement calls for well-designed macroeconomic policies and a regulatory framework that encourages domestic and foreign investment. To promote policy stability, moreover, the prevailing rules must have social legitimacy; that is, they must ensure a supply of jobs that pay more than the minimum wage and that provide some degree of protection. As Solow (1990) notes, the labor market requires a perception of fairness by both main actors. No efficiency wage or insider-outsider theories work when the labor outcome is perceived as unfair or when its determination lies with only one of the parties.

II. THE POLITICAL ECONOMY OF LABOR POLICIES

I would like to suggest an additional topic for discussion, concerning the costs and benefits of interventions to soften the costs of reforms. As Freeman points out, there is a consensus that state interventions during transition are functional to the reform. The difficult question is how much can be allocated to this objective without introducing serious distortions. Is a social fund amounting to

2-3 percent of GDP (like Ghana's or Bolivia's) acceptable? What about a food subsidy that costs 8 percent of GDP, like Egypt's. Would such a subsidy affect key prices during the adjustment process (UNDP 1990)? How should these programs be financed? So far, the only rules of thumb we have are that they should not increase fiscal deficits and that small programs are acceptable, while large ones (an undefined category) are not.

But if only small programs are feasible, they cannot be more than marginal. The equity result, and thus the social acceptance of the reform, will depend on how the macroeconomic package handles sequencing, timing, and instruments. *World Development Report 1990* (World Bank 1990) suggests protecting consumption levels during the early phases of the transition while sacrificing investment. Fiscal and price policies can also be reviewed from this perspective.

Finally, economists usually build a theory casting individuals as consumers or producers. Another perspective would incorporate society, which is not the sum of individuals. To be a citizen requires sharing some collective goals and adopting collective action to achieve them—action that is particularly important in relation to labor market institutions because they are the best way to ensure an equitable outcome in the real world. The building of civil society or the enhancement of its role is not an automatic result of the market; it requires investing in more than individuals. Human development alone is not sufficient to ensure equity.

There is a need for new programs. It was fashionable at one time to assume that government could do and guarantee almost everything. More recently, the assumption was that the market would handle it. The world has changed, and the changes affect individuals as well as society. Take, for example, collective bargaining and union strategy. There is a clear trend in collective bargaining toward decentralization at the enterprise level. At the same time, a new enterprise requires flexible regulation with elements of self-control and disciplined workers who consult with managers. Several changes apply in this new situation: enterprise decisions will increasingly be applied in a bipartite rather than tripartite manner; the regulatory system will move from protection of workers on the basis of legal norms to agreements reached in the context of collective bargaining; and the institutional framework will have to adapt to the new situation by reinforcing the government's capacity to guarantee private agreements.

There is thus a need for representative social actors, particularly trade unions, to ensure equity, good governance, and the permanence of the reforms. Unions have been weakened. Full employment is no longer a state responsibility. Creation of employment increasingly depends on private investment, and wages will have to conform to productivity levels. Union claims for wage increases that ignore these changes are doomed to fail. To be effective, unions should adapt their strategies to build up their presence at the enterprise and national levels, recognizing that society has social objectives that go beyond the particular interests of organized labor. The problems of the poor, of informal workers, and of

pensioners should become part of the unions' agenda. Workers should embrace a broader balance of interests that recognize the new economic realities but at the same time promote changes to achieve greater equity.

REFERENCES

- ILO (International Labour Organisation). 1991. *Employment Policies in the Economic Restructuring of Latin America and the Caribbean*. Geneva.
- Nelson, Joan M. 1991. "Organized Labor, Politics, and Labor Market Flexibility in Developing Countries." *World Bank Research Observer* 6 (1):37-56.
- Solow, R. M. 1990. *The Labor Market as a Social Institution*. London: Blackwell.
- UNDP (United Nations Development Programme). 1990. *Human Development Report*. New York.
- World Bank. 1990. *World Development Report 1990*. New York: Oxford University Press.

COMMENT ON "LABOR MARKET INSTITUTIONS AND POLICIES:
HELP OR HINDRANCE TO ECONOMIC DEVELOPMENT?" BY FREEMAN

Kwadwo A. Tutu

Professor Freeman's excellent paper presents the issues in a concise manner and identifies the mode of analysis and the yardstick for judgment. Because it takes a serious look at stereotyping and examines both positions from a theoretical and empirical point of view, it is easier to confront attackers and defenders and see the merits in each case.

The findings of the paper are instructive in that the empirical results do not support many of the neoclassicists' allegations of distortions caused by labor market institutions. The author's conclusion that we need to do more empirical work in individual countries is as important as the theoretical debate.

One theme that the author has neglected is the potential for labor to play a cooperative role in the production process (in addition to its traditional role as a factor of production). Unless we understand that this is appropriate behavior, the productivity effects of labor-enhancing policies arising from institutionalism will be largely ignored. It is also a mistake not to recognize that the notions of replacement cost and capital depreciation—which are standard investment writeoffs—also apply to the wear and tear incurred by the labor force. These and other characteristics impinge on labor's productivity. For example, there is a fundamental link between work effort and the employer-employee relationship. The three broad categories of work effort listed here provide a framework for understanding the behavior of labor.

- *Slave labor.* In this situation labor suffers under a tyrannical environment for little pay. In some countries payments to selected categories of labor are so low that they could be equated with slave labor. In this case strict supervision could lead to increased output, but no creativity can result from such labor.
- *Productivity-augmented labor.* In this case labor is given incentives that produce a high earning profile for the worker. Increased productivity could result from improved efficiency but is unlikely to lead to creativity or innovation.
- *Cooperative labor.* This is the optimal situation: there is increased productivity, and labor is not antagonistic. Respect develops when labor is consid-

Kwadwo A. Tutu is lecturer in economics at the University of Ghana.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

ered part of the management and production team and not just a factor of production. It is only in such circumstances that both productivity and creativity can occur. It is this form of labor that provides the rationale for the existence of institutions that are often considered distortionary.

In many less-industrialized countries formal earnings are below a basic living wage, resulting in low productivity and the migration of skilled workers. The establishment of minimum wages, nonwage labor benefits, and job security could increase profitability. This is the example displayed to some degree by the codetermination of Germany, the paternalistic labor practices of Japan, and the workers' industrial cooperatives in Spain. If productivity is a function of the quality of labor and higher minimum wage improves the quality of labor, higher productivity can result from minimum wage legislation.

Although the paper has made significant contributions to the debate, two other issues need to be addressed. The first concerns public sector employment. It has been argued that taxes and fiscal deficits that fund public sector employment crowd out investment in more productive private sector enterprises. This is not necessarily true. In Africa, where there has been significant deterioration of the economy, public investment in rehabilitation of infrastructure can be very productive. Moreover, the notion that all public sector enterprises are inefficient should be approached with caution. There are efficient—and inefficient—public enterprises. It is important to separate investment in the public sector from the management of that sector. If management is left in the hands of competent personnel without undue interference, it is possible to conduct business efficiently. Ethiopian Airlines is a good example; this parastatal company is one of the best airlines in Africa.

The second issue concerns labor unions. It is not useful to concentrate on the traditional role of unions by looking at ways to curtail their power or eliminate them. A more constructive approach would be to help transform them from antagonists to cooperative partners, not only in profit-sharing but also in the management of enterprises. In several countries, including Ghana, unions cooperated with government efforts to lay off workers in parastatals once they were convinced of the need. Research to address a more productive role for unions is essential.

While we argue about whether labor interventions are distortionary, it has to be pointed out that these laws are irrelevant. The earnings of most workers in less-industrialized countries are so low that there are serious costs to society resulting from shirking, stealing, lethargy, and rent-seeking. Obviously, these are influential factors in the efficient management of the economy. The costs of these activities to society are immense and would far outweigh the direct costs of paying labor decent wages. There is an urgent need for research in this area.

FLOOR DISCUSSION OF THE FREEMAN PAPER

Starting the floor discussion, a Nigerian economist said that he found Freeman's arguments full of flaws because of the selective way he used economic theory. He felt that the greater the prisoners' dilemma and similar problems are—as in many developing countries—the stronger the case for the institutionalist position.

Freeman was pleased that the speaker thought prisoners' dilemma, moral hazard, and agency problems were severe, but Freeman himself had not seen serious evidence of this. The speaker agreed with Freeman that views of liberal market interventions should be grounded less on models and more on case studies; in the cases of Taiwan (China) and Mauritius the government had intervened, yet there had still been economic growth. Freeman's observations about minimum wages, wage adjustment, and collective bargaining also held true in Nigeria, he added.

A participant made two comments. First, he stated that Freeman's paper focused too narrowly on trends in wage labor, which accounts for less than half of employment in many developing countries. Second, there could be two views on why the distortions made no difference: interventions such as the minimum wage might not be enforced or might be evaded, in which case one would not expect them to affect the economy. And if they were enforced, one would have, first, an enforcement cost, and second, the rent-seeking Freeman had talked about. The speaker said that he didn't know how to interpret this empirical evidence.

Freeman admitted ignoring nonwage labor. As for interpreting distortions, he said that in countries where interventions were determined domestically and where it was evident that maintaining the minimum wage during a massive decline would be disastrous, political economy would compensate. People would say, "Wait a minute, this is crazy." Freeman said he had given the example of Puerto Rico to show that he fully understood that one could have a distortionist nightmare in a particular economy. Puerto Rico was not a good example, however, because interventions and incentives introduced there were imposed by an outside authority, the United States. If Puerto Rico had not had U.S. subsidies through such vehicles as the U.S. welfare system and tax breaks—and if the one-third of Puerto Ricans who lived in the United States (who tended to be the less-educated Puerto Ricans) had been returned to Puerto Rico—we would not have seen the minimum wage in Puerto Rico, enforced or not, because there would have been a total disaster.

The session was chaired by Lyn Squire, chief economist, Middle East and North Africa Region, the World Bank.

Another participant insisted on the importance of timing in discussing wage flexibility: it made a difference whether flexibility came before or after a crisis and adjustment. In Chile the unemployment rate had gone from 8 to 25 percent in five quarters because a dictatorship mandated that real wages could not fall. Flexibility often comes after a crisis, she argued. To restructure in Eastern Europe and the former U.S.S.R., flexibility is needed.

Freeman agreed that the question of timing was important but said that he had not focused on particular countries. Instead, he had been struck by how much real wages had declined over time in many countries. And there did not need to be a dictatorship for that to happen. There had been no dictatorship in Costa Rica when, as one study claims, interventionist wage policies had actually helped adjustment in that economy. It was because trade unions in Eastern Europe and the former U.S.S.R. favored reform and the market economy that the old guard has not been able to cause trouble there, said Freeman, and the old guard was not usually found in the old Communist trade unions. The story might differ from country to country, he said, but one should not assume that the only way to get wage flexibility was through force.

A participant did not think it was useful at this point to go back to setting up straw men. He thought it was about time for labor economists at the World Bank and the International Labour Organisation (ILO) to collaborate on ways of finding a consensus on labor market policies. One fruitful area of research for the two groups would be how to increase voice options in the labor market without sacrificing the efficiency of market mechanisms. He stated that it was an oversimplification to think of issues as regulationist or deregulationist. Labor market regulations take many forms, and they should not be viewed as all of one type; to think of them as different would open up a tremendous research agenda. The same was true of institutions. He agreed with Freeman that there was a vast literature on collaborative, cooperative forms of labor institutions and thought it would be useful if both camps took stock of the costs and benefits of different kinds of institutions. Indeed, he felt that more such research was available than Freeman had indicated.

Larry Summers argued that although public sector wages were lower in some countries than they used to be compared with private sector wages, in dozens of countries nurses' wages were triple the average wage, yet the countries couldn't afford stethoscopes; or the student-teacher ratio was twelve-to-one, but students didn't have books—all because of the pressure to keep wages high. Governments were eager to privatize, but labor would not tolerate it, and labor's strength greatly compromised the public sector's effectiveness.

Second, said Summers, arguing that everything was okay because real wages had fallen spectacularly reminded him of comments on British public finance in the 1960s, when people contended that 93 percent tax rates hadn't produced much revenue, so they couldn't be doing much damage. If by restricting property rights labor could discourage investment, he argued, one could expect a reaction in the form of low real wages. In many countries, substantial real overvaluation

had grievously harmed the agricultural sector, where two-thirds of the labor force (the disproportionately poorer two-thirds) worked. Devaluing exchange rates under adjustment had increased equity and reduced urban workers' measured wages; labor had resisted devaluation, but if labor were less powerful and cohesive, there would be less resistance to these egalitarian adjustments. Summers asked Freeman if the evidence in his paper refuted the two propositions he had just made.

Freeman said he had known someone would come up with a strange country that provided evidence to counter his, but *dozens* of countries? He would like to know where nurses were so highly paid, especially if they were women—certainly not in the United States. He had not found those dozens of countries in the studies produced by the World Bank and the ILO. Freeman tended to agree with Summers that the fall in real wages was greatly exaggerated because people had large quantities of consumer durable goods, but he felt that nobody should take joy in falling real wages in any case because some people were getting a lot poorer. There was evidence, however—certainly for Hungary—that a third of labor earnings was coming from gray market activities, so it was missing the boat to look only at official real wage figures.

The World Bank's view, as Freeman read it, was that countries that imposed minimum wages could never devalue their currency: the labor market might (and he emphasized "might") mishandle an economically essential devaluation. But the evidence, he said—even in studies by people with a strongly distortionist perspective—was that these institutions had not stopped a devaluation that was necessary. A decline in real wages was an unfortunate consequence of devaluation that he was neither for nor against. He was willing to be convinced that real wages in agriculture had increased greatly because of devaluation, but in his reading, the day wages of agricultural laborers had not increased sharply, although possibly the incomes of farmers and those selling food (whom he had neglected in this paper) had improved.

A participant, prefacing her remarks by saying that it was not fashionable to argue the interventionist case, asked Freeman if he did not think there was a case for regulating at least workers' conditions—for example, forbidding child labor, guaranteeing women equal pay for equal work, or insisting that workers be paid the money they earned instead of having it doled out at the end of the month, after employers had subtracted an unfair amount in deductions. Freeman admitted that he had experienced some ambivalence about child labor laws in developing countries since his visit to a brick factory in Sri Lanka where he had seen a child of six or seven working. He had asked the employer why he was allowing such a child to work and the employer had said that otherwise the child would have starved to death. One might prefer that homeless children get money without having to work for it, but before one prohibited child labor one should be sure a system was in place to provide for such a child's welfare. In setting up stabilization and adjustment programs, it was also important, he said, to be sure that people could not only survive but possibly better themselves.

A participant said he wished Freeman had talked less about static efficiency and growth and more about dynamic efficiency and growth. If one shifted focus, the institutional issues were different. Investment, innovation, and incremental productivity were associated not just with formal rules and obligations but also with values and norms that shaped the patterns of discrimination, motivation, and commitment which might be important for productivity growth. He thought important research might be done to identify stable clusters of institutions (there might be alternatives), to see which were associated with egalitarian and which with inegalitarian growth patterns, and to tackle the issue of how one got from one to the other.

Freeman said he thought a logical case could be made for the issue of dynamic efficiency: that raising wages a bit through a minimum wage reduced incentives to invest, which had a negative effect on society. The evidence did not seem to offer a great deal of support for this argument, he said, but maybe that was because people hadn't worked as hard as they should have to make the case. Freeman agreed, however, about the importance of the rules of the game. The Aoki model of a Japanese firm, he said, was very different from the model of the firm economists usually deal with, in which the boss tells you what to do and you do it. Bringing labor creatively into the production process clearly affected production, said Freeman, and China seemed to be doing reasonably well in growth despite bizarre property rules about who owns what and how. Freeman added that productivity seemed to be pretty high in cooperative settings, judging from studies he had seen on profit-sharing in the United States and trade unions in Malaysia, among others. He had found no serious discussion of the tripartite national groups Victor Tokman (discussant) had mentioned. He wished rules had been established that everyone could have followed so that in Eastern Europe and Venezuela more players had been made to feel part of the reform process. If they had, perhaps there would not have been an aborted military coup (which did not upset the country because the people felt they had been left out of the process anyway).

In bringing the debate to a close, Lyn Squire (chair) said he had drawn from the discussion two conclusions about future research. First, the evidence—such as it was—suggested that labor market interventions might cost less if they were better implemented. If so, research should be focused on countries in which distortions are especially severe, and it should be country-specific research rather than cross-country regression analyses.

Expressing sympathy for Freeman's suggestion that it might be worthwhile to intervene in the labor market during reform if that would help reform succeed, Squire suggested an additional need for more empirical research and analysis of such issues as severance pay, unemployment benefits, and tripartite agreements.

Labor Market Adjustment in Transitional Economies

Jan Svejnar

This paper outlines the principal labor market developments in the transitional (formerly centrally planned) economies, presents a framework for conceptualizing the functioning of these markets, and draws conclusions with respect to public policy in the labor market. It assesses the effect of transition on wages, employment, and trade unions. The paper observes that although wage controls played an important role in stabilizing these economies, they also maintained and aggravated major distortions. The paper recommends the imposition of hard budget constraints and proposes that profitable enterprises in the business of producing tradables be permitted to set their own wages. With rapidly rising unemployment, the reallocation of redundant labor requires immediate attention. The paper argues that social—rather than private—welfare criteria should guide the behavior of state enterprises. Finally, trade unions and other employee institutions have not been involved in the formulation and implementation of market-oriented reforms. Although they have been amenable so far, the danger is that mounting job losses and economic hardship could turn these organizations into formidable adversaries of the transition process.

As the economic transformation of the formerly Communist countries unfolds, it is becoming increasingly evident that managing macroeconomic stabilization is far simpler than achieving an adequate microeconomic transformation. Poland, Hungary, and Czechoslovakia, for example, managed to bring inflation down to less than 2 percent a month within a few months of the start of the transformation program (Bruno 1992), but they have so far failed to commercialize state-owned enterprises, and only Czechoslovakia is carrying out large-scale privatization. Since evidence from the developing countries suggests that macroeconomic stabilization tends to unravel in the presence of inadequate microeconomic restructuring (Kiguel and Liviatan 1991), the early successes of the transitional economies may be misleading.

Jan Svejnar is professor of economics at the University of Pittsburgh and chairman of the faculty of the Center for Economic Research and Graduate Education, Charles University, Prague. The author would like to thank Marek Góra for useful comments on an earlier version and Julio Kipnis for his help in assembling the data for table 1. He is indebted to the National Council for Soviet and East European Research for its assistance.

An understanding of labor markets provides important insights into how economies operate, how their enterprises function, and how governments formulate microeconomic policies. This paper assesses the developments in the labor markets of the transitional economies, uses stylized facts to conceptualize the behavior of the main kinds of firms, and draws conclusions about public policies vis-à-vis the labor markets.

I. FEATURES OF THE LABOR MARKET

The formerly centrally planned economies have all embarked on the transition to free markets with a common set of labor market characteristics: zero open unemployment and visible excess demand for labor; a relatively low level and rate of growth of absolute wages; small wage differentials across most skills and industries; significant quit rates but low geographic mobility; virtually 100 percent membership in the official Communist trade unions; and the introduction of worker participation in enterprise management.

The system of soft budget constraints for state-owned enterprises, the shortages and unpredictable deliveries of material inputs, and the penalties incurred by managers and workers for underfulfillment of plan targets made it rational for firms to hoard labor. This system led to full employment with parallel vacancies and redundant labor. Marginal productivity of labor was low or zero (see Terrell 1992), work effort was generally low, and yet firms continuously advertised vacancies.

The system functioned because the authorities maintained absolute wages at a low level. Workers tolerated the system because it seemed egalitarian (wage differentials were small) and because it was promulgated and supervised by trade unions in a totalitarian system. The strength of the labor unions can be attributed to their affiliation with the Communist party apparatus (Windmuller 1971). Members enjoyed special bonuses in the form of vouchers for vacations at recreational facilities and spas, subsidized purchases of consumer durables, and so on (Freeman 1992).

The full employment-cum-vacancies system resulted in significant labor turnover because workers could advance somewhat by changing employers (Riveros 1990). Regional labor mobility was minimal, however, in large part because housing was under strict rent control and hence in short supply.

In the 1980s Poland, Hungary, and Czechoslovakia gave employees participatory rights in the selection of management. These rights were not as extensive as those enjoyed by workers in Yugoslavia, and under the Communist regime participation was more *de jure* than *de facto*. As a result, attempts to spur productivity by involving workers in the production process were by and large ineffectual. Enterprise (workers') councils were staffed mostly by party members, and influence over the nomination and recall of enterprise directors was limited.

A natural outcome of the system was a low-wage, low-effort equilibrium in the public sector and a marketlike behavior in the informal (private) sector. The

private sector gradually became more important as the Communist party began to lose control, and workers in state-owned enterprises increasingly engaged in informal—and sometimes illegal—activities.

The quality of the labor force in Central and Eastern Europe (especially Czechoslovakia, Hungary, and Poland) is relatively high: the proportion of the labor force with more than a basic education is comparable with that in the average member country of the Organization for Economic Cooperation and Development (OECD) (Boeri and Keese 1992). In higher education, the transitional economies do not rank as well, but they usually dominate some of the OECD countries. The type of education also differs; a higher proportion of secondary school students attend vocational schools, and the proportion of higher-level students in the field of engineering tends to be greater. Data on enrollment rates in secondary and postsecondary schools indicate that Central and Eastern Europe lags the West in the educational level of recent labor force entrants. Finally, perhaps the most important question is, to what extent are the skills acquired by Central and East Europeans obsolete, and how much retraining is necessary to renew their human capital?

II. TRANSITIONAL LABOR MARKET POLICIES AND SHOCKS

In all the transitional economies, independent trade unions, collective bargaining, and the right to strike are now legal. Pattern bargaining takes place at the level of a tripartite decisionmaking apparatus made up of trade union (con)federations, government representatives, and employer federations, but additional earnings adjustments take place at the company level.

While comprehensive price liberalization has been one of the pillars of the transformation, every country has imposed wage controls to prevent high (hyper-) inflation. These incomes policies have taken various forms. Poland, for instance, established an enterprise wage indexation scheme in 1990 that compensated in part for expected inflation. Any increase in the growth of the wage bill above the benchmark level was penalized by a prohibitive 500 percent tax. In 1991 the scheme was targeted to the average wage to minimize a potentially negative impact on employment. The government of Czechoslovakia established a similar, although graduated, tax scheme targeted to the average wage in 1991, forgot to renew it in the first half of 1992, and reimposed it in July of that year. In recent years Hungary, where wage controls have been in effect since the 1980s, has not taxed—or has taxed only lightly—wage increases if they correspond to increases in value added. Moreover, since 1991 the tax levied on wages has decreased in parallel with a decrease in the firm's employment, thus encouraging firms to lay off redundant workers.

Layoffs are now legal, as are short hours and part-time employment. Unemployment benefit schemes and active labor market policies have been introduced to cope with rapidly rising unemployment. Initially, these compensation schemes were relatively generous and provided workers with more than 60 percent of their previous wage for up to a year. Recently, budgetary problems

and rising unemployment have forced Czechoslovakia and Poland to scale down these generous benefits and reduce the eligibility period (Boeri and Keese 1992); other countries may have to follow.

A number of other policies as well as external shocks have altered labor demand and the operation of labor markets. Subsidies to enterprises have been cut substantially—as have commodity subsidies—making it difficult for firms to maintain employment and wages at previous levels. In some countries retroactive increases in interest rates on existing bank loans had the same effect. The dissolution of the Council for Mutual Economic Assistance (CMEA) and the collapse of the Soviet market led to a decline in regional trade and negative terms of trade shocks as transactions shifted from transferable rubles to convertible currencies. Rodrik (1992), for instance, estimates that the total 1990–91 Soviet trade shock amounted to a loss of 7–8 percent of gross domestic product (GDP) for Czechoslovakia and Hungary and 3.5 percent of GDP for Poland. Because these calculations are conservative (they do not include, for instance, the multiplier effect), they probably understate the external shock.

III. PRINCIPAL LABOR MARKET OUTCOMES

Table 1 contains comparative data on output, employment, and real wages for 1989, 1990, and 1991. In Hungary, where a gradual transition began earlier (Dervis and Condon 1992), the decline in employment and wages that could in part be attributed to external shocks and transition policies has been milder than in other economies. Poland's rather sudden transition in January 1990 resulted in a rapid decline in output, real wages, and (to a lesser extent) employment. Output fell in 1990 in Bulgaria and Romania as well—a reflection of the collapse of the CMEA market—but in both countries, and in Czechoslovakia, the principal shock occurred in 1991, when they had launched their economic strategies and were coping with the terms of trade shock associated with the shift of CMEA trade into convertible currencies.

A closer examination of the data in table 1 and of more disaggregated figures reveals a number of regularities. First, the decline in output at the start of the transition tends to be deep and fairly (though not completely) uniform, signaling a shock rather than intersectoral restructuring. Furthermore, the decline in employment is significant, especially when one considers not only the decrease in the number of employees but also the decline in hours worked. In Czechoslovakia, for instance, the average number of hours worked declined from 1,859 in 1989 to 1,817 in 1990 and 1,724 in 1991. With the possible exception of Hungary, however, cuts in employment have not matched reductions in output, and labor redundancy, which was estimated to range from 15 to 25 percent of the work force before the transition (Jackman, Layard, and Nickell 1992), has increased.

The fall in real wages has been comparable to the steepest declines recorded in the developing economies during stabilization and structural adjustment. In Poland the decline was in part offset by the real wage gains recorded in preced-

Table 1. *Production, Employment, Wages, and Unemployment in Selected Transitional Economies*

Country	Industrial production (percentage change)	Annual change in employment (percentage)	Real earnings (percentage change) ^a	Unemployment rate (year-end)
<i>Bulgaria</i>				
1989	-1.1	0.2	2.3	—
1990	-16.8	-8.2	-2.8	1.5
1991	-27.5	-17.1	-33.8	10.5
<i>Czechoslovakia</i>				
1989	0.7	0.3	1.0	—
1990	-3.7	-2.5	12.5	1.0
1991	-23.0	-12.0	-26.0	6.0
<i>Hungary</i>				
1989	-1.0 ^b	-4.0	0.9	0.5
1990	-9.0 ^b	-11.3	-5.1	2.0
1991	-21.5 ^b	-16.5	-5.8	8.0
<i>Poland</i>				
1989	-0.5	-1.4	—	—
1990	-24.2	-8.9	-37.2	6.3
1991	-14.2	-12.6	-8.4	11.4
<i>Romania</i>				
1989	-6.6	1.9	2.4	—
1990	-19.2	-1.3	4.6	1.0
1991	-13.0	-18.7	-24.9	2.2

— Nonexistent.

a. Real earnings for Bulgaria and Czechoslovakia; real wages for Hungary, Poland, and Romania.

b. Firms with fifty or more employees.

Source: Various government and World Bank data.

ing years, but in Czechoslovakia and Bulgaria real wages fell from a relatively low base. The wage differentials between Eastern and Western Europe have thus grown dramatically. Recent estimates, for instance, indicate that Volkswagen's hourly cost for production workers is more than DM 50 in Germany, DM 25 in Spain, and DM 8 in Czechoslovakia. Yet, the efficiency of the Czech and Spanish plants is reportedly comparable.

Unemployment, which was unknown outside of Hungary before 1989, emerged in 1990 and has risen rapidly. By the end of 1991, unemployment in Bulgaria, Poland, and Slovakia exceeded 10 percent of the labor force. In some countries (for example, Czechoslovakia) the unemployment rate fell in 1992, but much of the change was the result of a decreased eligibility for unemployment benefits.

There is no doubt that in an important way labor has borne the brunt of the early stages of transition. From the policy standpoint the interesting question is why there has been a more moderate decline in employment than in output and wages in most of these economies. There are several possible explanations. One, of course, is the tremendous decline in real wages that is documented in the table. With labor costs shrinking to between 5 and 10 percent of total costs, managers have focused on economizing in other areas (energy, for example) and

have tended to substitute labor for other inputs. Another reason is that managers, workers, and government officials view the maintenance of employment as an important goal during transition. Indeed, such behavior is consistent with the pursuit of rational objectives by state-owned enterprises, taking into account the falling shadow wage in the economy (Svejnar and Terrell 1991), the expected utility maximization by workers, and the cost-minimizing behavior of managers in the presence of costs of labor adjustment.

Private Firms versus State-Owned Enterprises

Unlike state-owned enterprises, domestic and foreign private firms and joint ventures appear to economize on employment. They also pay considerably higher wages and elicit greater effort and quality of work from their employees. The private sector thus behaves consistently with profit maximization and possibly pays efficiency wages to generate higher effort. Stabilization policies severely constrain the ability of state-owned enterprises to pursue an active wage policy and stimulate their employees to exert greater effort. They are increasingly victims of brain drain as the best workers in all skill and occupational categories leave for the private sector. Unfortunately, although the private sector has expanded in all the countries in this group, it is still too small, and the inflow of Western capital has been too limited, to offset the decline of state-owned enterprises.

The evidence emerging from the transitional economies points to a number of common patterns. The enterprise sector includes private firms that behave consistently with profit maximization and state-owned enterprises that emphasize employment and try to evade hard budget constraints. Unless bound by a prior agreement with the government, Western firms that acquire local enterprises tend to slash employment; they usually employ a small labor force, frequently on short-term contracts. State enterprises, in contrast, maintain a large work force even in the face of sharp declines in demand. This situation is only tolerable because of declining labor costs and subsidies from the government (or the ability to increase the enterprise's indebtedness). The state-owned firms are releasing excess labor, albeit at a rate that is much slower than a rapid adjustment to a profit-maximizing level would dictate, and the private firms are absorbing the released workers and the new entrants to the labor force. The main problem is that in the early phases of transition the private sector is too small to absorb all the workers who have been released by state-owned firms. The result is a gradually emerging and growing unemployment.

Unions, Wages, and Employment

The artificial 100 percent unionization rates characteristic of the former Communist economies have vanished, and membership is now estimated at between 30 and 75 percent (Freeman 1992). Surprisingly, however, most workers have opted to remain in the old unions rather than join new independent trade unions, apparently because of the advantages of incumbency. The old unions

responded to the challenge posed by transition reforms by expelling the Communist leadership, taking control of the unions' (considerable) resources, and effectively representing their members in the workplace. The new unions face major disadvantages: limited financial resources; inadequate staff; and, frequently, an intellectual, rather than a blue-collar, orientation (Freeman 1992). The fragmentation of the union movement and the questionable legitimacy of the successor unions has enabled governments in most of the countries in transition to dominate the tripartite deliberations and reduce the unions to consultative—rather than bargaining—partners in wage settlements. The unions are, however, reported to be gaining influence within firms, where decisions about employment, layoffs, and wage increases are made.

The behavior of wages during the Polish (1990) and Czechoslovak (1991) big bang provides an instructive pattern. The nominal wage increases observed in the first six months of 1990 in Poland and the first six months of 1991 in Czechoslovakia fell well below the wage (control) norm, although real earnings declined dramatically in both countries (see table 1). In the second six months of the transition (July to December 1990 in Poland and July to December 1991 in Czechoslovakia) wages started to catch up with the norm in both countries. In the second year average wages in Poland substantially exceeded the norm, and enterprises paid the stiff penalty tax.

These developments are interesting because they reveal a different approach to wages and employment at the national and enterprise levels. At the national level the decision to set the wage well below actual inflation reflected an underestimation of future inflation, an attempt to achieve price stability, and the desire of the social partners (especially government) to limit the negative impact of the transition on employment. Within the firm, the initial (below-norm) wage increase reflected the desire of trade unions, management, and workers to limit layoffs in a period of uncertainty and existential anxiety. Nevertheless, when real wages fell precipitously and it became clear that reductions in employment could in large part be accomplished through attrition, hiring freezes, and the termination of fixed-term contracts, unions and workers pushed for higher wages. The wage norm hence seems to be unsustainable in the medium term in the presence of a sharp decline in real wages.

Worker Participation in Management

Employee participation in management—and profits—is controversial and has been criticized by government officials and external advisers. Indeed, some observers (see Hinds 1991) have argued that the poor economic performance of Poland, Hungary, and Yugoslavia in the late 1980s reflected the strong control exerted by employees in these economies.

The potential abuse of power by workers inside the firms should not be underestimated. But economic theory and empirical evidence do not support such a negative view of participation. Economic theory on participatory and unionized firms indicates that in a competitive environment these firms do not

behave very differently from their profit-maximizing counterparts. Moreover, under conditions of redundant labor, participatory and unionized firms tend to allocate labor by the same criteria as profit-maximizing firms even in the presence of extranormal profits or rents. Econometric evidence also indicates that the economic effects of these schemes are nonnegative (see the surveys in Blinder 1990).

More casually, economic performance worsened not just in the socialist economies that introduced participation (Poland, Hungary, and Yugoslavia), but also in those that did not—Czechoslovakia, the former German Democratic Republic (GDR), and the former U.S.S.R. Moreover, neighboring capitalist economies (Austria and the Federal Republic of Germany) fared well in the postwar period, with significant worker participation in management. The former GDR adopted codetermination after German unification with no obvious ill effects. And a recent study by Prasnikar and Svejnar (1991) found that government intervention, a lack of competition, and the absence of clearly defined property rights—rather than employee participation per se—appeared to be the principal shortcomings of Yugoslavia's economic system.

The Social Safety Net

As noted above, fiscal problems have already forced Czechoslovakia and Poland to reduce the level and duration of unemployment benefits, and other governments are likely to follow. The ability to sustain social safety nets will become increasingly limited. Active labor market policies, including training schemes, public works programs, and subsidies for new employment have been introduced, but so far they account for only a small proportion of total labor market expenditures and do not appear to be very effective. As the financing of unemployment compensation becomes a growing priority, these policies may become marginalized.

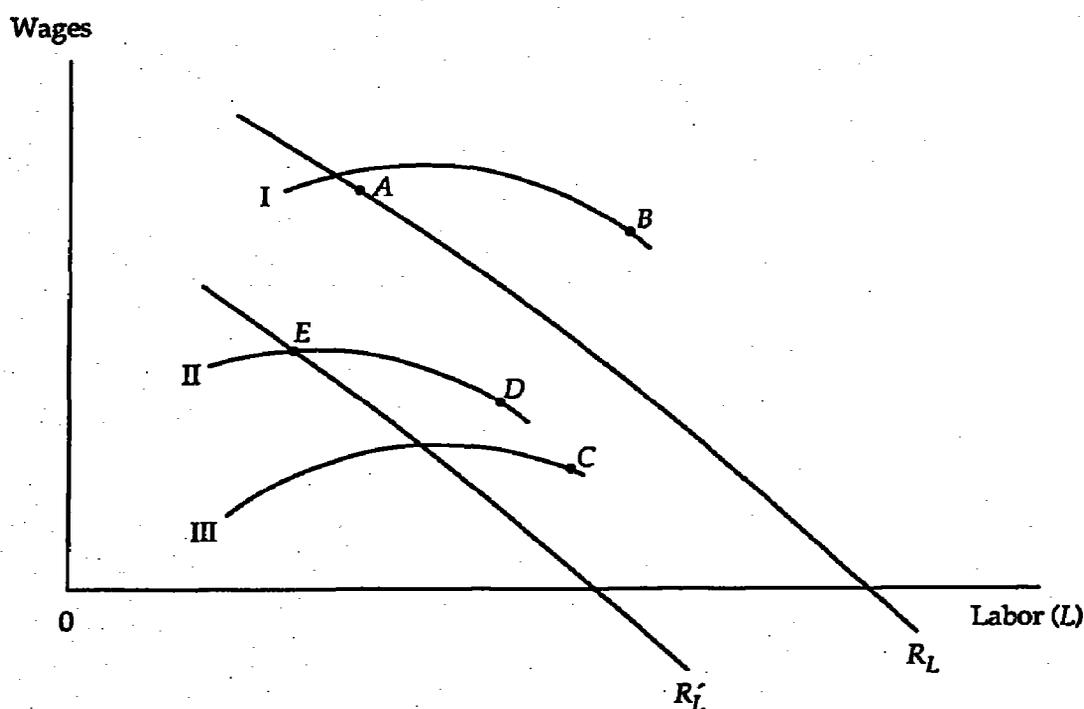
IV. POLICY IMPLICATIONS

A relevant model of the labor market behavior of enterprises in transition can be cast in the form of a tripartite bargaining framework in which the government, managers, and workers jointly maximize an objective function that contains profit, wages, and employment (see Svejnar 1986; Prasnikar and others 1991). As the transition proceeds, the weight that the three parties place on employment gradually falls, while the emphasis placed on profitability and wage growth rises. Figure 1 depicts in a simple way the variety of wage and employment outcomes that can be observed in the transitional economies. Let R_L be the marginal revenue product of labor, W the wage, and curves I, II, and III the isoprofit curves (all wage-employment combinations on a given isoprofit curve yielding the same level of profit; the higher the curve, the lower the profit). A labor market outcome corresponding to the private (profit-maximizing) firm is, for instance, $R_L = W$, depicted by point A in the figure. Now, most enterprises started the transition with redundant labor, thus implicitly placing emphasis on

employment and displaying $R_L < W$. In the figure this means that the outcome is to the right of the marginal revenue product curve of labor at a point such as *B*. Because the pretransition situation was often characterized by subsidies (negative profits), the outcome is depicted in the figure as lying relatively high in the northeast—lying on a low- (possibly negative) profit isoprofit curve and displaying relatively high wages and employment. The transition and the negative external shocks shift the marginal product curve to the left as demand falls, and they also result in a reduction in subsidies to the state-owned enterprises. This effect is captured in the figure as a move to point *C* above the new marginal revenue product curve R'_L . The outcome is characterized by somewhat lower employment, significantly lower wages, and greater labor redundancy (more labor with low or zero marginal product). In the later phase of the transition one then observes a rise in wages and a further decline in employment (that is, a shift from *C* to *D* and later to *E*). The characterization of the transition in terms of enterprise employment and wage behavior hence consists of the accumulation and reduction of redundant labor and the absorption of the released labor by the emerging private firms.

From the standpoint of economic efficiency, a fundamental question that arises with respect to the transition policies is whether the wage (bill) controls (norm) ought to be maintained. The principal justification of the norm is that it induces price stability and prevents workers from according themselves unwarranted wage increases and thus decapitalizing state-owned firms. The disadvan-

Figure 1. *Wages and Employment in Transitional Economies*



tage of the norm is that it maintains much of the previous distortion and prevents the transforming economies from overcoming low effort and poor workmanship by a skilled and relatively well-educated labor force.

Since the transitional economies in Central and Eastern Europe are resource poor, the negative incentives imposed on the human factor are a major hindrance to a successful transition. This latter aspect appears so important in the context of the transforming economies that one ought to think seriously about alternative measures. Since the economies are now relatively open to trade, one possible scheme would be to allow state-owned firms in the tradable goods sector to pay profit-sharing bonuses if they are able to (a) operate without subsidies, (b) reinvest a given proportion of capital, and (c) pay the government a set real rate of return. A scheme of this kind retains the incentive for the firm to increase productivity, improve marketing, maintain prices at world market levels, and reward key employees. Incentive schemes for managers (and possibly workers) should also be based on the company's successful long-term performance (for example, shares of stock that are nonredeemable for a number of years).

The second policy issue concerns the optimal approach to redundant labor. Measures encouraging labor mobility are clearly desirable. One often observes, however, an exclusive emphasis on reallocating resources and closing unprofitable operations. Far less attention is paid to increasing the efficiency of resources, given their allocation. In particular, the question arises as to whether resources ought to be reallocated to a massive extent in the short run. For instance, it is not clear how many workers would be employed in Czechoslovakia's steel industry under perfectly competitive conditions. Their productivity may be low, but as the earlier discussion indicates, it is low elsewhere as well. It may not have made sense to develop the steel industry to begin with, but with a limited inflow of Western capital in the short run, the relevant issue is how to use labor optimally, given the existing capital structure. To the extent that governments maximize a social welfare objective such as gross national product, the state-owned enterprises ought to employ labor to the point where labor's marginal product equals the shadow wage (productivity of the marginal employee elsewhere in the economy) rather than the actual wage. This makes it optimal from society's standpoint to maintain higher employment levels in state-owned firms than would be dictated by the profit-maximizing criterion alone because the total product of these firms cannot be increased by reallocating labor among them. By implication, as long as actual wages exceed shadow wages, too few workers are employed by private firms from society's perspective. Finally, in deciding to close unprofitable enterprises, it is important to take into account the fact that the profitability of a firm may not be indicative of its efficiency unless it is evaluated at shadow prices and wages. (For a detailed discussion of these issues see Svejnar and Terrell 1991.)

My third point concerns the role of trade unions and the industrial relations system. The early successes of transition have come in part from the weakness of

the unions and their tolerance of stabilization measures. Some governments have effectively ignored the unions and have pushed through labor market measures without "codetermination" under the framework of the established social partnership. This approach is effective in the short term and, if unionization does not take off in a major way, it will work in the long term as well. The approach may backfire, however, if unions become stronger and workers become less tolerant of the transition measures. Since transition promises to be a protracted and painful process, it is likely that workers will sooner or later start questioning the merits of the transition program. The trade unions, which have in many cases been channeling membership dues and significant income from property rentals to the strike funds, may become powerful adversaries. With private employers and governments likely to be relatively weak, the entire transition process could be endangered. A more cooperative approach to industrial relations from the very start of the reform may hence be a superior strategy. Note that there are obvious advantages to a participatory rather than an adversarial system; participation has historically worked well in Austria and the Federal Republic of Germany. Workers' councils or similar mechanisms emerged spontaneously in Central Europe whenever central controls were lifted (after World War II, as well as during the Communist era) in the former GDR and Poland in 1953, in Hungary in 1956, in Czechoslovakia in 1968, and in Hungary and Poland in the 1980s.

Finally, unemployment has been growing rapidly and is likely to become an increasingly divisive social and economic problem. With fiscal problems worsening over time, this aspect of the labor market will require special attention.

REFERENCES

- Blinder, Alan S., ed. 1990. *Paying for Productivity*. Washington, D.C.: Brookings Institution.
- Boeri, Tito, and Mark Keese. 1992. "Labour Markets and the Transition in Central and Eastern Europe." Organization for Economic Cooperation and Development, Paris.
- Bruno, Michael. 1992. "Stabilization and Reform in Eastern Europe: A Preliminary Evaluation." International Monetary Fund, Washington, D.C.
- Dervis, Kemal, and Timothy Condon. 1992. "Hungary: An Emerging Gradualist Success Story?" Paper presented at the conference on "Transition in Eastern Europe," National Bureau of Economic Research, Cambridge, Mass., February 26-29.
- Freeman, Richard B. 1992. "What Direction for Labor Market Institutions in Eastern and Central Europe?" Paper presented at the conference on "Transition in Eastern Europe," National Bureau of Economic Research, Cambridge, Mass., February 26-29.
- Hinds, Manuel. 1991. "Issues in the Introduction of Market Forces in Eastern European Socialist Economies." In Simon Commander, ed., *Managing Inflation in Socialist Economies in Transition*. EDI Seminar Studies. Washington, D.C.: World Bank.
- Jackman, Richard T., Richard Layard, and Stephen Nickell. 1992. "Unemployment in Eastern Europe." Paper presented at the conference on "Transition in Eastern

- Europe," National Bureau of Economic Research, Cambridge, Mass. February 26-29.
- Kiguel, Miguel, and Nissan Liviatan. 1991. "Stopping Inflation: The Experience in Latin America and Israel and Implications for Eastern Europe." In Vittorio Corbo, F. Coricelli, and J. Bossak, eds., *Reforming Central and Eastern European Economies*. A World Bank Symposium. Washington, D.C.
- Prasnikar, Janez, and Jan Svejnar. 1991. "Workers' Participation in Management vs. Social Ownership and Government Policies: Yugoslav Lessons for Transforming Socialist Economies." *Comparative Economic Studies* 33(4):27-46.
- Prasnikar, Janez, Jan Svejnar, Dubravko Mihaljek, and Vesna Prasnikar. 1991. "A Test of Enterprise Behavior under Yugoslav Labor-Management." University of Pittsburgh, Department of Economics, Pittsburgh, Pa.
- Riveros, Luiz. 1990. "Labor Market Policies and Market-Oriented Reforms in Socialist Economies." World Bank, Country Economics Department, Management and Growth Division, Washington, D.C.
- Rodrik, Dani. 1992. "Foreign Trade in Eastern Europe's Transition." Paper presented at the conference on "Transition in Eastern Europe," National Bureau of Economic Research, Cambridge, Mass., February 26-29.
- Svejnar, Jan. 1986. "Bargaining Power, Fear of Disagreement and Wage Settlements: Theory and Evidence from U.S. Industry." *Econometrica* 54(5):1055-78.
- Svejnar, Jan, and Katherine Terrell. 1991. "Reducing Labor Redundancy in State-Owned Enterprises." Policy Research Working Paper 792. World Bank, Infrastructure and Urban Development Department, Transport Division, Washington, D.C.
- Terrell, Katherine. 1992. "Productivity of Western and Domestic Capital in Polish Industry." *Journal of Comparative Economies* 16(3):494-514.
- Windmuller, John, P. 1971. "Czechoslovakia and the Communist Union Model." *British Journal of Industrial Relations* 1(1):33-54.

COMMENT ON "LABOR MARKET ADJUSTMENT
IN TRANSITIONAL ECONOMIES," BY SVEJNAR

Marek Góra

The transition to a market economy in Central and Eastern Europe has proved to be much more complicated than economists, politicians, and private citizens had predicted. The macroeconomic stabilization programs instituted between 1989 and 1991 should be seen as elements in the first stage of the transition. But different issues are involved in the subsequent—microeconomic—stages that must follow. Jan Svejnar's paper discusses this next stage and comments on the pressures that labor markets exert on the transition process. I agree with his diagnosis; we need to recognize that these pressures have a crucial influence on economic decisionmaking.

In the pretransition period, the accepted view was that Eastern Europe was in economic distress because of the command system and Soviet domination (which was true in part). Hence it was expected that removing these national and international obstacles, establishing democracy, and converting to free markets would lead almost immediately to prosperity. This assumption, unfortunately, was an illusion, but it helped advance the proposition that the more radical the reforms, the faster would be the transformation to a prosperous future. Thus radical stabilization programs were introduced and, in general, successfully completed. Since then, however, mounting economic problems have generated resistance, and the public has been less than enthusiastic about the next stage of the transition process.

At this stage, however, when changes in work habits, living standards, and ways of thinking are essential, we must face the fact that most people have very little—if any—commitment to reform left. Yet the behavior patterns associated with a command economy are not appropriate in a market-driven economy. What incentives can be devised to persuade the labor force to adjust to the new economic environment?

I. THE HERITAGE OF THE PAST

Professor Svejnar notes that a number of features of the command system tended to insulate labor markets, making adjustment to competitive markets

Marek Góra is adjunct professor of economics at the Warsaw School of Economics.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

difficult. First, full employment, based on artificially high demand for labor, was guaranteed. Second, workers spent a substantial amount of time and effort outside their official workplace and received part of their income from this kind of activity (swapping goods and services, for example). Third, the state provided many free goods and services and subsidized numerous others. And fourth, industrial relations were highly centralized, with very little, if any, scope for talented individual workers.

I would add to Svejnar's list another negative influence: the lack of motivation to work. Under the command system generations of workers had nothing to lose and nothing to gain from the work effort they expended. If they worked harder, they were not paid more. If they had been better paid, they would have not worked harder. In my opinion, this legacy of indifference is the most destructive element in the situation.

Although these features were common throughout Eastern Europe, their scale and intensity varied from place to place. In any discussion of the potential of these economies, it is important to remember that the past exerts a strong influence on the labor market today.

II. ADJUSTMENT IN THE LABOR MARKET

Svejnar's model of the employment and wage-setting process in the transitional economies yields a formula for the marginal revenue product of labor in these economies. On this basis he offers a two-stage outcome of transition. In the first stage, employment and wages fall, and labor redundancy rises. This theoretical result is fully supported by empirical evidence in all the economies in transition. In the light of Svejnar's analysis, what we observe is not the result of policy mistakes or a random outcome of the change. Labor market adjustment seems to be an element of the move from a centrally planned economy toward a market-oriented economy.

The model concludes that the outcome of the second stage will be a further decline in employment accompanied by a rise in wages. Although I would like to believe this scenario, I am afraid that another outcome is also possible. If transition does not proceed quickly enough (particularly as regards privatization), we may see a kind of economic dualism. State-owned firms may play the role of the "noncapitalist" or "indigenous" sector and thus serve as a reservoir of labor for the "capitalist" sector. As a result, newly established private firms will operate in a context of unlimited labor supply. Low productivity in the state sector will prevent the wages of those workers from rising, while wages in the private sector—although higher—will not increase because, at a ruling-wage rate, the supply of labor will be greater than demand.

Svejnar also suggests that a reduction of redundant labor, which is unavoidable (and may be beneficial for state-owned enterprises), will parallel the absorption of that labor by new private firms. I would note, first, that the scale of the expected reduction in employment in the state sector will significantly exceed the demand by new establishments and, second, that the structure of

this new demand is unlikely to correspond to the structure of the labor supply. This mismatch is likely to affect labor markets.

In some ways the labor market situation may be even worse than the statistics suggest. On the demand side, the end of labor hoarding meant that even if production had not fallen, redundant labor would have been dismissed. But in fact production fell much more than employment, so labor hoarding actually increased. Effective labor demand decreased more than the numbers on employment and vacancies suggest.

On the supply side, the fall in real wages induced an increase in the labor supply as a result of a strong income effect and a weak substitution effect. These, in turn, are a consequence of the elimination of many elements of the traditional parallel economy. Hence people are ready to supply more labor than the unemployment figures suggest. These effects may have significant lagged results in the future.

III. PARTICIPATION

Svejnar argues that labor should be involved in management affairs. It is difficult to know whether participation is a good or a bad option for transitional economies. It is probably a good one, but it may also be an obstacle. In Poland, for instance, workers were granted extensive rights in the early 1980s (paradoxically, just after the imposition of martial law). But because these rights were de jure only, labor had no influence on the situation. Then, virtually overnight, the de jure position became de facto; trade unions and workers' councils claimed real power. In some state-owned enterprises, particularly the largest ones, labor's influence slowed or even blocked necessary reforms. Workers, fearing that a change would mean substantial layoffs and slight, if any, increases in wages for those who remained, were anxious to maintain the status quo.

If labor is to play an active role in East Europe, its participation should be based on clearly established roles. The legacy of the command system of industrial relations may not promote the most desirable outcome.

IV. POLICY IMPLICATIONS

Svejnar's paper draws policy-oriented conclusions in regard to the productivity of state-owned enterprises and their employees and the importance of avoiding undesirable economic, social, and political externalities. I would like to stress that the transition process has winners and losers. Right now, a much larger number of people (usually employees of large state-owned enterprises) are convinced that they are losers. This group, which constituted the driving force for political change, is in a strong collective position. The labor situation, particularly for this group, is bad and is getting worse. We should not overlook the likelihood that these "losers" will resist the new ideology. This may make the transition either endless or biased. In both cases, the outcome will be destructive.

To avoid this possibility, I agree with Svejnar that a more cooperative approach to industrial relations is crucial. The question is how to convince workers and managers to shed their antagonism and work cooperatively in the process of transition.

Information Flows and Discrimination in Labor Markets in Rural Areas in Developing Countries

Andrew D. Foster and Mark R. Rosenzweig

A variety of implications of a pervasive labor market information problem—the inability of employers to perceive perfectly the skills of heterogeneous workers—are considered. The principal objective is to provide empirical evidence on the effects of imperfect information in order to achieve a better understanding of how workers are rewarded and to identify potential interventions that may reduce the costs associated with such problems. The emphasis is on the casual labor markets that characterize most rural areas of developing countries. Among the topics addressed are the informational assumptions of nutrition-wage theory; the processes by which employers learn about differences in worker capabilities and their implications for changes in wage inequality over the life cycle of workers; and worker mobility, information-based efficiency costs, and the identification of employer discrimination among workers on the basis of worker traits unrelated to employer-perceived productivity.

Two principal issues in the study of labor markets with heterogeneous workers concern how employers select workers for employment and how worker contributions are rewarded. An essential ingredient of the processes by which heterogeneous workers are matched to jobs and rewards are allocated to workers is the information available to employers about workers. Employers must be able to discriminate among workers in order to match optimally worker characteristics with particular tasks. Thus, the quality of information available in the labor market is an important factor in the efficiency of the market, and because it determines which workers are rewarded well and less well, it can have a significant effect on the distribution of income.

A number of theories of labor markets have focused on information problems in the labor market. But they have for the most part been concerned with moral hazard and with effort incentives for unmonitorable, but homogeneous, workers. Much less attention has been paid to the sorting of heterogeneous workers among employers and tasks in markets in which employers have imper-

Andrew D. Foster is assistant professor and Mark R. Rosenzweig professor in the Department of Economics at the University of Pennsylvania.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

fect knowledge about a worker's potential productivity. Moreover, the empirical evidence on the importance of employer ignorance and its relevance to developing countries and to the process of development have received little attention. Many empirical studies of labor markets have failed to take into account the problems of information barriers, with potentially misleading results.

In this essay, our principal aim is to provide empirical evidence on the effects of information problems in labor markets to better understand how workers are rewarded and to identify potential interventions that may reduce the costs associated with such problems. We focus on the casual labor markets that characterize most rural areas of developing countries. These markets are important in most of the world and readily permit the empirical study of information barriers. By understanding the contrast between the features of these markets and of those characterizing industrial countries, it is possible to assess the implications of information costs for development.

In the first part of the paper we discuss the informational assumptions of nutrition-wage theory. The empirical evidence on this theory is problematic because of the absence of attention to information problems. We also discuss the important implications of the processes by which employers learn about differences in workers' capabilities for changes in wage inequality over the life cycle of workers. These processes affect how changes in the age structure and in worker mobility within a labor market alter income inequality and labor market efficiency. The efficiency costs implied by the asymmetry of information between workers and employers are discussed in the context of efficiency wage models incorporating adverse selection. Finally, we discuss the information problems faced by employers to identify whether wage inequality is also determined by employers' preferences for or against workers on the basis of traits unrelated to productivity.

In the second part of the paper we employ data sets from the Philippines, India, and Pakistan to test several implications of information theory in rural labor markets of developing countries. Our evidence indicates the existence of considerable heterogeneity in workers' contributions to output, as measured by the interworker variability in piece-rate wages. Whereas a substantial portion of individual workers' skills is rewarded, the evidence suggests that there is also considerable ignorance among employers about individual differences in workers' abilities. One example is the evident lack of employer knowledge about individual differences in the food consumption of workers.

The data also indicate that there is adverse selection in casual labor markets, as implied by information theory. Evidently because employers are not fully able to identify the better workers, such workers tend to participate less in jobs paying time wages. Employers learn about worker abilities as workers' participation ("exposure") in the local labor market and job tenure increase. This finding implies that equally able workers with less participation and employer attachment are *ex ante* more costly to hire. Thus, women are at a disadvantage

because on average they have less time on the job than men. But in our tests of taste discrimination, which incorporate the limits on employers' information about workers, we find no evidence of taste discrimination against women in either the Philippine or Indian data. There is, however, statistical discrimination: among workers with the same productivity, women are paid lower time wages. That there is both statistical and taste discrimination against low-caste workers in the Indian village from which we have data becomes apparent only when attention is paid to employers' imperfect knowledge about workers' capabilities.

The empirical evidence presented in this paper is not meant to be conclusive regarding the effects of information barriers in labor markets in developing countries. It suggests, however, that further attention to these issues may have significant payoffs for understanding the consequences of economic growth and designing policies that contribute to equity and efficiency in labor markets. We discuss some of these implications in the concluding section.

I. LABOR MARKET THEORIES AND INFORMATION PROBLEMS

Many models of labor market processes assume that fundamental problems in wage setting are that employees are heterogeneous in their abilities or contributions to output and that employers do not have full knowledge of each worker's productivity. If we let the productivity of worker i be μ_i , then worker i 's productivity as perceived by employers, μ_i^* , is based on a set of observed worker characteristics:

$$(1) \quad \mu_i^* = E(\mu_i | Z_{mi} \forall m \in M_E).$$

Equation 1 states that worker i 's productivity as perceived by employers (μ_i^*) is the expected value of worker i 's productivity, based on the employer-observed characteristics of the worker, where Z_{mi} is the m th characteristic of worker i in the set of characteristics observed by the employer (M_E). The relation between the productivity of worker i and worker i 's productivity as perceived by employers is given in equation 2:

$$(2) \quad \mu_i = \mu_i^* + u_i$$

where u_i is the component of actual productivity that is unknown to employers. There is imperfect information in the labor market if $\text{var}(u_i) > 0$. The import of equation 1 is that rewards paid to workers by employees will depend on μ_i^* . The rewards depend on true productivity (μ_i) only to the extent that μ_i^* and μ_i are correlated. The wedge between rewards paid to workers and their productivity represents an allocative cost to an economy. It also has implications for the distribution of earnings and for the types of contractual arrangements between employees and employers.

A theory of labor markets is incomplete without explicit assumptions about the limits on information about workers and about the processes that form employers' perceptions of worker productivity. Thus, we need information on

what observable worker characteristics enter into M_E and how the information varies with respect to contractual arrangements between employer and employee and other features of the economic environment.

Nutrition Efficiency Wage Theory and Information

One of the most influential models of wage setting in developing countries, nutrition efficiency wage theory, makes a very strong assumption about the employers' information set. This theory posits not only that there is a significant relationship between worker productivity (μ_i) and food consumption that is known to all employers, but also that employers know the food consumption of each individual worker and offer time wages on the basis of that knowledge. The theory is important because it provides an explanation for a so-called natural floor on wage rates in rural areas in developing countries. The theory thus has strong implications for the relationship between efficiency and inequality in such settings, both within and across households. The wage minimum exists in this model because there is some level of consumption below which an increase in the time wage increases worker productivity by more than the wage increase.

Even without the hypothesized nonlinearity in the relationship between worker productivity and consumption, which sets a minimum wage, the model predicts that wages paid will be closely tied to consumption. It predicts that better-fed workers receive higher wages in competitive labor markets and lower wages in monopsonistic markets (since such workers will be more efficient than less well fed workers at the same wage). Thus, the operation of the labor market either exacerbates or reduces income inequality. By contrast, in conventional models (with perfect markets), wealth (and thus the ability to finance consumption) is independent of wage rates. Moreover, if consumption at the margin is differentially rewarded across jobs, the distribution of foods within households will depend on the occupational distribution of household members (Pitt, Rosenzweig, and Hassan 1990).

The nutrition-based efficiency wage model has intuitive appeal because of its potential relevance to developing countries, where consumption levels are low and are likely to be closely tied to income. The strong informational assumption of the model is that the individual consumption of workers is part of employers' information set, M_E . Although critical to the model, this assumption is not appealing if employers do not know which job applicants have a higher level of consumption. Employers then have no basis for rationing jobs among particular workers, and wage rates paid on a time basis cannot be related to consumption. Despite this information problem, only the assumption of the model concerning the productivity-consumption relation has received appropriate empirical attention.

Strauss (1986) tested the productivity-consumption relation on the basis of Sierra Leone data, and Deolalikar (1988) tested it using data from the semiarid tropical region of India. They estimated farm-level production functions incorporating worker consumption to test the effect of food consumption, nutritional

status, or both on workers' productivity. Both studies found that increased calorie consumption did significantly affect productivity (the obvious simultaneous relation between consumption and production was taken into account). Neither study, however, tested appropriately the second assumption of the model, that daily time wages paid by employers were influenced by worker consumption. In the Strauss study it was noted that wage rates, by sex, did not vary across workers.

Deolalikar attempted to test directly whether calorie consumption affected wage rates. He failed, however, to distinguish between time-wage rates and piece-rate wages. Piece rates will always reflect worker productivity. If most of the wage rates paid are on a piece-rate basis, Deolalikar's finding of a positive relation between worker consumption or health and wage rates may merely indicate that consumption augments productivity; it does not say anything about employers' knowledge of individual worker consumption. It is known, for example, that in the survey areas harvest operations, which employ the most people, are almost exclusively paid on a piece-rate basis (Walker and Ryan 1990). Indeed, if food consumption is only rewarded at harvest time, it may exhibit seasonal patterns that are wholly the result of changes in payment regimes over the year.

The distinction between the effects of workers' food consumption on time-rate wages and piece-rate wages provides evidence concerning the information set of employers. Strauss's observation that time wages in his sample area of Sierra Leone were more or less uniform, despite heterogeneity in consumption levels and consumption-productivity effects, is suggestive of the importance of the information problem. Experimental evidence reported in Basta and others (1979) also illustrates the consequences of constraints on employers' abilities to discern consumption and the importance of distinguishing between payment methods. In that experiment, randomly selected anemic plantation workers in Indonesia were given either an iron supplement or a placebo. The study found that in relation to the workers given the placebo, the workers given the iron supplement and paid on a piece-rate basis received higher wages, while those paid on a time basis did not. Given that the supplement was random and temporary, it is not surprising that employers paying time wages had little or no basis on which to discriminate among workers with respect to their iron consumption.

In nonexperimental local labor markets, employers may have better information for predicting workers' consumption over the long term. Employers' ability to discern and reward workers using time wages based on workers' consumption is a necessary condition for the distributional implications of nutrition efficiency wage theory. But there is little evidence on this condition in rural labor markets of developing countries.

Life-Cycle Inequality in Earnings and Information Costs

The now-conventional model of the evolution of earnings over the life cycle emphasizes workers' and firms' investments in human capital (Mincer 1974). In this framework, labor markets may be competitive and there may be no differences in workers' abilities, but workers may choose different life-cycle investment

paths. Earnings inequality would first decrease, on average, with worker time in the labor market (work experience) and then increase until worker retirement. In a world with no human capital investments and a labor market information problem, however, the life-cycle trajectory of earnings inequality is quite different. It is a well-known result from basic probability theory that the variance of equation 1 will be larger the more information employers have about μ_i (the larger is employers' information set, M_E).¹ The labor market experience of workers may lead not to increased worker skills but to the transmission to employers of better information on workers' time-invariant abilities. Then the variance in worker earnings, based on employer expectations, will increase monotonically with worker experience, which in this case might be better named worker exposure. Intuitively, employers' priors on the unobserved ability of workers will be relatively homogeneous. It will converge to the true distribution of abilities, which are less homogeneous, with increased observations of workers.

With respect to the time path of earnings (wage rates) for individuals, in the Mincer world of homogeneous (preinvestment) abilities, earnings of all workers will either be flat or will rise over the life cycle. In a heterogeneous world without investments, however, some workers (those with below-average μ_i) will experience declines in wages paid, while others (with above-average productivity) will experience wage increases as worker abilities are revealed. On average, in the human capital model with no heterogeneity, earnings rise with age; in the heterogeneous worker model with information problems, average wage rates are constant over the life cycle, or a major part of it. Moreover, in the Mincer-style model with no heterogeneity, lifetime earnings are the same for all workers. In the imperfect information model workers with greater ability have larger lifetime earnings, but because of information barriers, the differential between the lifetime earnings of low- and high-productivity workers is attenuated the greater and the more persistent is employer ignorance. This comes about because initially (among workers with little exposure in the labor market) high-productivity workers are underpaid and low-productivity workers are overpaid. In the extreme, when employers know nothing and learn nothing about worker productivities, all wages are identical.

In a world in which there is both human capital investment and employer learning, the implications of information considerations for life-cycle wage inequality hold within observable investment groups. Glaeser (1992) directly tests some of the implications of information theory in labor markets in the United States. He shows that residuals from wage regressions on worker schooling and

1. That is, if e is measurement error that is uncorrelated with any of the elements in the information set and M_{E1} is a larger set than M_{E2} , then

$$E(\mu_i | Z_{mi} \forall m \in M_{E1}) = E(\mu_i | Z_{mi} \forall m \in M_{E2}) + e$$

and

$$\text{var}[E(\mu_i | Z_{mi} \forall m \in M_{E1})] > \text{var}[E(\mu_i | Z_{mi} \forall m \in M_{E2})].$$

potential work experience increase with worker experience. In a more sophisticated analysis based on Bayesian models of learning with the same longitudinal data, Glaeser also shows that employer updates of workers' productivity shrink faster the longer a worker has been with a particular employer. The updates do not shrink very much, given job tenure, with total labor market experience. Thus, in an industrial economy such as that of the United States, there appears to be employer-specific accumulation of knowledge about individual workers that is not readily transferable, which may have implications for job mobility.

It is unclear how relevant findings from industrial countries on the role of information are to the casual labor markets in developing countries. In the latter, worker-employer attachment is rare, job tenure intermittent, and workers' geographic mobility relatively low. Thus, worker exposure in the local labor market may be closely related to worker time-wage compensation through general employer learning in such environments. Moreover, in casual labor markets, work is mostly based on rudimentary, physical tasks. It is likely that learning by doing is of considerably less importance than in industrial countries, and it may be less important than accumulation of knowledge by employers. This means that the implications of information theory for life-cycle wage inequality may be more salient in developing countries. Interventions dealing with information problems in those contexts may have dramatic effects on the distribution of earnings.

Tests for employer learning may also be clearer in casual labor markets. In industrial countries human capital accumulation is important, but worker abilities and worker investments may be correlated, making the identification of employer learning more difficult. In industrial countries wages rise faster for the inherently more able workers. Thus, wage deviations from average wage growth rates increase for the more able and decrease for the less able. This mimics the patterns expected if employers learn about worker abilities and if worker investments are either nonexistent or are orthogonal to ability.

The two distinct causes of the life-cycle growth in wage inequality are employer learning and investment in human capital. If the growth in wage inequality over the life cycle of workers is the result of employer learning, that is evidence of information barriers and, perhaps, of market failure. Such barriers may be reduced by means of appropriate labor market policies, with possible overall efficiency benefits to an economy (although reductions in information costs, as noted, exacerbate lifetime wage inequality). The life-cycle growth of wage dispersion as a result of complementarities between worker abilities and human capital investments does not obviously require policy attention, but such attention may be needed if the correlation is caused by failures in markets that provide the resources for investments in human capital.

Efficiency Costs of Information and Adverse Selection

In addition to affecting the distribution of wages in a population, incomplete information is costly to an economy and can have important implications for efficiency. In particular, inefficiency can arise from four sources: misallocation

of workers to tasks and sectors; supervisory costs; unemployment; and the inability fully to capture returns to investment.

Misallocation of workers to tasks and sectors. In the absence of complete information, it may not be possible to allocate a worker to a task or sector of the economy in which he or she has comparative advantage. Consider, for example, a model in which a worker works for a wage w_k , performs task k , and receives utility $U(w_k - k)$. If workers are paid on a piece-rate basis, it is easily seen that the most productive workers will allocate themselves to the high k tasks. All workers experience the same cost of doing high k tasks, but the more productive workers earn more than less productive workers earn on a piece-rate basis. Thus, for example, if the piece rates for two tasks are such that less productive workers are indifferent between the low and the high k task, the more productive workers will strictly prefer the high k task and will allocate themselves accordingly. In effect, the high-productivity workers have comparative advantage in the production of more difficult (that is, high k) tasks, and as a result specialization by them in these tasks is efficient.

If workers are paid a time wage, however, and if employers cannot distinguish between more and less productive workers, it will no longer be the case that the most productive workers will be allocated to the high k tasks. With time wages there is no way of differentially rewarding more productive individuals. In this case, whenever the less productive workers are indifferent between the low and high k tasks, given the time wage paid for that task, the more productive workers will also be indifferent, and thus there is no incentive for workers to allocate themselves on the basis of their comparative advantage. The inability to allocate workers to tasks in which they have comparative advantage leads to a loss of efficiency.

Supervisory costs. It is argued that a piece-rate payment system will be fully efficient even in the presence of employer uncertainty. But why are time wages paid when there is employer uncertainty about worker abilities? The answer is that piece-rate employment has associated costs that do not arise in the context of time-wage employment. It has been argued that the costs of monitoring output quantity and quality will be higher in piece-rate employment for technological or other reasons (Roumasset and Uy 1980). Foster and Rosenzweig (1991) show that for low levels of employment, supervisory costs are higher when piece-rate employment is used than when time wages are used. The cost of supervising piece-rate workers is thus both a barrier to the efficient allocation of workers by task and a source of inefficiency. Of course, supervisory costs in the time-wage sector also represent a loss in efficiency that may be attributable, in part, to problems of information.

The possibility that informational asymmetry can lead to equilibrium unemployment as well as to an inefficient allocation of workers across sectors of the economy has been examined in detail by Weiss (1980). Individuals in rural areas

of developing countries face the opportunity to work in the time-wage labor market, in which unknown productive differentials cannot be compensated, as well as the opportunity to work on their own farms or on a piece-rate basis. But time wages underpay the most productive workers, while the latter types of employment fully reward productivity differentials. Therefore, the composition of workers willing to work for a particular time-wage employer will be affected by the level of time wages. From an alternative perspective, the reservation wage for time-wage employment will be an increasing function of the level of productivity. Although the Weiss model focuses on understanding unemployment in industrial countries, it is arguably a more appropriate characterization of labor markets in developing countries. Indeed, as Weiss points out, "the positive correlation between acceptance wages and productivity is easily justified in the context of a less developed country where acceptance wages are determined by output in agriculture and handicrafts" (Weiss 1980, p. 29).

The simplest version of the Weiss model incorporates the assumption that productivity differentials are completely unobservable to the employer but are known to the employee. Under these circumstances we may let $q(w)$, where w is the wage offered in the time-wage sector, denote the efficiency of the worker who is indifferent between working for time wages and working elsewhere. The assumption that more productive workers have higher reservation wages implies that $q'(w) > 0$. Because only workers with efficiency $q \leq q(w)$ will be willing to work in the time-wage sector if they are offered a wage of w , the average efficiency of workers in the time-wage sector $[Q(w)]$ may be written:

$$(3) \quad Q(w) = \frac{\int_0^{q(w)} q dF(q)}{F[q(w)]}$$

where $F(q)$ is the distribution function of worker efficiency in the labor market. From equation 3, $Q'(w) > 0$: a higher wage draws more productive workers into the time-wage sector and thus raises the average productivity in that sector.

Adverse selection occurs when wage rates affect the distribution of workers. Its empirical implication is that individual workers with higher productivity will participate less in the time-wage market when they have an opportunity to work on a piece-rate basis. We are unaware of any empirical evidence demonstrating that phenomenon, however, despite its central role in this influential model.

Unemployment. Unemployment is one possible result of the adverse selection equilibrium. Under asymmetric information, equilibrium in the labor market depends on supply and demand conditions at the wage that minimizes the cost per efficiency unit of labor, w^* , where $w^* = w/Q(w)$ and $Q(w^*) = w^*Q'(w^*)$. If labor demand exceeds supply at this wage, competition among firms will bid up the wage, which then attracts more productive workers until the labor market equilibrates. If labor supply exceeds demand, however, there will be queues in the time-wage labor market. Some workers will strictly prefer work in the time-wage sector over work in the piece-rate sector at the going time-wage rate.

Employers will not employ these workers at a lower wage because to do so would raise the cost per efficiency unit of labor.

The adverse selection model also yields implications for the distribution of jobs across workers differing by observable characteristics that signal productivity, such as sex or age. Suppose that potential employees can be divided into a number of different groups with similar observed characteristics but that there remains substantial heterogeneity in productivity within each group. As Weiss shows, equilibrium in this more complex labor market can be identified by first ranking each group by the minimum cost per efficiency unit of labor. Employers will then choose workers on the basis of their observed characteristics, selecting first that group with the lowest cost per efficiency unit. If labor supply in this group exceeds demand, its members will face job queues, and members of other groups will be excluded entirely from the time-wage labor market. Otherwise, other groups may be employed in inverse order of the minimum cost per efficiency unit of labor.

Under these circumstances the allocation of workers to jobs will depend not on differences in average productivity but on the distribution of perceived productivity within each group. This has an important implication for discrimination. An application of the envelope theorem to the cost-minimization problem yields the result that a mean-preserving increase in the spread of the distribution of unobserved productivity which lowers the proportion of workers willing to work at the cost-minimizing wage will raise the minimum cost per efficiency unit of labor.² Informally, the latter condition is likely to be met under the same circumstances that result in the rationing of jobs in the time-wage sector: a relatively high cost-minimizing wage. Thus a group of workers for which information is less precise will be more likely to be excluded from the time-wage sector than one for which information is more precise. In addition, as established above, employer information will be more precise for individuals who are more active in the labor market. This result implies that individuals with less labor market experience as a group will find it more difficult to find time-wage employment even if their lack of work experience has no effect on their actual productivity.

2. Let $c^*(\eta)$ denote the minimum cost per efficiency unit of labor arising in a group with an unobserved productivity distribution $F(q;\eta)$, where η parameterizes the distribution. An application of the envelope theorem to the cost-minimization problem yields the following result:

$$\frac{dc^*(\eta)}{d\eta} > 0 \leftrightarrow F[q(w^*)] \int_0^{q(w^*)} G(q;\eta) dq - G[q(w^*)] \int_0^{q(w^*)} F(q;\eta) dq > 0$$

where

$$G(q;\eta) = \frac{\partial F(q;\eta)}{\partial \eta}.$$

The first term on the right side is positive whenever $\eta' > \eta$, which implies that $F(q;\eta')$ second-order stochastically dominates $F(q;\eta)$. Thus, if $G[q(w^*);\eta] < 0$, a mean-preserving increase in spread will raise the minimum cost per efficiency unit of labor.

The implications of the Weiss model for efficiency stem largely from the question of whether groups that are excluded from time-wage labor have a comparative advantage in that sector. For example, more productive individuals may have comparative advantage in planting, but for technological reasons, planting tends to be paid on a time-wage basis. There would be a gain in efficiency if productivity differentials were known and these workers could be allocated to the time-wage sector.

The inability to fully capture returns to investment. The final efficiency cost associated with informational asymmetries arises because investments in human capital may not be fully rewarded if the resulting productivity differentials cannot be perfectly measured. Not only does this mean that there will be underinvestment in human capital, on average, but the allocation of human capital may also be less efficient. The returns to human capital investment for an individual will in part depend on the distribution of human capital among individuals with similar observed characteristics. But it is possible that more able individuals in a group for which signals of productivity are noisy will invest less than will less able individuals in a group for which the signals are more accurate (Lundberg and Starz 1983).

Taste and Statistical Discrimination

The adverse selection model shows that labor market information problems have implications for the interpretation and analysis of discrimination on the part of employers. This is especially relevant to the important issue of distinguishing between statistical and taste discrimination. Statistical discrimination, implied by equation 1, results from employers' efficient use of their information set. Taste discrimination refers to the notion that employers have a preference for employing certain kinds of workers. Net of *perceived* productivity, μ_i^* , as given by equation 1, such workers will receive higher wages.

In the absence of taste discrimination, a simple model with a competitive labor market yields the result that workers should be paid according to their expected productivity. In linear form, with or without some employer ignorance, but with taste discrimination, wages are paid according to

$$(4) \quad w_{ii} = \sum_{m \in M_D} \alpha_m^* Z_{mi} + \alpha_\mu^* \mu_i^*$$

In equation 4, w_{ii} denotes worker i 's wages under taste discrimination, Z_{mi} represents the m th characteristic of individual i , and M_D describes the set of characteristics of an individual subject to taste discrimination.

In the empirical literature there is no information on μ_i^* and no knowledge of the set of worker characteristics (M_E) that employers use to predict worker productivity. These are fundamental problems in the identification of whether the α_m^* in equation 4 are nonzero. Most studies that attempt to measure the effects of taste discrimination include the variables concerning which employers may have preferences. They also include a vector of characteristics that are

hypothesized to be the set M_E which employers use to predict productivity. If employers know more than the researcher and the omitted component of μ_i^* is correlated with characteristics that are hypothesized to be subject to taste discrimination (that is, are preferred by employers given μ_i^*), then incorrect inferences result. If, for example, height is a determinant of productivity that is known to employers but not to the researcher, it may appear that men receive higher wages than women, net of other observed characteristics.

The existence of imperfect knowledge on the part of employers means that identification of taste discrimination (α_m^*) is not possible without knowledge of the employers' estimate of productivity, μ_i^* . This is true even if the researcher can measure workers' productivities exactly (μ_i is known) because worker compensation depends only on that part of μ_i that can be predicted on the basis of the characteristics known to the employer. Foster and Rosenzweig (1991) analyze a regression of time wages on actual productivity of the following form

$$(5) \quad w_{ti} = \sum_{m \in M_D} \alpha_m Z_{mi} + \alpha_\mu \mu_i$$

where M_D is the set of characteristics hypothesized to be subject to taste discrimination. They show that the coefficients in equation 5 will not yield consistent estimates of the parameters of equation 4, which can reveal taste discrimination.

Consider a population containing four kinds of individuals in equal proportions but with different levels of productivity. For example, the four groups can pick the following number of baskets of corn in an hour: productive men, four; unproductive men, two; productive women, three; and unproductive women, one. Note that men are on average more productive than women but that more productive women can pick more in an hour than less productive men.

Assume further that employers cannot distinguish between productive and unproductive individuals of the same sex, that monitoring costs preclude the use of piece-rate payments, and that labor is supplied inelastically by the four groups. Under these circumstances a nondiscriminating employer in a competitive labor market will pay individuals according to their expected productivity: all men will receive a wage of three and all women a wage of two. A regression of wage on actual productivity and sex (equation 5) will yield a coefficient of one on male and a coefficient of zero on productivity because all men receive a higher wage than all women. Thus, ignoring problems of information, the incorrect conclusions might be drawn that women are subject to discrimination and that productivity differentials are totally unrewarded.

Although estimation of such a regression can be misleading, it is nonetheless of some interest. The regression indicates that a man would have a higher wage than an equally productive woman. The result that efficient employers tend to pay higher wages to individuals who come from groups with above-average productivity is what is implied by statistical discrimination (see, for example, Aigner and Cain 1977). To distinguish this form of statistical discrimination from discrimination arising from the exercise of employer preferences that are

orthogonal to perceived worker abilities is the principal objective in the empirical literature on discrimination in labor markets.

II. THE DATA

Information on worker productivity is a critical feature of data that are needed to learn about employers' perception of, and the allocation of rewards to, workers' performance and abilities. A common practice in agricultural labor markets is the use of piece-rate payment methods in at least some agricultural operations. Given common payments per piece across workers and farms, piece-rate wages should provide a good measure of workers' productivities. In some environments data are available on piece-rate and time-based wages for each unit of time worked for the same workers. It is thus possible to directly assess the extent to which employers are cognizant of individual productivity differences across workers. Moreover, it is possible to examine the determinants of the accuracy of and the amount of bias in employer assessments of differences in worker productivity by comparing the time wages offered to workers with their productivity as measured by their piece-rate performance.

Of course, piece-rate wages for each time period may differ from time wages for the same worker in the same activity because of the moral hazard problems associated with time wages. We would expect, however, that piece-rate performance is the best predictor of productivity under time wages. It is reasonable to assume that more able workers in a task earn higher piece-rate wages and incur less disutility from work in that task. Thus, they are more productive in that task, compared with other workers, and less prone to shirk if supervision is lax.

We have selected data sets from the Philippines, India, and Pakistan to explore the issues of information asymmetries and employer bias. The data sets from the Philippines and India permit the comparison of piece-rate and time wages for individual workers. The data set from Pakistan provides information on time wages in an area in which there is evidently significant in-migration.

The Data Set from the Philippines

The data set from the Philippines is from a stratified random panel of 448 households in Bukidnon in northern Mindanao. The households were interviewed in four rounds at four-month intervals in 1984-85 as part of a study on the nutritional effects of export cropping. The study was carried out by Bouis and Haddad (1990) under the auspices of the International Food Policy Research Institute (IFPRI). For each round information is available on general household demographics, individual food intake, own- and off-farm labor activities, and wages for all family members by activity, crop, and payment method (piece rate or time).

Three important characteristics of labor markets in the survey area in the Philippines make that area particularly well suited to the study of information problems in casual labor markets. First, because of evident scale economies in supervision costs and differences in the scale of operations (land size) across

employers, both piece-rate and time-wage payment methods are used for the same operations in the same time period (Foster and Rosenzweig 1991). Second, because of a relative absence of seasonality, the same crop activities occur at different times of the year across different employers. Third, few workers work for the same employer for any length of time; they move from employer to employer as operations are finished. Thus, the labor market is almost exclusively a spot market, and most laborers are casual. Most important, because the employers who engage workers on a piece-rate basis are not the same as those using time wages, unbiased discrepancies between time wages and piece-rate productivity measure the extent of the diffusion of information on worker abilities across employers.³

The Data Set from India

The data set from India provides information on piece-rate and time wages from the Indian Village Level Studies of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). These data provide detailed information for up to ten years at approximately three-week intervals on all transactions, and on labor supply and wage rates for all individuals residing in a sample of forty households in each of ten villages in the semiarid tropics of India. As in the Bukidnon survey area in the Philippines, almost all workers are hired on a daily basis and work in their home village. The survey data provide information on the basic demographic characteristics, work time, and wage rates of all workers. In addition, there is information on the caste "rank" of all the households, based on a detailed classificatory study by a resident anthropologist (Walker and Ryan 1990). Thus, it is possible to study employer discrimination with respect to both the sex and the caste of workers.

One important shortcoming of the Indian data set—a feature it has in common with most other data sets—is that no direct information was elicited on the kind of payment method used in the labor market. For most of the Indian villages, however, agricultural operations are highly seasonal (unlike the Philippine case), and the survey provides information on the kind of activity (agricultural operation). For example, workers engaged in harvest operations in this setting are compensated exclusively on a piece-rate basis. Thus, it is possible to infer payment method and to identify, for the same workers, piece-rate and time wages.

Another unfortunate limitation of the Indian data set is that the information on worker activity (harvest, planting, and the like) is only provided for household transactions that cannot be linked to individuals. In three of the ten villages (Aurepalle, Rampura, and Shirapur) harvest operations did not overlap significantly with other operations during certain periods of the year because of the highly seasonal nature of this monsoon economy. Thus, in our tests on Indian

3. Piece-rate performance could differ across farms depending on the levels of complementary inputs. In the survey area, however, farms do not differ very much in capital equipment (harvest knives, rakes) used. The survey farms do differ in the scale of operations.

data we only use the information from these three villages. For tests performed using the Indian data set, we assume that a worker's harvest productivity, measured by his or her harvest piece-rate wages, is a relevant measure of productivity in the preharvest operations, which are paid on a time-wage basis.

Many employers will be able to observe the piece-rate performance of a worker whom they might hire at a different time in the year for time wages, albeit for a different agricultural task. This could imply that information on individual worker productivities is better known in the Indian context than in the Philippine setting, in which payment regimes differ across employers.

An important advantage of the Indian data set is that it has a significant longitudinal component. For two of the three villages from which piece-rate wage rates can be most precisely inferred, consistent time-series on individual wage rates and market participation are available over the years 1979–84. For one village (Rampura) such information is available for the period 1980–84. The longitudinal data permit assessments of how workers' exposure in the village labor market influences employers' ability to discern accurately worker productivity. Finally, in one of the village areas (Shirapur) a major governmental employment program nearby accounts for a significant share of the total wage employment of village residents. To the extent that this program offers a non-discriminatory (in the taste sense) opportunity for all workers, we can assess its impact on wage differences caused by the taste discrimination of the village employers, if we find evidence of such discrimination.

The Data Set from Pakistan

Another IFPRI data set, the Pakistan Food Security Survey, provides information on worker (time) wage rates in an area in which there is evidently significant in-migration. This survey was based on a stratified sample of 926 households in three major wheat-growing provinces of Pakistan—Punjab, Sind, and the Northwest Frontier Province. The sample was followed in twelve rounds over the period from July 1986 through September 1989, with information on length of residence in the village for each sample individual collected in the fifth round. Unfortunately, no information on piece-rate payments was collected in the survey, so it is not possible to compare true and employer-perceived productivity across workers classified by their years of settlement. We can, however, examine wage variances by length of residence.

III. THE EVIDENCE

In this section we employ the data sets from India, Pakistan, and the Philippines to test several implications of information theory for labor markets in developing countries. We present the results of statistical tests in six areas. First, we use data from the Philippines to test whether the information set of employers in a casual labor market includes the individual consumption of workers. Second, we use data from India and Pakistan to test the hypothesis that inequality in time wages among heterogeneous workers rises as the workers age.

Third, we use data from the Philippines to test whether the tenure of workers with specific employers provides informational benefits. Fourth, using data from India and the Philippines, we analyze the relation between perceived productivity and a series of employer-known measures of worker productivity and estimate the known and unknown components of the variance in worker productivity. Fifth, we use data from the Philippines to test for the presence of adverse selection. Sixth, we use data from India and the Philippines to test for discrimination based on sex and caste.

Is Worker Calorie Consumption Known by Employers?

We first test whether the information set of employers in a casual labor market includes the individual consumption of workers, a necessary condition for the existence of nutrition-based wage setting. As noted, the data set from the Philippines provides information on individual calorie consumption for all sample household members, including those who participated in the labor market as wage workers, in all survey rounds.

To test whether employers are able to discern and thus reward (or exploit) worker consumption, we need to establish that increased calorie intake augments actual productivity. Given that piece-rate wages reflect worker contributions to output, we can ascertain whether calorie consumption and piece rates are positively correlated. As is well known, however, wage rates may cause calorie consumption to increase. To eliminate this simultaneity problem, we employ instrumental variables. We assume that a worker's landownership (a proxy for wealth) affects his or her calorie consumption but does not, given consumption, directly affect wages. Either employers do not know worker landownership, or, more plausibly, it is irrelevant, given calorie intake, to productivity. We also assume that the sample period (round) affects calorie consumption because of seasonality in prices and imperfect ability to smooth consumption.

Table 1 reports the results of weighted two-stage least-squares estimation of the effects of calorie consumption on harvest wage rates by method of payment. (The weights are total days of employment.) In the first column, as in most of the econometric studies, we do not distinguish between the piece rate and time wages. The estimates indicate a positive relationship, net of the sex of the worker and his or her height and age. This would lead researchers to conclude that there is support for the competitive version of nutrition-based efficiency wages. We chose harvest wages, however, to get the most precise estimate of the effects of calorie consumption on productivity. Harvesting is characterized by the highest incidence of piece-rate wage payments (about 60 percent of total employment days) of all the agricultural operations in the Bukidnon area of the Philippines. The estimates in the first column may thus only reflect the productivity effects of calorie consumption. They do not necessarily indicate employers' recognition of worker consumption as reflected in the time wages rewarded.

Table 1. *The Effects of Calorie Consumption on Harvest Wage Rates, by Method of Payment, in the Philippine Survey*

Variable	Harvest wages			Difference between worker's within-round piece-rate and time wage
	All	Piece rate	Time	
Calories consumed ^a ($\times 10^{-3}$)	0.211 (2.36)	0.438 (3.13)	-0.0153 (0.22)	0.316 (2.59)
Age	0.0208 (0.82)	0.0153 (0.43)	0.0567 (2.02)	-0.074 (1.38)
Age squared ($\times 10^{-3}$)	-0.142 (0.38)	-0.216 (0.41)	-0.721 (1.70)	1.09 (1.41)
Height	1.04 (2.37)	0.446 (0.74)	0.040 (0.30)	1.91 (2.18)
Male	-0.366 (2.37)	-0.357 (2.66)	0.285 (0.61)	-0.628 (3.59)
Constant	0.747 (0.90)	1.40 (1.21)	1.59 (2.22)	-1.72 (1.15)
F(S,d.f.)	2.81	2.17	2.73	2.72
Degrees of freedom	322	286	108	188

Note: Results are from weighted two-stage least-squares estimation of the effects of calorie consumption on log wage, using total days of employment as weights. The absolute values of asymptotic *t*-ratios are in parentheses.

a. Endogenous variable. Identifying instruments include land owned by worker and dummy variables for round.

Source: Authors' calculations.

In the second column we report the estimates of calorie intake on piece-rate harvest wages. These indicate a significant positive productivity effect of increased calorie consumption. In the third column, however, we see that there is no effect of calorie consumption on harvest wages when they are paid on a time basis. The estimated time-wage calorie effect is not statistically significant and does not have the expected sign. Thus, when attention is paid to payment method, we see that there is no support for nutrition-based efficiency wage theory. Employers do not appear to be able to differentiate workers according to their calorie consumption, even though it evidently affects workers' contributions to output.

In the last column we report a test based on regressing the difference between a worker's piece-rate and time wages in a survey round on his or her (predicted) calorie consumption in that round. This test reveals whether the effect of increased calorie consumption on a worker's productivity, measured by the piece-rate wage, is greater than the effect on the worker's time wages. The estimate confirms what is indicated by the estimates in columns two and three: calorie intake augments productivity significantly more than it increases time wage rates. The information set of employers thus does not appear to contain worker consumption. This constraint on information makes it impossible, at least in the Philippine context, for employer wage setting to be nutrition based, even though it evidently would be advantageous for employers to take worker consumption into account.

Time in the Local Labor Market

If the achievement of more accurate assessments of worker productivity depends positively on the number of employer observations of workers' performance, then, in the absence of important investment effects, inequality in time wages among heterogeneous workers rises as the workers age. We test this using the longitudinal Indian data set and measuring productivity with piece rates. The general information effect of labor market experience is possible in the casual labor market because most of the workers are employed by farmers residing and farming in the same village (Rosenzweig 1988). Unlike Glaeser (1991), we do not need to make assumptions about the homogeneity of workers across experience groups. We can assess directly whether the discrepancy between piece rate (actual productivity) and time wages (employers' perceptions of productivity) for any individual worker decreases with the labor market exposure of the worker.

As noted, employer ignorance of worker ability decreases with the worker's labor market exposure. This would imply that the variance in the time wages of a group of workers that is heterogeneous with respect to productivity would increase with the age of the workers to the extent that age and average cumulative labor market exposure are positively correlated. The variance would increase even if labor market experience did not augment workers' productivity.

It is likely that in the agricultural environment characterizing the survey areas in both India and the Philippines there is little scope for learning by doing among workers because most operations are straightforward and mechanical in nature. Worker experience is thus likely to have little effect on either time or piece-rate wage rates, on average—a proposition we can test using the piece-rate productivity measure. Rather, some workers will experience a downward path of time wages and others a rising time path. In both cases the time wage rate approaches the time-invariant piece-rate wage if there is learning by employers.

We again employed a two-stage estimation procedure for two reasons. First, wage work experience is clearly endogenously related to the market wage of the worker. Second, the Indian data provide only a segment of any worker's total wage employment history, corresponding to the 1979–84 or 1980–84 survey intervals. We first regressed the actual number of days a worker participated in the wage labor market over the sample period on his or her age, age squared, sex, and landholdings. We then used this predicted wage market experience indicator in regressions in which the dependent variables were the piece-rate wage, the time wage, and the absolute value of the difference between the log piece-rate wage and the log time wage for the worker. The third dependent variable is a measure of employer ignorance of worker productivity.

The estimates of wage market experience on the three dependent variables are reported in table 2, which also includes estimates for controls for the sex and caste of the worker. The estimates indicate that neither piece-rate wages nor time wages are influenced by workers' market exposure; there appear to be no

Table 2. *The Effects of Wage Market Experience on Employer Ignorance, Piece Rates, and Time Wages in the Indian Survey*

<i>Worker characteristic</i>	<i>Employer ignorance^a</i>	<i>Log piece-rate wage</i>	<i>Log time wage</i>
Cumulative days in market ($\times 10^{-3}$)	-0.285 (2.27)	0.217 (0.79)	0.000954 (0.04)
Male	-0.0107 (0.43)	0.534 (9.85)	0.484 (10.8)
Low caste	-0.0348 (1.14)	-0.0378 (0.56)	-0.0681 (1.23)
Literate	-0.0392 (1.30)	0.0214 (0.33)	-0.0133 (0.25)

Note: Results are from two-stage least-squares estimation of the effects of cumulative days in the wage labor market by worker on employer ignorance of worker productivity, log actual worker productivity (measured by piece-rate wages), and log time wages. There are 114 observations (one randomly chosen time period per wage worker). Specification also includes village dummy variables. The absolute values of asymptotic *t*-ratios are in parentheses.

a. The absolute value of the difference between the log piece rate and the log time wage.

Source: Authors' calculations.

direct experience effects on productivity. Consistent with the hypothesis that workers who have spent more time in the wage labor market are better known by employers, however, increased exposure in the wage labor market does reduce employer ignorance. In particular, for each year (300 days) a worker is in the wage labor market, although working for different employers, the absolute value of the difference between his or her piece-rate wage and time wage declines by 6.5 percent.

The estimates also imply that, net of labor force exposure, there is no difference in employer uncertainty about productivity across men and women or across low-caste and other workers. Because women in these economies on average spend less time in the local labor wage market than do men, however, the estimates in table 2 imply that there will generally be more uncertainty about any given woman's productivity compared with that of any man with the same observable characteristics. Information theory suggests that employers would therefore find it more costly on average to hire women even if they were on average as productive as men.

The results in table 2 imply that in the casual labor markets of these Indian villages, wage inequality monotonically increases on average as the population ages without any significant skill augmentation by manual workers. Unfortunately, there are too few observations either over time or in a given year to describe accurately the time-wage distribution by worker age.

An additional reason why some workers are less known by employers, apart from their labor market participation rates and age, is that some workers arrive at different times in a local labor market. Newly arrived heterogeneous in-migrant workers should have lower variability in time wages than lifetime residents who are otherwise compositionally similar in productivity. The variance in wage rates should rise with length of residence, given age, for any group of heterogeneous workers. But the Indian villages have few wage labor in-migrants

Table 3. The Relationship between Years Settled in Village and Wage Dispersion for Male Wage Workers in the Pakistan Survey

<i>Years in village</i>	<i>Wage rate (1988 rupees)</i>		<i>Number of workers</i>	<i>Test statistic^a (d.f., d.f.)^b</i>
	<i>Mean</i>	<i>Standard deviation</i>		
0-5	24.3	5.23	15	2.70 (165,14)
6-10	24.3	7.93	16	1.36 (165,15)
More than 10	26.2	9.25	166	n.a.
10 or fewer	24.3	6.80	31	1.85 (165,30)

n.a. Not applicable.

a. Test of difference in wage standard deviation of row group with workers settled more than 10 years in village.

b. Significant difference at 0.02 critical value.

Source: Authors' calculations.

significantly above age twenty.⁴ Therefore we use the IFPRI Pakistan survey data to analyze the effect of worker in-migration on employers' information sets.

Using the data set from Pakistan, table 3 provides the average and standard deviation in wage rates for male daily wage workers by the number of years the workers had resided in the village. These figures, consistent with learning by employers in the villages in which these workers currently reside, show a monotonic rise in the standard deviation in wage rates as the average length of residence of the workers increases. The average wage rate of workers residing more than ten years in a village is 7.8 percent greater than the average wage of workers with ten or fewer years of residence, but this difference in mean wages is not statistically significant. The 36 percent differential in the standard deviation in the wage rates of these two groups of workers, however, is statistically significant. Wage inequality is smaller among new in-migrants than in the resident population, as implied by imperfect information theory.

Job Tenure Effects

The results from the Indian and Pakistan data suggest that information about workers with little or no attachment to particular employers becomes diffused among employers as the workers increase their participation in that market. The data set from the Philippines enables us to test whether the tenure of workers with specific employers also provides informational benefits. In that environment a proportion of the workers is more or less permanently attached to employers. We can compare the variance in time wage rates for these workers with that for the more prevalent temporary workers. If the continuity of employment with particular employers increases the speed at which worker productivity is discerned and appropriately rewarded, permanent workers should be characterized by a higher time-wage variance, given the same variance in pro-

4. The in-migrants are almost exclusively women who marry village residents (Rosenzweig and Stark 1989).

ductivity. We cannot obtain enough piece-rate wage observations for the small group of permanent workers to assess whether the variance in actual productivity differs across workers classified by their attachment to employers. We can, however, control for observable worker characteristics and kind of work (agricultural task and crop) by estimating a regression of time wages on these variables and comparing the residual variances.

Table 4 reports the gross and residual variances of the time wages of permanent and temporary workers. The gross wage variances that do not control for differences in worker heterogeneity are no different across the two groups. But the variances net of differences in observable worker characteristics and in kinds of work are, as expected, higher for the permanent workers, and the difference in variances is statistically significant. Thus, there is support for the hypothesis that worker tenure is negatively related to employer ignorance, although it is possible that there are remaining differences in the variances of worker productivity across these kinds of workers which are not captured by age, sex, or work task.

How Much Do Employers Know about Worker Productivity?

To measure the extent of employer uncertainty about worker productivity, we obtain an estimate of the variance in worker productivity. We then decompose the variance into the components that are known and unknown by employers. Estimates of the effect of a unit percentage increase in perceived productivity on the wage (α_μ^*) and the parameters of the measurement equations (λ_m) are presented in table 5, where the λ_m are coefficients associated with the measurement equations relating employer-perceived productivity to the measured characteristics in the data, Z_{mi} :

$$(6) \quad Z_{mi} = \lambda_m \mu_i^* + \epsilon_{zmi}$$

Because employers are assumed to use information efficiently, the residual component of the measurement equation (ϵ_{zmi}) must be uncorrelated with the unobserved component of productivity. Thus the measurement equations imply that Z_{mi} is a valid instrument for μ_i^* in estimating equation 4. In the data sets from both India and the Philippines, higher perceived productivity is associated

Table 4. *Gross and Residual Wage Variances for Permanent and Temporary Workers in the Philippine Survey*

Item	Category of worker		P-value
	Permanent	Temporary	
Wages	7.40	7.67	0.323
Residuals ^a	7.15	5.94	0.007

Note: There were 133 observations for permanent workers and 240 for temporary workers.

a. Residuals obtained from regression of time wages on crop, task, age, age squared, sex, and category of worker.

Source: Authors' calculations.

Table 5. The Relation among Time Wages, Employer-Observed Worker Characteristics, and Perceived Worker Productivity in India and the Philippines

<i>Item</i>	<i>India</i>	<i>Philippines</i>
Log time wage, ln W_i	0.960 (19.252)	1.476 (2.414)
Age	1.562 (0.399)	10.053 (1.508)
Age squared ($\times 10^{-2}$)	0.759 (0.175)	—
Education	1.094 (5.232)	1.558 (0.734)
Male	1.453 (19.897)	1.582 (2.500)
Height	—	0.197 (2.122)
Shirapur village ^a	-0.287 (2.530)	n.a.
Aurepalle village ^a	0.307 (4.964)	n.a.
Low caste	-0.176 (2.434)	n.a.

— Not available.

n.a. Not applicable.

Note: Results are from maximum likelihood estimation. The absolute values of asymptotic *t*-ratios are in parentheses.

a. Reference village is Rampura.

Source: Authors' calculations.

with significantly higher time wages.⁵ A number of other factors, including education, sex, village, and caste, are also associated with the employer-predicted productivity. There is nothing structural about these estimates; they simply reflect the covariance between these employer-observed attributes and productivity. Thus, for example, the reason for the covariance between education and perceived productivity may be that sex is correlated with productivity and education is correlated with sex. Moreover, these correlations only reflect the sample of individuals who work in the labor market and thus are affected by selection into or out of the labor market.

It is apparent from these estimates that men in both labor markets are more productive than women and that taller individuals (who are more likely to be men) are also more productive. Of greater interest are the estimates of the variance decomposition, given in the top two rows of table 5. In both countries there is evidence of significant uncertainty about productivity differentials on the part of employers, although the magnitude of the effect is quite different. In the case of the Philippines only about 32 percent of the variance in productivity is known by employers, whereas the corresponding figure in India is 88 percent.

The information set of employers appears to be more complete in the Indian context. This may be because fewer employers in the Philippine sample employ

5. The higher estimate of α_i^* in the Philippines than in India is attributable to the fact that the productivity estimate in the Philippines is net of crop and task.

Table 6. *Estimates of Employer Knowledge of Worker Productivity in India and the Philippines*

<i>Component of the variance in worker productivity</i>	<i>India</i>	<i>Philippines</i>
Known by employer, $\sigma^2(\mu^*)$	0.061 (11.933)	0.015 (1.637)
Unknown by employer, $\sigma^2(u)$	0.008 (3.282)	0.032 (2.007)
Ratio of known to total, $\sigma^2(\mu^*)/[\sigma^2(\mu^*)+\sigma^2(u)]$	0.884	0.319

Note: Results are from maximum likelihood estimation. The absolute values of asymptotic *t*-ratios are in parentheses.

Source: Authors' calculations.

the same workers under the two payment regimes compared with employers in India. Thus, in the Philippines farmers have less opportunity to observe workers on the basis of their piece-rate performance, unlike almost all farmers in the Indian villages. The relatively high degree of employer information among the Indian farmers may also indicate that harvest piece-rate productivity is a good predictor of productivity in other activities. Piece-rate productivity estimates in India, however, were constructed by selecting wages from periods when harvesting predominated. Therefore some of the piece-rate wages that are used may actually be time wages. If this is indeed the case, the estimate of the variance in true productivity (μ_i) in the Indian villages will be too low, but the estimated variance of the known component of productivity (μ_i^*) will be unaffected. This will lead to an overestimate of the proportion of productivity that is known.

There are other possible reasons for the difference in the information known by the employers in the Indian and Philippine samples. For example, the mobility of the work force may be different in the two areas. As noted above, the extent of employer ignorance will depend on the exposure of individuals in a particular labor market. Thus, increased mobility will lead to greater unknown variance in productivity. Similarly, the size of the labor market may affect the extent of employer uncertainty; if villages are relatively isolated and individuals are unlikely to move from village to village to find employment, information on even casual workers should be relatively complete. A thorough investigation into the causes of differences in employer knowledge is beyond the scope of this essay, but it is clearly a topic of importance for identifying remedies for information problems.

Is There Adverse Selection?

Having decomposed worker productivity into the components corresponding to those known and unknown by employers, we can estimate the extent, if any, of adverse selection in the market. We estimate the covariance of the share of labor force time spent in piece-rate work (*PRS*) with the component of productivity that is *not* known by the employer.⁶ The test of adverse selection can be carried out using the Philippine data because in the Philippines both types of

6. Specifically, $\sigma(w_{pi}, PRS) = \sigma(\mu, PRS)$ and $\sigma(w_{ii}, PRS) = \alpha_i^* \sigma(\mu^*, PRS)$, and thus $\sigma(u, PRS) = \sigma(w_{pi}, PRS) - \sigma(w_{ii}, PRS)/\alpha_i^*$.

payment methods exist contemporaneously. We cannot test for adverse selection using the Indian data because, as noted, in the Indian sample villages piece-rate wage activities and time wage activities do not occur at the same time of the year.

The estimated covariation between the difference between actual productivity and productivity as perceived by employers, μ , and the share of labor force time spent in piece-rate work, PRS, in the Philippine data is 0.021. The associated asymptotic t -ratio, 1.75, is significant at the 0.08 level, suggesting that there is adverse selection in the labor market in the Philippines. This estimate and the variance estimate of μ (0.032) in table 6 suggest that a 10 percent increase in the unobserved component of productivity in the Philippines leads to a 6.6 percentage point increase in the share of market work time that a worker devotes to piece-rate activities ($([\sigma(\text{PRS}_{\mu_i})]/[\sigma^2(\mu_i)]) 10 = 6.6$).

Do Employers Discriminate against Women and by Caste?

The Indian villages and the Philippine survey area both exhibit patterns of time wages by sex that are similar to those in many areas of the world. Table 7 reports the average wage rates of men and women in the two surveys, classified by method of payment. In the Indian sample, male daily time wages are on average 62 percent higher than those of females; in the Philippine sample, male time wages are 24 percent higher than female time wages. These differentials by sex in time wages, however, are not very different from those in piece-rate wages in the two samples. The differential in piece-rate wages is 87 percent in favor of males in the Indian villages and 20 percent in favor of males in the Philippine survey area. Strong differentials in work activities across men and women are also evident in the two areas. For example, according to Foster and Rosenzweig (1991), 26 percent of the days men spend in the wage labor market in the Indian

Table 7. Average Time- and Piece-Rate Wage Rates by Sex and Caste in the Indian Survey and by Sex in the Philippine Survey

Item	India		Philippines	
	Time wage	Piece-rate wage	Time wage	Piece-rate wage
Men	7.87 (1.91)	7.17 (1.84)	23.3 (8.75)	29.7 (13.3)
Women	4.85 (2.35)	3.83 (0.82)	18.8 (5.61)	24.8 (12.1)
Low-caste men	8.16 (1.74)	6.30 (2.42)	n.a.	n.a.
Low-caste women	3.60 (0.61)	3.91 (0.71)	n.a.	n.a.

n.a. Not applicable.

Note: For the Indian data, wage rates are in rupees per day and the piece-rate wage is based on the harvest wage. For the Philippine data, wage rates are in pesos per day from the fourth round. Data for both countries are for 1984-85. Standard deviations are in parentheses.

Source: Authors' calculations.

villages are devoted to plowing, the highest-paid activity (24 percent for men in the Philippines), whereas only 1 percent of the Indian women's labor market days are spent in plowing (3 percent in the Philippines). The wage patterns displayed in table 7 are less clear for differentials by caste in the Indian villages. The time wages of low-caste women are evidently lower than those of other women, but the time wages of low-caste men are higher than those of all men.

How much, if any, of the time wage differentials by sex and caste exhibited in table 7 reflect *individual* productivity differentials, employer's use of proxies for individual productivity, or employer prejudice? To identify these factors, we use the information on piece-rate and time wages to carry out a test for the presence of *statistical* discrimination. This involves the equivalent of a regression of the time wage on a series of employer-observed characteristics and on actual productivity (μ_i). We use two observations for the same worker on his or her piece-rate wages (to eliminate measurement error). These estimates are presented in table 8, where we have used variables that can be observed by the employer and are available in the data. These include the worker's age, schooling, and sex for both samples; height for the Philippine sample; and caste and village for the Indian sample.

In the equations determining the existence of statistical discrimination, the presence of a positive coefficient for a particular characteristic implies that individuals with that characteristic will receive a higher wage than equally productive individuals without that characteristic. As noted above, the difference could arise because of the presence of either taste discrimination or statistical

Table 8. *Estimates of Statistical Discrimination in India and the Philippines*

Item	India	Philippines
Age	0.579 (3.387)	0.424 (1.552)
Age squared	-0.474 (3.056)	—
Education	-0.005 (0.393)	0.006 (0.673)
Male	0.110 (2.910)	0.150 (2.844)
Height	—	0.256 (0.710)
Shirapur village	-0.046 (1.743)	n.a.
Aurepalle village	0.014 (0.537)	n.a.
Low caste	-0.072 (3.986)	n.a.
Actual worker productivity, μ_i	0.690 (9.879)	0.761 (2.281)

— Not available.

n.a. Not applicable.

Note: Results are from maximum likelihood estimation. The absolute values of asymptotic *t*-ratios are in parentheses.

Source: Authors' calculations.

discrimination. The most striking feature of table 8 is that, among workers with the same productivity, men are paid a higher wage than women. This differential is significant and of similar magnitude in the Indian and Philippine villages. Thus, there is either taste or statistical discrimination in favor of males. Our findings of significant employer ignorance (table 6) and a covariance between perceived productivity and sex (table 5) for both populations suggests that statistical discrimination may play an important role. The estimates also indicate that among workers with the same productivity, low-caste workers in India receive a lower wage. Again, because caste was observed to be correlated with productivity, this result may reflect statistical discrimination, although taste discrimination may also be involved.

Table 9 provides a clear resolution of these issues. In contrast to the equation for which estimates are reported in table 8, the estimates reported in table 9 control for *perceived* worker productivity rather than actual productivity. Sex does not affect the time wage net of perceived productivity in either sample. Thus we may conclude that the significant effects of sex observed in table 8, and some of the wage differentials by sex exhibited in table 7, are the result of statistical discrimination but not taste discrimination. A rather different picture emerges with regard to the effects of caste. Our estimates suggest that among workers whose productivity is perceived by employers to be the same, low-caste individuals are paid a wage 6.8 percent lower than the wage paid to workers of higher castes. Thus, at least some component of the caste differential in table 8 appears to arise from taste discrimination—a result not very apparent in the gross time wage differentials by caste in table 7.

The presence of a government employment program in Shirapur provides an opportunity to examine the hypothesis that taste discrimination, which evi-

Table 9. *Estimates of Taste Discrimination in India and the Philippines*

Item	India		Philippines
	With control for male	With dummy variable for Shirapur interacted with caste	
Male	0.030 (0.691)	—	-0.143 (0.830)
Low caste	-0.068 (3.655)	-0.083 (4.131)	—
Low caste × Shirapur	—	0.064 (1.384)	n.a.
Worker productivity as perceived by employers, μ^+	0.878 (8.956)	0.946 (18.545)	2.77 (2.13)

— Not available.

n.a. Not applicable.

Note: Results are from maximum likelihood estimation. The absolute values of asymptotic *t*-ratios are in parentheses.

Source: Authors' calculations.

dently exists in the study villages, can be reduced or eliminated through the provision of alternative employment. In table 9 the second set of estimates from the Indian sample is from a specification that includes a dummy variable for Shirapur interacted with caste. Although the resulting interactive coefficient is significant only at the 0.15 level, the point estimate suggests that there is substantially less taste discrimination in Shirapur. Indeed, while one can soundly reject the hypothesis that there is no taste discrimination in the other villages, one cannot reject this hypothesis in the case of Shirapur. Of course, there may be relevant characteristics other than the presence of the government program that lead to this difference in wages for low-caste individuals across the three villages, but exploring this issue requires data with more environments. Clearly, however, the causes of differentials in the extent to which taste discrimination affects wages are important topics for investigation.

IV. CONCLUSIONS

It has long been argued that employers' inability to completely discern workers' productivity plays an important role in determining the structure of labor markets in rural areas of developing countries. Little evidence, however, has been advanced concerning the importance of this problem. In this paper we have used survey data from the Philippines, India, and Pakistan to show that information asymmetries in labor markets in developing countries may be substantial. These asymmetries are an important influence on the distribution of wages across workers and over time for individual workers.

Information problems interact in a number of important ways with the process of economic development. One of the main features distinguishing labor markets in developing and industrial countries is the extent of permanency of labor contracts. The difference can be attributed to a number of factors, including the relative importance of firm-specific human capital, as well as the highly seasonal nature of labor demand in agriculture. This difference in labor contracts has implications for both efficiency and equity because employers with longer ties to their workers are better able to distinguish differences in workers' abilities. The equity implications arise because differentials can be rewarded only when employers know or can measure productivity differentials. The efficiency implications arise because an employer with better knowledge of workers' abilities will do a better job of allocating workers to tasks.

Informational asymmetries also have important implications for worker mobility. Unless there is an efficient mechanism for transmitting information about worker productivity from one village to another, very little will be known about the productivity differentials among recent immigrants into an area. As a result, the problem of adverse selection will reduce both the hiring of recent immigrants in the time-wage labor market and the average quality of those who do migrate. Thus, information problems impose a kind of tax on mobility, particularly of high-productivity individuals, and regional differences in wages may persist longer than might be the case if differences in individual productivity

could be readily assessed. A further implication is that family connections may be important in facilitating migration. As the process of development proceeds, informational barriers to mobility are likely to lessen because of regional integration of firms (and thus information) as well as the increased importance of readily observed signals of productivity such as education, regardless of whether education affects productivity.

Information (and therefore wage inequality) increases with age because employers learn about worker productivity as workers gain exposure in the labor market. This finding has implications for the distribution of income. In particular, as population growth slows and the population ages, the inequality of wages will increase even if there is no systematic relationship between average wages and age. As the economy develops, increases in the returns to experience may eventually swamp age-related information effects. Nevertheless, these effects may be important in the early stages of development.

Informational asymmetries can also reduce the private returns to human capital investment and so attenuate economic growth. From an equity perspective, the extent of the wedge between the private and social returns for an individual will depend in part on the magnitude of informational asymmetry among individuals with similar observed characteristics. For example, if, as a result of lower attachment to the labor force, productivity differentials among women are less likely to be rewarded than those among men, women will be especially likely to underinvest in human capital. This, in turn, leads to greater male-female wage differentials. The magnitude of the informational effect on investment is likely to diminish with economic development, as human capital investment takes a form that is readily measured (such as educational attainment).

The results in this paper also suggest that data collection efforts should include information on the form of payment. Because differences in productivity can easily be compensated when wages are paid on a piece-rate basis, information problems have less impact on the piece-rate labor market. As a result, inference that ignores the distinction between forms of payment may yield misleading conclusions, as we showed in the case of a regression of wages on calories. Moreover, having information on both piece-rate and time wages can be extremely useful for measuring the extent of employer ignorance and how it changes. Another implication of this research for data collection is the need to obtain data on the relationship between workers and particular employers; our findings imply that job tenure decreases information inefficiencies and thus has a distinct and important effect on wage levels and wage inequality.

Development policy should also be sensitive to the implications of informational asymmetries in labor markets. There are a number of issues that we believe to be of particular relevance. First, because the ability to monitor work output will in part depend on the production technology used in a particular setting, employers' inability to utilize piece-rate payment methods under certain technologies may be a significant barrier to technological change. Under these

circumstances it would seem relevant to consider the implications of monitoring costs for the design of agricultural extension projects.

Second, in addition to reducing the extent to which employer prejudices will lead to lower wages for certain groups, our results suggest that employment programs can influence the structure of information in the economy. Because employer ignorance about worker productivity in a particular group depends in part on the labor market experience of that group, certain groups may be systematically excluded from the time-wage labor market because of information problems. Thus, employment schemes targeting women or others with low attachment to the labor force may lead to efficiency increases if this employment provides employers with better information about the group. Public employment programs, however, may have a perverse effect on the wages of disadvantaged groups if they lead to increased ignorance among private employers by reducing the exposure of the individuals to private employers.

Third, one solution to information problems in labor markets is to increase the permanency of employment. The seasonal nature of agricultural employment is a significant barrier to permanency because the employer may not have sufficient work for permanent workers during the off-peak periods. The development of activities with different seasonal patterns of labor demand may lead to an increase in permanency and thus reduce the extent of informational asymmetries.

Fourth, the problems of employer ignorance are reduced to the extent that individuals work for themselves rather than for employers. Increased small-scale credit, by increasing the ability of low-income individuals to be self-employed both in the agricultural and in the nonagricultural sector, may reduce some of the efficiency costs associated with employer ignorance. Increasing the availability of land to poor workers either through land reform or through a shift toward fixed rental or share contracts will have the same effect.

Finally, because informational asymmetries can lead to underinvestment in human capital, information problems provide a justification for subsidizing human capital. The efficacy of educational attainment as a mechanism for providing information to potential employers will depend on the development of consistent standards of education across schools.

APPENDIX. DECOMPOSITION OF THE VARIANCE IN WORKER PRODUCTIVITY INTO ITS KNOWN AND UNKNOWN COMPONENTS

In this appendix we establish the information needed to measure the extent of employer uncertainty about worker productivity and to test for the presence of adverse selection in the labor market. The objective is to obtain an estimate of the variance in worker productivity, $\sigma^2(\mu)$, and then to decompose this variance into its known component, $\sigma^2(\mu^*)$, and its unknown component, $\sigma^2(\mu)$.

To estimate the component of the variance in worker productivity that is unknown by the employer, we assume that the piece-rate wage at any particular

point is equal to actual productivity plus an error term that reflects measurement error and any idiosyncratic component of the piece-rate wage:

$$(A.1) \quad w_{pi} = \mu_i + \epsilon_{pi}$$

where subscript p indicates piece-rate wages. Assuming that the error terms ϵ_{pi} are uncorrelated, using two piece rates for each individual, w_{p1} and w_{p2} , $\sigma^2(\mu) = \text{cov}(w_{p1}, w_{p2})$. If piece-rate data include information on different tasks, as in the case of the Philippines, task controls may be estimated using a fixed-effects procedure, and a noisy productivity estimate, net of task, may then be constructed. The rest of the procedure remains unchanged.

To estimate the component of productivity that is known by the employer, we compute the covariance of one piece rate and one time wage, w_{t1} , for each individual. Using equations 2, 4, and 7 (assuming there is no taste discrimination in equation 4), along with the fact that u_i is uncorrelated with μ^* , and assuming that the measurement error in the piece-rate wage is uncorrelated with measurement error in the time wage, we obtain the following: $\alpha_\mu^* \sigma^2(\mu^*) = \text{cov}(w_{t1}, w_{pi})$. Thus, if an estimate of α_μ^* , the effect on time wages of a unit increase in perceived productivity, can be obtained, it will be possible to solve for the known component, $\sigma^2(\mu^*)$ and, since $\sigma^2(\mu) = \sigma^2(\mu^*) + \sigma^2(u)$, it will be possible to solve for the unknown component, $\sigma^2(u)$, as well.

The principal problem that arises in estimating α^* is that, as noted above, a regression of the time wage on actual productivity, μ , will not yield a consistent estimate because μ is a noisy measure of the component of productivity that actually affects the wage, μ^* . The implication is that if one can identify variables that are correlated with μ^* but not with $u = \mu - \mu^*$ or time wages net of predicted productivity, instrumental variables may be used to obtain consistent estimates of the parameters in equation 4.

Instruments are readily available as long as the researcher has access to worker characteristics that are correlated with productivity, that are known by the employer, and that can be argued on a priori grounds not to be subject to taste discrimination. As argued above, an optimizing employer will form expectations of productivity on the basis of any worker characteristics that he or she knows. Thus, any characteristics that are known by the employer must be uncorrelated with the unknown component of productivity, u . As long as these characteristics are correlated with productivity and are not subject to taste discrimination, they will be correlated with expected productivity but not with time wages net of expected productivity.

Instead of computing a series of separate covariances, as suggested by the above discussion, we use an alternative approach to estimation that involves fitting the theoretical covariance matrix implied by the model and parameters to the data covariance matrix. There are several advantages to this approach. First, simultaneous estimation of the different components of the model leads to an increase in efficiency. Second, if errors are distributed as multivariate normals, the approach is maximum likelihood and is therefore fully efficient. Third,

standard errors can be computed for variance estimates, making it possible to test hypotheses about the structure of information. These and other standard errors can be corrected so that they are robust to misspecification of the distribution generating the data.

REFERENCES

- Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in the Labor Market." *Industrial and Labor Relations Review* 30:175-87.
- Basta, Samir S., M. S. Soekirman, Darwin Karyadi, and Nevin S. Scrimshaw. 1979. "Iron Deficiency Anemia and the Productivity of Adult Males in Indonesia." *American Journal of Clinical Nutrition* 35:127-34.
- Bouis, Howarth E., and Lawrence J. Haddad. 1990. *Agricultural Commercialization, Nutrition, and the Rural Poor: A Study of Philippine Farm Households*. Boulder, Colo.: Lynne Rienner.
- Deolalikar, Anil B. 1988. "Nutrition and Labor Productivity in Agriculture: Estimates for Rural South India." *Review of Economics and Statistics* 70(3):406-13.
- Foster, Andrew D., and Mark R. Rosenzweig. 1991. "Unequal Pay for Unequal Work: Asymmetric Information, Sex Discrimination, and the Efficiency of Casual Labor Markets." University of Philadelphia, Department of Economics, Philadelphia, Pa.
- Glaeser, Edward L. 1992. "Who Knows What about Whom: An Empirical Analysis of Heterogeneous Information in the Labor Market." Harvard University, Department of Economics, Cambridge, Mass.
- Lundberg, Shelly J., and Richard Starz. 1983. "Private Discrimination and Social Intervention in Competitive Labor Markets." *American Economic Review* 73:340-47.
- Mincer, Jacob. 1974. *Schooling Experience and Earnings*. New York: National Bureau of Economic Research.
- Pitt, Mark M., Mark R. Rosenzweig, and M. N. Hassan. 1990. "Productivity, Health and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries." *American Economic Review* 80(5):1139-56.
- Rosenzweig, Mark R. 1988. "Risk, Implicit Contracts, and the Family in Rural Areas of Low Income Countries." *Economic Journal* 98:1148-70.
- Rosenzweig, Mark R., and Oded Stark. 1989. "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India." *Journal of Political Economy* 97:905-26.
- Roumasset, James, and Marilou Uy. 1980. "Piece Rates, Time Rates, and Teams: Explaining Patterns in the Employment Relation." *Journal of Economic Behavior and Organization* 1:343-60.
- Strauss, John. 1986. "Does Better Nutrition Raise Farm Productivity." *Journal of Political Economy* 94(2):297-320.
- Walker, Thomas S., and James G. Ryan. 1990. *Village and Household Economies in India's Semi-Arid Tropics*. Baltimore, Md.: Johns Hopkins University Press.
- Weiss, Andrew. 1980. "Job Queues and Layoffs in Labor Markets with Flexible Wages." *Journal of Political Economy* 88(3):526-38.

COMMENT ON "INFORMATION FLOWS AND DISCRIMINATION IN LABOR
MARKETS," BY FOSTER AND ROSENZWEIG

Nancy Birdsall

This carefully prepared paper makes imaginative use of the limited information available to employers about their employees. (I'm tempted to say that if the employers were as clever as the researchers in sorting out the characteristics of the employees, there would be no information problem in these labor markets.)

Yet I have a sense of dissatisfaction about this paper. The authors do well what they set out to do; they assess and dissect the effects of poor information in labor markets in low-productivity, rural economies. But they do not broaden our understanding of why economies are underdeveloped, nor of what could be done to spur the development process. The paper brings the full panoply of sophisticated economic thinking and careful econometrics to a question that may not be the correct one—a bit like hitting a small nail with a sledgehammer. Is this a fruitful area of research, either for improving our fundamental understanding of development or for guiding development policy? I do not think so. This may reflect a failure of my own imagination, and I'll return to that point later. But first let me illustrate why I do not think so.

Findings in the paper indicate that at least two problems can arise because of employers' poor information about their employees: job immobility and statistical discrimination. These two problems impinge on the smooth and efficient operation of the labor market in poor rural areas. But note that the efficiency costs which these problems create are largely static—they are inefficiencies that make low-productivity agriculture marginally lower than it might otherwise be. Are these static costs the fundamental development problem or question? As with the small Harberger triangles that Richard Freeman has mentioned, the static costs studied in this paper under a high-powered microscope amount to asking, why is per capita income in rural India \$100 a year, when with a one-time gain from eliminating static inefficiency it might rise to \$102 or \$103 a year?

In contrast, we could address questions of efficiency costs in a more dynamic sense. Why is productivity growth in rural India or in the Philippines 1 or 2 percent a year instead of the 6 or 10 percent recorded in East Asian countries

Nancy Birdsall is director of the Country Economics Department of the World Bank.

© The International Bank for Reconstruction and Development / THE WORLD BANK

when they were transforming themselves into societies with much higher incomes?

Take the example of job immobility. The authors show that there will be a tax on migration for the most productive workers. These workers are reluctant to move around because when they do so, they separate themselves from employers who have better information about their characteristics and who thus are willing to pay them higher wages (when they are paid on a time-wage basis).

This seems to me to be a classic example of a static efficiency loss. If all the workers in rural India could be put exactly where they ought to be by a one-time reallocation, thus correcting in one fell swoop the job immobility that arises from the information problem, there would be a one-time efficiency gain. But surely for development in the larger sense, this one-time efficiency gain would be minimal. Such a one-time gain through reallocation might be important in Eastern Europe; if we were able to move people around and get them reallocated optimally to the right industries in the right regions, the one-time increase in productivity would be substantial and would set the stage for subsequent dynamic growth. But it does not strike me as central in the low-productivity, rural labor markets of the poorest parts of the world.

The gains that could be realized in India or the rural Philippines from that one-time reallocation would surely be less than the gains that could be achieved by giving every worker in these parts of the world an additional year of education or by ensuring—if we could—the correct package of macroeconomic and structural reforms to get these economies moving and growing.

What if job immobility is itself a cause of underinvestment in human capital? This does seem more costly in a dynamic sense because it could be self-reinforcing from one generation to the next. But, as the authors are careful to point out, the differences in human capital among the workers are not associated with formal schooling. This is part of the problem; these are differences in human capital that the employer cannot observe. Once levels of school achievement increase in a population, it turns out that there is usually a correlation between unobservable human capital (for instance, innate ability) and schooling, and the problem of unobservable (to the employer) human capital is eclipsed. A straightforward approach to the employers' information problem, and thus to any costs of labor immobility, might be simply to increase the supply of schooling. More education would resolve much of the problem of concern to the authors—that is, the employers' lack of information about their employees.

Consider another example: statistical discrimination. The authors propose a nice method for distinguishing taste discrimination from statistical discrimination. For women in the sample for rural India, they find no evidence of taste discrimination. They do find, however, that there is statistical discrimination and that it has a substantial effect on the earnings of women in that sample.

The negative effect of this statistical discrimination on the logarithmic wage of women, as shown by the coefficients, is similar in absolute terms to the positive

effect we expect from an additional year of schooling, as seen in a standard regression of earnings on schooling. In short, statistical discrimination in rural India reduces a woman's wage about as much as one extra year of schooling increases the wage in most developing countries. On the one hand, this is a large effect, especially in societies where average schooling is as low as two to three years. On the other hand, the story of development over the past few decades has been one of rapid expansion of schooling, so that now we see in most developing countries virtually 100 percent enrollment in primary school. These may not be very good primary schools, and not all students complete primary school. But compare the effect of statistical discrimination in terms of a static efficiency loss to the static and potential dynamic gains a society reaps by educating girls (see Summers 1992). The existence of statistical discrimination reveals welfare losses for women as a group. But its elimination in rural India would not in itself matter much for development—not the way that eliminating lack of schooling would.

It is worth considering the dynamic costs that may occur as a result of statistical discrimination. For example, such discrimination may reduce the willingness of parents to invest in girls' education, so that from one generation to the next there is underinvestment in girls' education (see Birdsall and Sabot 1991). But again, the real issue and point of leverage is to increase access to girls' education.

In their concluding section, the authors do point out some policy implications of their research and touch indirectly on dynamic costs. They note that programs such as land reform and better access to credit would increase the number of small landowners; this would reduce the efficiency costs associated with labor immobility, since small landowners do not have the monitoring and information problem of large landholding employers. These are steps that might bring larger and more dynamic gains. Similarly, they note that agricultural technology that minimizes the costs of monitoring workers (given the information problem) should be developed and encouraged in rural areas and that certain approaches to targeting public programs could help to reduce the statistical discrimination problem.

In the end, I remain skeptical about the value of this work for understanding critical development questions. But my experience with the work of Mark Rosenzweig and his various coauthors over time is that these microanalytic studies, because of their emphasis on understanding human behavior in response to the economic environment, insinuate themselves into our thinking in important ways—and eventually do turn out to have implications for the way in which we address development questions.¹ So my dissatisfaction with this paper may ultimately reflect my own failure of imagination.

1. Similar microeconomic analysis has greatly increased our understanding of the links between development and family choice in such areas as education, investment in girls versus boys, fertility, and the links among factor markets in poor rural economies. See, for example, the review articles and citations in Rosenzweig (1988), Behrman and Deolalikar (1988), Birdsall (1988), and Schultz (1988).

REFERENCES

- Behrman, Jere R., and Anil B. Deolalikar. 1988. "Health and Nutrition." In Hollis Chenery and T. N. Srinivasan, eds., *The Handbook of Development Economics*. New York: North-Holland.
- Birdsall, Nancy. 1988. "Economic Approaches to Population Growth." In Hollis Chenery and T. N. Srinivasan, eds., *The Handbook of Development Economics*. New York: North-Holland.
- Birdsall, Nancy, and Richard H. Sabot, eds. 1991. *Unfair Advantage: Labor Market Discrimination in Developing Countries*. World Bank Regional and Sectoral Studies. Washington, D.C.
- Rosenzweig, Mark R. 1988. "Labor Markets in Low-Income Countries." In Hollis Chenery and T. N. Srinivasan, eds., *The Handbook of Development Economics*. New York: North-Holland.
- Schultz, T. Paul. 1988. "Education Investments and Returns." In Hollis Chenery and T. N. Srinivasan, eds., *The Handbook of Development Economics*. New York: North-Holland.
- Summers, Lawrence. 1992. "Investing in All the People." Paper prepared for the Quad-i-Azam lecture at the 8th annual general meeting of the Pakistan Society of Economists, Islamabad, January. World Bank, Office of the Vice President, Development Economics, Washington, D.C.

FLOOR DISCUSSION OF THE SVEJNAR AND FOSTER-ROSENZWEIG PAPERS

In response to discussant Marek Góra's comments about the legacy of poor motivation to work under the command system, Svejnar said he thought Western economists probably underestimated that factor because they think people will always work harder given reasonable incentives. He said that he had been struck by the inertia in Central Eastern Europe, which he saw now was reflected in the slow emergence of markets. One positive aspect of the situation was that unemployment was not linked to poverty so much as to waiting things out. In the parts of Czechoslovakia that had the highest unemployment rates, for example, workers were being imported from the Ukraine because wages in the booming construction sector were so low that local workers didn't want to do the work. So the unemployment figures didn't necessarily mean hardship and privation, although that might come. In response to Góra's question about if, and how, people could be motivated to work harder, Svejnar wondered how it could be done in the context of state ownership, given that privatization was proceeding very slowly.

A participant from the World Bank said that he was struck by the apparent conflict between Svejnar's discussion and what Larry Summers had said earlier that day. Svejnar had said that worker-controlled enterprises did fine if they operated in a competitive environment and that Eastern Europe's experience under worker control was satisfactory. But Summers had said that things got worse if you went from central control to markets without incentives—a position Svejnar found too simplistic. Manuel Hinds of the World Bank, said the participant, had suggested in his research that workers given control of an enterprise with \$10 million in capital would be tempted to simply decapitalize as quickly as possible, which would soon leave the country with little or no capital stock. What one saw in countries with worker control was centralized regulation to prevent that kind of abuse, so it wasn't really pure worker control. One also saw pressures for subsidies or loans at negative interest rates. Codetermined solutions were useful, said the participant, but only under privatization, when the workers were part owners, so they didn't have an incentive to decapitalize. The Bank had a proposal on the table, he continued, that would privatize enterprises in the Russian Federation and give workers roughly 25 percent ownership. He asked if Svejnar thought that was an appropriate balance.

This session was chaired by Gershon Feder, chief, Agricultural Policies Division, Department of Agriculture and Rural Development, the World Bank.

Svejnar responded that managers of state-owned enterprises told him that their hands were tied because wage levels were regulated. Was it possible to control state enterprises while at the same time providing strong incentives? The wage norm was imposed out of concern that managers in these firms would decapitalize and siphon off wealth, ultimately destroying the enterprise. If the decision was that distorting an important market was worthwhile for macro-economic stabilization and long-term gain, why not impose such a constraint directly in a way that wouldn't distort the labor market? Why not impose binding taxes and force the owner (in this case, the state) to reinvest and achieve a certain rate of return? In a way, said Svejnar, this would be a commercialization of state-owned enterprises, which the Bank has looked at in various countries.

As to what role worker participation in profits could play, in a recent article in *Comparative Economics* Svejnar and a coauthor had examined the Yugoslav experience and had found that outcomes had been greatly affected by social ownership and government intervention (for example, banks being owned by enterprises), two factors that were not necessarily features of worker participation. He thought that if you isolated the issue of worker participation—whether in management, ownership, or profits—you would not see negative effects.

The codetermination example was a good one, said Svejnar. The question always has to be, how do you impose the constraints that are needed? As for the specific question about Russia, evidence about the benefits of worker ownership was very weak. If one were to ask whether it was better to go for profit-sharing, worker ownership, or worker participation in management or decisions, the evidence on ownership was in some sense nonnegative, but it was not very strong either way. Svejnar said he would use different criteria, worrying about such things as how to boost productivity.

An economist from the World Bank who had studied the worker participation issue said there really wasn't a contradiction because in Yugoslavia it had only appeared that workers were the owners and decisionmakers. Political elites were the real owners, so it was no wonder that workers would opt for money in their pockets if they were given a chance. He had observed another interesting feature of worker participation in Eastern Europe and the former U.S.S.R.: worker participation did not preclude shedding labor. He had visited a steel mill where the workers themselves, through the workers' council, had laid off 2,000 workers, about a third of the total. He asked for a more thorough evaluation of what the Bank calls proactive labor market policies—for example, training and public employment programs; wage subsidies; job creation or unemployment reduction measures (such as direct government spending or investment tax credits); and public employment programs.

Svejnar said that there were many examples like that of the steel mill just mentioned. German workers had okayed Volkswagen's moving its plant to the United States, for example, after the long-term benefits were explained to them. Svejnar also agreed that more attention should be paid to proactive labor poli-

cies. Eastern Europe should be focusing on income maintenance; they were giving these policies less attention by default. He had always argued that the Czechoslovak government should ask workers to build highways and similar projects, but things were at a crossroads in these areas, and you could not get workers to work even though unemployment was rising fast.

Gershon Feder (chair) asked if any studies had distinguished between the two effects of worker participation in nonagricultural enterprises: the effect on management decisions (such as what the wage rate should be and what investments should be made, and for how much) and the effect on worker effort. Did Manuel Hinds' work criticize worker participation on one count or both?

Svejnar said he thought worker participation had been criticized on both counts but especially the first. The answer depended on how sophisticated a model you used. Very simple models say that the larger the group, the more you have a free-rider problem; in the cooperative model, workers self-monitor. The outcome varies with the approach. The best empirical studies examine enterprise-level panel data, looking at the performance of firms with different degrees of worker participation, ownership, decisionmaking, and profit. The weakness of the data is the amount of variation; firms either have worker participation or don't have it, but ownership may range from 10 to 60 percent. Svejnar said they hoped in the ongoing World Bank research project to get before-and-after data on enterprises in Central and Eastern Europe so that they would effectively have a laboratory experiment.

Nancy Birdsall's (discussant) criticism of Rosenzweig's paper elicited a strong reaction. A participant from the World Bank felt the implications of the paper were serious. Rosenzweig had shown information problems to exist in a labor market in which workers and employers lived in the same village, knew a lot about each other, and were evaluating rather simple tasks, such as how much cotton you can pluck in an hour. If you found both taste and statistical discrimination there, consider how much more serious they would be in complex situations where employers didn't live near or know workers and where jobs were more difficult to evaluate.

A participant from the University of Minnesota agreed that the paper raised important issues for development—not so much for the efficiency loss as for equity of access to employment. The problem was probably more serious in urban labor markets employing casual craft labor, such as the construction industry. He thought the paper provided a theoretical basis for segmenting labor markets by such criteria as gender, ethnicity, and kinship. The paper showed that poor people with a gender or kinship connection associated with low productivity were shut out of labor markets. Alleviating the information problem might help alleviate poverty.

Rosenzweig came to Birdsall's defense. He said he'd been asked to do a paper on discrimination and his response had been that he hadn't thought it was an important issue for development for some of the same reasons Birdsall had given. He had thought it might be interesting, though, to look at discrimination

in the broader context of information problems, since they were able to compare data on people who had worked for both piece rates and time wages in the same area. He agreed that the implications were even worse for more complex work, where skill differentials were more subtle. Choosing the wrong economist or the wrong corporation president would have far broader ramifications than choosing who was going to do the weeding.

On the one hand, the cost of information-related problems in casual rural labor markets must not be high, or employers would do something about it. Rosenzweig saw a certain irony there. It had cost economists hundreds of thousands of dollars to interview rural workers to understand the data on piece-rate performance and time-pay performance, and in principle the farmers could do the same thing. But they don't do it, for a good reason: despite large discrepancies between true productivity differences among workers and what employers know about them, the consequences of their misallocations are not that serious.

On the other hand, in thinking about economic transitions, one should anticipate the emergence of institutions and problems. It is important, for example, to understand the informational implications of reduced population growth and an aging population. Wage inequality will increase; the question is, whether to do something about it or not. Inequity of distribution is exacerbated by information problems. Moreover, as skills become more subtle, more important, and more differentiated, different payment forms may emerge, and employers may have to do things to cope with these problems, through tenure, tests, or other instruments. A lot of research in labor economics is focusing on what the optimal forms of compensation are, and information problems are part of what one looks at to explain different payment methods. We need to think more, said Rosenzweig, about the consequences of information problems—about whether they will become more important to development, whether they will hinder it, and what we should expect to see as labor markets alter.

Birdsall said that as she had predicted, the discussion had revealed her own failure of imagination about information problems. The challenge, she said, was to explore dynamic efficiency losses and their unpleasant connection with the equity problem. Statistical discrimination deserves more thought because it discourages the next round of investment in human capital for the groups experiencing discrimination—including women and people of lower castes—leading to another round of perhaps more intensified discrimination, so that we could end up with a serious self-reinforcing situation.

The paper had tentatively suggested that employment programs influence the structure of information programs in world labor markets, so one speaker thought there was something to be said for pushing employment programs and examining the implications and effects of individual programs in specific settings.

A Bank participant said he had always been puzzled by the term "statistical discrimination," which had never struck him quite as discrimination. If two groups differ in average productivity, they should be paid different wages. The problem would come if discrimination affected workers' motives to improve themselves,

which was a much more subtle thing. Rosenzweig said that while it was fine if employers used all the information they had to reward employees, problems arose when employers couldn't perceive productivity differences among workers and where there was a gap between true productivity and an employers' assumption about productivity for workers with less exposure in the labor market—women, especially, who were therefore discriminated against as a group in the market.

To clarify the problem of statistical discrimination, another participant gave the example of a university that was hiring professors. If you were at such a university, he asked, would you be more likely to hire a professor who you knew was a solid scholar, or a professor who had a 50 percent chance of being Einstein and a 50 percent chance of being an idiot? It depended, of course, on whether you had an option to fire people. If you could, then you might hire the possible genius. If he turned out to be a genius and could prove it to everybody, he might stay, but he might also move on. If you could keep private the information that he was terrific, he would be more likely to stay. Using that argument, it seemed wrong to argue that uncertainty about productivity would lead to discrimination. Rosenzweig agreed with the speaker in that context but said that in casual labor markets you didn't get a chance to find out, and information was kept private.

Nancy Birdsall posed this situation: if one had a son and daughter who were equally able, if there were a positive interaction between ability and returns to investment on schooling, and if there were statistical discrimination against women (if employers perceived productivity X for all women and productivity X plus Y for all men), then, given scarce resources, all other things being equal, one would be inclined to invest in more human capital for the son than for the daughter. Another participant said that because of such discrimination, the opportunity cost would be lower for the daughter than for the son. In the lively discussion that followed, one participant observed that these phenomena are quite subtle.

Rosenzweig tried to clarify the main point he was making by saying that in a particular situation—say, harvesting—you might observe that women were 20 percent less productive than men. In that situation, to pay women less would be rational. But when men and women are observed to be equally productive at harvesting, women are still paid 20 percent less. The nice thing about their study was that it compared the same workers in essentially the same task but under different payment schemes. Nobody had a problem with payment under the piece-rate scheme; inequity set in under the time pay scheme. And that was why Rosenzweig had made the distinction in his paper about “good” (statistical) and “bad” (taste) discrimination.

A participant from the International Labour Organisation (ILO) linked the two papers by reporting that in ILO studies of labor markets in the transitional economies, particular groups were suffering disproportionately—including women, minorities, older workers, and the disabled. Something like 78 percent of the recognized unemployed in Russia were women, he said, and many of the unemployed were highly qualified. He agreed about the need to complement the current analysis with a consideration of different labor market policies.

Knowledge, Property, and the System Dynamics of Technological Change

Paul A. David

Technological change and its relationship to the growth of knowledge are considered here from a general systems-theoretic perspective. The traditional linear model that has influenced economic thinking and policy analysis suggests a unidirectional flow of causation, from exogenous fundamental discoveries in science leading eventually to technological inventions, innovations, and the diffusion of new products and production techniques. Scientific and technological advance should be approached, instead, from a general evolutionary viewpoint, as a phenomenon of "organized complexity" that results in cumulative and irreversible transformations in knowledge and use of economic resources.

This paper examines some of the system effects of various institutional solutions to the so-called appropriability problem affecting the production of information. It points out some of the science-technology interactions that have often been overlooked and discusses the implications of positive and negative feedbacks between the dynamics of innovation and diffusion. It concludes by considering what these may imply for discussions of North-South differences over the policy of strengthening protection for intellectual property rights.

For a long time, most economists' conceptualizations of technological progress have been held in thrall by one or another expression of the linear, reductionist, analytical approach that has dominated the discipline as a whole. There is a lot to be said for trying to understand the workings of a system as complicated as an economy by examining the behaviors of its component parts and trying to characterize the equilibrium states of the various subsystems that can be identified. This is what we do in neoclassical microeconomics and related branches of analysis. Difficulties do tend to arise, however, in grasping the ways in which these subsystems can interact. This is especially true when some of the subsystems are driven far from the neighborhoods of their more readily charac-

Paul A. David is William Robertson Coe Professor of American Economic History in the Department of Economics at Stanford University. This paper was originally drafted while the author was a fellow at the International Center for Economic Research (ICER), Villa Gualino, Turin, Italy. The author extends his gratitude to ICER and its staff for their hospitality and to Cristiano Antonelli, Weston Headley, Paul Romer, and Francisco Sagasti for their helpful comments.

terized equilibrium states. The behavior of the system as a whole frequently diverges from what one is led to expect by simply adding together the effects identified in analyses of its constituent subsystems.¹ Economists are rather belatedly taking a more serious interest in the analysis of nonlinear dynamic systems. Our colleagues in the natural sciences long ago saw that the paradigm offered by classical (Newtonian) mechanics was not the only—or, indeed, the most useful—way of thinking about dynamical processes.

Yet old habits die hard, especially habits of the mind. Consider the persistent influence of what I refer to as the Simplest Linear Model, or SLIM. This elementary and familiar construct depicts technological change and productivity growth as the end results of a unidirectional causal sequence, often graphically represented by a series of boxes, each connected to the next by a single arrow pointing from left to right. The system flow-chart tells us that (1) fundamental science yields discoveries, which lead to (2) experimental findings of applied science, which lead to (3) acts of invention, which provide the stimuli and basis for (4) entrepreneurial acts of innovation (commercial introduction of novel products and production methods), which incite (5) imitation and so bring about (6) diffusion of the new technology into general use. From diffusion will flow changes in productivity and welfare improvements but also quite possibly profound alterations in market structure and untoward effects such as the displacement of workers, the downward valuation of assets rendered economically obsolescent, and the demise of firms that fail to adapt to the competitive pressures unleashed by more efficient methods and better-quality products.

I am by no means the first person to notice that this idealized sequence of events vastly oversimplifies and distorts the workings of the world.² It is widely appreciated that for much of the world's history new technologies had little indebtedness to what we would call "science." Even today, inventions do not necessarily follow from applied scientific discoveries. Furthermore, important developments in applied science—consider recent breakthroughs in high-temperature superconductivity, as one example—often occur well in advance of an understanding of the fundamental processes underlying the phenomena in question. Technological mastery may run far ahead of science and is in many regards both a stimulus to scientific inquiry and the means whereby such inquiries can be conducted.

Most of the criticisms that have been lodged against SLIM focus on three glaring deficiencies. The first problem, which has just been reviewed, is the inadequacy at an epistemological level of the account it provides of the evolution

1. See Rapoport (1986) for a useful discussion of the strengths and limitations of the analytic and holistic, or gestalt, approaches to cognition and an introduction to general system theory. The application of a systems approach to science and technology policymaking has been discussed by others, notably in a thoughtful monograph by Sagasti (1972).

2. See, for example, the detailed critique of the classic linear model by Kline and Rosenberg (1986), which gives special attention to the complicated ways in which product and factor demand conditions and interrelated developments in scientific and technological knowledge impinge on the innovation process.

of the stocks of scientific and technological knowledge. The second is the depiction of science as neatly separated into fundamental and applied compartments, with the activities carried on in the first compartment being exogenous to the economy and therefore appearing in the role of the driver of the entire sequence of activities and events. The determinants of induced invention and the institutional conditions affecting market-oriented investments in research and development (R&D) thus receive no explicit notice. Instead of extending an influence backward into the search for scientific principles that will help guide a profit-motivated research in quest of prespecified new products or production techniques, would-be inventors in the world of SLIM await their cues from the realm of autonomous science.

The third major distortion of reality is that changes in the technological opportunity set available to producers are conceptualized as resulting from discrete advances or research breakthroughs. There is a great deal of evidence, however, that most long-run increases in technical efficiency and declines in the price-performance ratio of products in an industry have been the result of the cumulation of a myriad of small improvements. These incremental modifications are usually based on experience gained in actual production operations and in the repeated interactions between the users and the manufacturers and vendors of complex products. In short, endogenous experience-based learning, which is predicated on having gotten beyond the innovation stage, is an important source of the technical developments that SLIM would ascribe to an anterior stage of invention.

These three lines of criticism hardly exhaust the list of features of SLIM that are sufficiently problematic to deserve notice at the outset. Even more glaringly absent from SLIM is a recognition of the way that past, current, and anticipated future economic conditions determine decisions affecting the allocation of resources in the several activities distinguished by the schema. The interconnections between the generation and the diffusion of innovations need to be articulated in ways that the unidirectional flow of influence will not admit.

Although seemingly innocuous, this way of organizing one's thinking about the process of technological change has had some subtle and pernicious influences which persist despite critiques of the model. Using SLIM, each stage of the dynamic sequence is seen as characterized by a distinctive resource allocation problem that can be separately addressed for purposes of policy analysis. Thus, government subsidies for organized science are designated as the "policy fix" for public goods problems connected with fundamental scientific knowledge. Intellectual property rights policies are seen as the appropriate choice of a policy instrument to stimulate invention and innovation (commercialization), although at the undesirable sacrifice of wider diffusion. Different policy measures are addressed to the separate stages—as indicated by designations such as basic science policy, science-technology transfer policies, innovation (company-financed R&D support) policies, and, in the United States recently, technology diffusion policies—and their effects are implicitly presented as separable and additive. This approach does not

address policies that have side effects or knock-on effects that may show up as perverse consequences elsewhere in the system.

Economic analysis needs to move beyond SLIM to articulate the interdependence and interactions among the subprocesses in the overall system governing the production, distribution, and utilization of scientific and technological knowledge. Scientific and technological advance may more usefully be approached from a general evolutionary viewpoint, as a phenomenon of "organized complexity" that results in cumulative and irreversible long-run change. A systems-theoretic view would emphasize the interchangeability of supposed causes and effects as the context is varied. It would notice the existence of intricate feedback loops, including positive feedbacks that will, if sufficiently dominant, lead to the existence of multiple equilibria and to nonlinearities that may cause catastrophe-like discontinuities in the dynamic responses to policy interventions. A program of this kind would enable us to understand technological change as a channeled historical process that exhibits strong aspects of inertia and continuity but that also is capable of generating abrupt changes in pace and direction; it would enable us to begin to frame policies for differently situated economies accordingly.

That, however, is not a simple program. Nor is it one that can be systematically pursued here. What I shall attempt, therefore, is to resurvey some of the familiar terrain with an eye for features that cannot be readily seen from the SLIM perspective but appear more salient from a complex systems perspective. In section I, I examine some little-noticed interactions, dynamic feedbacks, and nonlinearities between science and technology that affect the generation of knowledge through organized R&D processes. I try to show why it is important to achieve the contextually appropriate mix of public patronage, governmental procurement contracting, and protection of intellectual property rights.

In section II, I identify various feedback mechanisms (positive and negative) that link technological innovation and diffusion. Their implications are shown to include various possibilities that perverse results will arise from the interactions of policy interventions which often are supposed to be separable and additive in their effects. In section III, I reconsider from the systems perspective thus gained the current proposals for global harmonization of intellectual property protection regimes. I conclude that even if it were legitimate to identify the interests of industrial and developing countries with the earlier and later stages, respectively, of the linear model's activity sequence—with invention and innovation in the case of the industrial nations, and with imitation and diffusion in the case of the developing world—that is not sufficient to indicate the areas of conflict and conciliation between them in regard to the protection of intellectual property rights.

I. THE GROWTH OF SCIENTIFIC AND TECHNOLOGICAL KNOWLEDGE

The economics of information and its bearing on the production and distribution of scientific and technological knowledge is a logical starting point for our discussion. To avert later confusion, a few definitions and terminological dis-

tinctions will be introduced before plunging into the substantive issues. A number of terms employed throughout this section have precise meanings that differ somewhat from those assigned them in ordinary language usage.

Information is knowledge reduced to messages that can be transmitted to decision agents. This accords with the standard information-theoretic view in economics that such messages have information content when receipt of them causes some action or alteration in the subjective or objective state of an agent.³ Transformation of knowledge into information is thus a necessary condition for the exchange of knowledge as a commodity. When a knowledgeable agent puts what he or she knows into a form with legally protected use-rights, we may say that the commodity in question is information. This is the case with a proprietary report for a client, a copyrighted publication, or a patent describing an invention.

Information can be disclosed (made available by one agent for transmission to others) without valuable consideration being demanded. In other words, information is not an economic commodity by its very nature because it also may be left as a free good. The transformation of knowledge into information commodities does not suffice, however, to remove the special features of knowledge (and information) that oblige us to distinguish it from the run of ordinary commodities exchanged through conventional markets. Correspondingly, it will be seen that the production and distribution of additions to the stocks of scientific and technological information call for reliance on some special resource allocation mechanisms.

Codified and Tacit Knowledge and the Properties of Information as a Commodity

Knowledge can be gained through accidental discoveries, or it can emerge from systematic, rational inquiry and observation, but in all cases knowledge products are distinctive rather than homogeneous goods. In addition to being highly differentiated, knowledge is characterized by an extreme form of indivisibility: it is sufficient, at least in principle, to acquire a specific piece of knowledge once. There is not much social advantage to repeating the process of its acquisition; although the wheel does seem to have been invented on several occasions, the documented instances occurred in different and isolated societies. But the most important characteristic of knowledge is the possibility of its being possessed and enjoyed jointly—even simultaneously—by different individuals. This property in a commodity may be referred to as *perfect expansibility*.⁴

3. Shannon's well-known measure of the quantity of information was formulated to serve the specific needs of communications theory (see Shannon and Weaver 1949) and has been found unsuitable for application in economics (Arrow 1969; Marschack 1971).

4. Following Thomas Jefferson (Koch and Peden 1972, p. 629), who wrote of ideas being "like fire, expansible over all space, without lessening their density at any point," I suggest *expansible* as an alternative to the term *nonrival*. The latter term has been introduced by Romer (1990 and in this volume)

Knowledge, in being expansible, displays one of two properties commonly associated with the general category of public goods. The other defining attribute of a public good—the *nonexcludability of individuals* from its use—is not necessarily present in the case of knowledge because it remains quite possible to enforce exclusive use. Intellectual property rights are social contrivances that prevent nonowners from using information for certain purposes even though they have gained possession of it.

While these latter observations form the point of departure of the classic analysis of the economics of R&D (Arrow 1962), they hardly can be claimed as modern insights. Consider the following passage in a letter written in 1813 to Isaac McPherson, a Baltimore inventor, by Thomas Jefferson:

If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it. *Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me . . .* Inventions then cannot, in nature, be a subject of property. Society may give an exclusive right to the profits arising from them, as an encouragement to men to pursue ideas which may produce utility, but this may or may not be done, according to the will and convenience of the society, without claim or complaint from anybody. (Koch and Peden 1972, pp. 629–30, emphasis added)

Codification of knowledge is a step in the process of reduction and conversion that renders the transmission, verification, storage, and reproduction of information especially easy. Codified information typically is organized and expressed in a format that is compact and standardized to facilitate and reduce the cost of such operations. The demand among any community for codification of knowledge in a given form will be influenced in part by the recognizability of the encoding conventions employed. The encoding conventions affect the decipherability of the compressed message and the ease of making a translation from an abstract, generic formulation to one that will have specific applicability. The degree of comprehensiveness, or sufficiency, of the codified communication also matters, as does the reliability of the source and the availability of means of certifying the accuracy of the information. If essential data are omitted or

to describe the possibility of a given piece of information being used concurrently by a number of independent users. Nonrival, however, is a confusing term in this context because although it is intended to indicate that the nature of the good itself does not *require* competition among individuals for its possession, such individuals well may be economic rivals in the market and may therefore compete for exclusive possession of the information in question.

thought to be inaccurate and are costly to validate and to replace, the value of the message to the receiver must be reduced.

On the supply side, whether or not knowledge is put into codified form is in part a question of how costly it is to do that and in part a matter of whether and to what extent there are rewards for the extra effort entailed. Certain kinds of knowledge that are tedious and costly to describe precisely in words may be easily and cheaply conveyed in pictorial or other modes. Graphical representations can be directly displayed and are less subject than expressions in natural language to errors of translation. Consequently, it would seem that the first-order effects of improvements across the range of information technologies would be to increase the extent of codification, as well as to expand potential international access to the stock of codified scientific and technological information.⁵ Yet, somewhat paradoxically, because the process of codification enhances the convenience of transactions involving information commodities, it interacts with technologically driven reductions in the costs of transmitting and processing data to render the knowledge they convey all the more subject to joint possession. It also makes the retention of exclusive use more dependent on the interposition of artificial obstructions to access. I shall return to consider the implications of this for both the producers and the users of such knowledge.

The foregoing discussion of codified knowledge should not create the impression that knowledge of all kinds can now be transmitted at negligibly small marginal cost or that the private and social costs of filtering, interpreting, and utilizing information are insignificant. In juxtaposition to codified knowledge, the concept of *tacit knowledge* refers to the common perception that we all are often generally aware of certain objects without being focused on them.⁶ That does not make them less important: they form the context in which focused perception becomes possible, understandable, and productive.

Like the knowledge that is codified and packaged as information, tacit knowledge can, in principle, be freely shared or exchanged. The transfer process usually involves demonstrations, personal instruction, and the provision of expert services (such as advice and consultations) by those who possess the knowledge that remains in an uncoded form. But the arrangements for effecting such transfers, and the extent to which they occur among scientists and between scientists and members of other communities—such as industrial and military technologists, health care providers, and the polity at large—are various and remain to be studied more fully.

5. Eisenstein (1980) emphasizes the important effects of the wider circulation of scientific findings (both valid and nonvalid, as it turned out) that flowed during the seventeenth century from the reduced costs of printing books and pamphlets containing pictures accompanied by textual explanations and comments. The modern analogue is the revolution that is in progress in the computer generation and transmission of scientific and engineering information in graphical forms. See, for example, National Academy of Sciences (1989).

6. This follows Polanyi's (1966) conceptualization.

Recent discussions of the economics of R&D and technology transfers, however, do recognize the significance of tacit components of technological knowledge. They emphasize that the information contained in papers, patents, blueprints, and other forms of codified knowledge often will not be sufficient to implement the technology in question. Complementary *know-how* is required, and to acquire it will be often an expensive proposition.⁷ The worlds of the technologist and the scientist resemble one another more closely in this regard than has been supposed by some of the leading philosophers of science—and those who have paid them heed. Nagel (1961), Hempel (1966), and Popper (1968) emphasized the epistemological bases for distinguishing between the scientific and other modes of human understanding, stressing formal methodological rules of scientific procedure. By contrast, Polanyi's (1966) insights concerning the role of tacit knowledge have led more recent observers, such as Latour and Woolgar (1979), to see the practice of science as a craft. That is, science draws in an important way on field- and laboratory-specific sets of skills and techniques—expertise in scientific know-how. These forms of knowledge are acquired experientially and are transferred by demonstration, rather than being reduced immediately or even eventually to conscious and codified methods and procedures. Thus both codified and tacit knowledge are generated by scientific researchers, and both are used to produce further knowledge.

Feedbacks and Interactions between Advances in Technology and Science

Advances in the domains of science and technology appear entangled with one another in complicated ways. It will become evident in the next subsection that the interdependent relationships between the conduct of scientific research aimed at publication, on the one hand, and the pursuit of proprietary knowledge for commercial purposes, on the other hand, must be affected by the mix among the various institutional arrangements established in a given society to support research activities. The mix of institutional arrangements also affects the balance between the growth of the stock of scientific knowledge that is codified and made public and the accumulation of tacit knowledge. Much of the knowledge that we are apt to designate as belonging to the realm of technology is likely to take a form that is specific as opposed to general, and idiosyncratic as opposed to universal. Although it may be codified so that property rights to its use can be secured, much of it will remain tacit.

Before discussing the institutional complications, it is worth noting that technological progress impinges in a causal way on the growth of the stock of scientific knowledge. The connections I have in mind here are essentially epistemological, for they have to do with the way in which scientific knowledge is gained. I have in mind here something additional to the point that artificial

7. See Pavitt (1987), Rosenberg (1990), Nelson (1990), and, with special reference to the role of tacit knowledge in technology transfers involving developing countries, Arora (1991). The early work of Teece (1976) recognized the problem of tacit knowledge as a major source of the surprisingly high costs of technology transfers within large multinational corporations.

phenomena may pique the curiosity of the purest of scientists—as the effects observed in the cycle of the steam engine's operations provoked the nineteenth-century development of thermodynamic theory, beginning with the work of Sadi Carnot. I want instead to emphasize what may be called the “instrumentation interconnection” between science and technology.

Practicing experimental scientists have to make observations, measurements, and calculations. Some of the apparatus and technical skill that they develop in the course of their work is quite simple, and some of it represents the solutions to enormous technological challenges. High-energy particle physicists—at facilities such as CERN in Geneva, FermiLab in Batavia, Illinois, and SLAC in Stanford, California, for example—have constructed huge and intricate pieces of scientific apparatus to generate, detect, record, and analyze the products of particle collisions. Just as such technologies pose explanatory challenges to the scientists, the work of the scientist poses technological problems, the solutions to which may find applications in industrial spheres.⁸ In this respect the expansion of the scale and diversity of scientific research activities has contributed in a very direct and immediate fashion to generating new technologies. The use of laser techniques for aligning laboratory apparatus, and their subsequent wider application in surveying and construction work, is one instance. Similarly, the computational needs of natural scientists (perhaps even more than those of engineers) have remained a force driving the technical development of high-speed electronic computing.

That is only half of the story. The other half, involving the reverse flow of causation along the instrumentation interconnection—from the improvement and standardization of instruments for measurement and data analysis to the practice of science—is no less important. The standardized definitions of physical units, and the associated concepts and law-like relationships among them, permit scientists working in different laboratories situated in different countries throughout the world to communicate their findings readily and so cooperate in adding to the stock of verified knowledge. Many of the standardized definitions are operationally grounded in standardized measurement procedures. These procedures can be implemented with standardized apparatus (for example, pieces of equipment as simple as the voltmeter and the ultracentrifuge or as complicated as the electron microscope) that can be produced in long production runs and marketed at far lower unit costs than the laboratory-constructed prototypes. Thus standardization has a dual influence on the progress of scientific knowledge.

First, the introduction of such equipment almost invariably raises the task productivity of scientific researchers by permitting faster and more accurate measurements and data analysis. Of course, qualitative advances have resulted from being able to perform experiments that previously were infeasible because

8. For some illustrative details in the case of high-energy particle physics, see David, Mowery, and Steinmueller (1988, 1992).

the range of measurement error was too great. But the quantitative impact of technological advances that are embodied in industrially produced scientific instruments also has been impressive. As a result of improvements in instrumentation techniques, including the development of generic computer software for data processing, storage, retrieval, and network transmission, the unit variable cost of performing some research tasks has declined by as much as two orders of magnitude, and in some instances three orders of magnitude, over the course of the past thirty years (Moulton, Young, and Eberhardt 1990). While significant increases in the fixed capital costs of doing science have been experienced in many areas of research, advances in instrumentation technologies have been dramatically reducing the price-performance ratios of many types of scientific equipment.

I would contend that in some part the long-run international tendency toward relative expansion of the volume of R&D—vis-à-vis other activities conducted in the private and public sectors—has reflected the supply-side push effects of falling unit costs of R&D performance. Growth in R&D does not simply respond to derived demand generated by the growth of (final) demand for novel goods and services and for more efficient processes of production. This is not a view that can be held if one maintains the traditional conceptualization of one-way causal relationships, such as are represented by the traditional linear model of scientific and technological change.

Second, the tendency toward increasing capital intensity in scientific research activities gives rise to equipment that embodies instrumentation techniques that are increasingly standardized. Once standard equipment becomes available, the research procedures that are designed to make use of it become easier to replicate at different sites, and new scientific findings can be more quickly transferred from one laboratory for applications in others. The trend toward decentralized collaborative pursuit of scientific knowledge (“globalizing” the conduct of research) has thus been furthered by the greater availability of standardized procedures embodied in new instruments of generating and analyzing data, as well as by the availability of high-speed digital communications networks for linking spatially separated researchers.

The discussion to this point has stressed the falling unit costs of performing selected research tasks. It is also important to acknowledge that in many areas of basic science and industrial R&D the physical facilities and scientific apparatus required are not only technologically complex but also have absolutely large and indivisible fixed costs. The capital sunk in such facilities may strain the financial resources of large corporations and even national governments, creating pressures for the formation of consortia to undertake collaborative funding of shared research facilities. The need for some centralized management of the facilities located at collaborative research sites, and of the social interactions among the researchers, raises the costs of having the participants maintain individual proprietary control of information through secrecy. Hence, the technological developments affecting the conduct of company-financed research have

been conducive to the adoption of cooperative R&D strategies, especially where, as in the United States, the antitrust laws are relaxed to allow so-called precompetitive research collaborations.

The feedbacks from the advances in technology to the practice of science through the instrumentation interconnection also may have organizational effects that are biased toward collaborative research. It is true that in some contexts standardization of research techniques favors decentralization and the geographic dispersal of proprietary R&D activities by transnational corporations. But in my view standardization more strongly supports the collective pursuit of *public* knowledge.⁹ That is, the drift of technological progress is tending to increase the potential power of the cooperative mode of inquiry that has been associated historically with the norms and institutions of "open science." In open science new findings are quickly disclosed and widely disseminated so that they may be verified and used by others to generate further additions to the stock of public knowledge. Yet the treatment of scientific findings as public goods—which is required to reap the benefits of this technologically enhanced mode of collaborative inquiry—is itself a highly problematic feature in the organization and funding of modern research.

*Appropriability Problems and Institutional Arrangements for Science
and Technology: "The Three Ps"*

Nonrival possession (made possible by the "perfect expansibility" of ideas), low marginal cost of reproduction and distribution (making it more difficult to exclude others from access to information), and substantial fixed costs of original production are three properties generally associated with public goods. And, as is well known, competitive markets cannot be relied on to perform well in allocating resources to the production of goods with those characteristics. Where prices are driven toward marginal costs, the revenues received by competitive suppliers will not even cover their full costs, much less approach the use-value of the goods to the consuming public. Indeed, the attempt to make the beneficiaries pay for value received would so reduce demand as to result in an inefficiently low level of consumption. All this was appreciated three decades ago in the path-breaking work of Nelson (1959) and Arrow (1962) on the "appropriability problem" and the economics of R&D. From that time forward, the principal economic rationale offered for public policy interventions affecting R&D activities has been the putative failures of competitive markets (a) to elicit revelation of the actual demand for new scientific and technological knowledge and (b) to provide private individuals and organizations with sufficient incentives to induce the socially optimal amount of investment in the production of such information.

In the literature of public finance economics, alternative allocative mechanisms have long been recognized as solutions to the public goods problem.

9. See Pavitt (1992) for a discussion of recent trends in the internationalization of R&D activities and their significance.

There are three principal alternatives. One is that society should provide independent producers with *subsidies* financed by general taxation and should require that the goods be made available to the public freely or at nominal charge. A second solution would have the state levy general taxes to finance its direct participation in the processes of production and distribution, *contracting* where necessary with private agents to do the work. The objective in this approach, again, is to supply the good without having to charge a price that recovers its costs of production. The third solution is to create a publicly regulated *private monopoly* and to allow it to charge customers prices that will yield a normal rate of profit. There is a striking correspondence between this set of solutions for the standard public goods problem and the three main institutional arrangements that have been devised to cope with the appropriability problems which arise when competitive markets are left to guide the production of knowledge and pure information goods.¹⁰ I will refer to the latter arrangements as "the three Ps": Patronage, Procurement, and Property.

The term *patronage* stands for the system of awarding publicly financed prizes, research grants based on the submission of competitive proposals, and other subsidies to private individuals and organizations engaging in scientific discovery and invention, in exchange for full public disclosure of their findings. Patronage characterizes the pursuit of open scientific inquiry and is the dominant institutional and social mode of organization associated with the conduct of academic science in the democratic societies of the West.¹¹ *Procurement* is associated with governmental contracting for intellectual work generally and for scientific research performance in particular. Whether or not the information produced will be disclosed for public use is an important policy issue, but a secondary one, the determination of which is left in the hands of the procuring agency. Thus, defense-related R&D typically is carried on by government employees and private contractors under secrecy restrictions in secure facilities, whereas much public-contract R&D and the scientific work of governmentally operated laboratories and agricultural experiment stations is undertaken with the intention of disseminating the findings rapidly and widely.

The third institutional solution is for society to grant private producers of new knowledge exclusive private *property* in the use of their creations, thereby forming conditions for the functioning of markets in which the originators will be able to collect (differential) fees for the use of their work by others. Here we come to the specific legal contrivances that define and institute intellectual property: patents, copyrights, and, somewhat more problematically, trade secrets.¹²

10. This parallelism is treated more fully in Dasgupta and David (1988); the discussion here also draws on the more detailed presentation made in David (1992a) concerning the drawbacks and tradeoffs among the alternative arrangements applied in the case of knowledge production.

11. On the connections between patronage and the historical emergence of the institutions of open science, see David (1991b) and references therein.

12. Unlike patent and copyright statutes, the law of trade secrets (even when given statutory structure)

None among "the three Ps" provides a complete and perfect solution to the problem that they all address. Some field of useful employment has been found for each type of institutional arrangement, but no one has emerged as clearly superior to the others in all contexts.

In most practical applications both the patronage and the procurement solutions will be marred by the problems of informational asymmetries. Theoretically, of course, one can conceive of establishing criteria for the award of prizes and research grants or of drawing up procurement contracts that would reward researchers commensurately with the anticipated social use-value of their discoveries and inventions. Unfortunately, however, the public authorities generally do not know enough *ex ante* to set efficient contractual terms for such awards and prizes. In thin markets where the expertise about the likely costs and benefits of particular research projects is unevenly distributed, contracting will entail high transaction costs even to arrive at imperfect agreements. Ever since Adam Smith,¹³ economists have delighted in noting that this particular drawback can be avoided simply by granting intellectual property rights and letting the economic rewards be determined *ex post* through the workings of the market, as is done where patentholders are allowed to sell licenses to use an invention.

Alas, as economic critiques of the system of statutory protection for private rights in intellectual property repeatedly point out (for example, Plant 1934; Machlup 1958; Kaufer 1989), the property solution, unlike patronage and procurement, inherently entails restricting the extent of useful application of the new knowledge by permitting the imposition of license and royalty charges on the users. The more secure is the patent monopolist or the copyright holder in possession of the right to exclude others from using the information in question (even though it has been publicly disclosed), the higher are the charges that can be levied. This reduces the benefits that would have accrued to society at large, and to consumers in particular, had the information been made available for competitors to exploit in the form of new products or production processes. These lost benefits have been referred to in the conventional jargon of economists as the deadweight burden of (patent and copyright) monopolies.

Relying on the property solution brings additional drawbacks, which have grown more worrisome as greater importance has been attached to the utilization of scientific and technological findings as intermediate (informational) inputs into other R&D projects. The maintenance of secrecy, of course, raises the

is rooted in common law principles, including theories of contract and tort, as well as property concepts. Indeed, the general tendency among modern legal scholars is to deemphasize rights to property in information that is held secret but to argue for the appropriateness of the indirect protection that is provided to those who invest in the origination of such knowledge by enforcing their rights to establish relationships of *confidentiality*—for example, between industrial firms and their employees, including workers engaged in research. See Jager (1991, pp. 4–9) and David (1992a, pp. 14–15) for further discussion.

13. See the citation in Smith (1776/1776, p. 754, n. 69) to the discussion in his *Jurisprudence*, A.ii: pp. 31–33.

costs of the search for new knowledge borne by the researchers themselves and, ultimately, by the whole of society. The withholding of information, even temporarily, leads to social inefficiencies by making it difficult, if not virtually impossible, to coordinate the planning of research pursued by different entities and so avoid delay and duplication by optimally sequencing the solutions to the myriad subproblems that make up a research problem of any realistic degree of complexity.¹⁴ In order to enforce property rights under the laws of trade secrecy, firms are obliged to take deliberate measures to keep their practices secret. But even where it is more advantageous eventually to disclose research findings to obtain the benefits of statutory protection under patent or copyright law, those modes of appropriation create strong incentives to hide the process of one's research from rivals.

This latter source of social inefficiency (in the conduct of the search for new knowledge itself) is a direct consequence of the payoff structures that characterize not only intellectual property rights systems but also the organization of open science under patronage arrangements. At the root of the trouble is the tendency of property and patronage to draw researchers into racing against one another for *priority* of invention or discovery. To secure the benefits of a rapidly accumulating stock of knowledge, it is desirable to elicit speedy disclosure of new findings so that these may be disseminated, verified by replication, and put to use by others. Only in this way can the fullest scope be provided for the realization of the cumulative, interactive process through which ideas proliferate and generate still more ideas. To achieve that goal, full and prompt disclosure must be sought under the patronage and the property approaches, and both, therefore, are impelled to base their assignment of rewards in some way on the establishment of priority. In the case of patronage this is achieved by offering prizes and by awarding subsidies in the form of research grants to those who have established reputations for research success. Among academic scientists in the West, such reputations are won by being able to claim priority in discoveries and inventions that expert peer groups deem useful contributions to the collective pursuit of knowledge.¹⁵ Under the modern patent system the role of priority is equally central: it appears in the criteria of originality (having originated with the inventor in question) and novelty (not having been invented independently by another) which must be satisfied if a patent is to issue.¹⁶

14. Simulation studies by Fölster (1985) suggest that the losses from suboptimal sequencing of the solutions to the constituent parts of a large research problem may be quantitatively more important than the duplication of effort caused by the common pool problem, discussed below.

15. On the importance accorded to matters of priority in scientific communities, which has long been recognized by sociologists of science, see Merton (1973); Lamb and Easton (1984) provide an historical survey of priority races and priority disputes. The functional significance of priority in the organization of science is examined from the viewpoint of economics by Dasgupta and David (1987, 1988), from which the present compressed treatment is drawn.

16. Under the U.S. Patent Act (35 U.S.C., Sec. 1-376) a patentable invention must also satisfy the criterion of nonobviousness (not being obvious beforehand to a person having ordinary skill in the pertinent art). Priority of invention is what is required under the U.S. patent system, but in other (Western

A common implication of priority-based reward systems is that they give rise to competition—whether among reputation-seeking scientists or among patent-seeking inventors—that is fueled by the winner-takes-all structure of the payoffs for those who participate.¹⁷ The contestants consider only what they individually stand to gain, not the effect of their participation on the expected outcomes for all the other competitors. The situation therefore resembles the inefficient resource allocation patterns that emerge when there is a common pool problem: seen from the societal viewpoint, there will be too many contestants entering the races for priority in discovery and invention (see Dasgupta and Stiglitz 1980; Wright 1983). Furthermore, in addition to the crowding of the field with contestants, private rents tend to be dissipated in accelerating the race for priority. The private value to the researcher(s) of arriving at a new finding a little sooner than the second-place contestant is likely to greatly exceed the benefit that society as a whole will derive from the slight consequent advance in the date of discovery or invention.

Whether or not the inefficiencies caused by excessive investment in competitions for priority are a serious defect of the patronage and property solutions, the winner-takes-all structure of the payoffs from “research races” results in the potentially more wasteful failure to coordinate the research efforts of rival investigators (Fölster 1985). The effort to win races leads researchers of all stripes to shroud their ongoing efforts under a cloak of secrecy. This is a deviance from idealistic social norms subscribed to by the members of ostensibly open-science communities. Nevertheless, such behavior is encountered frequently enough to have gained widespread toleration.

Public Policy and the Funding of Basic and Applied R&D

The argument generally offered for public policy interventions to enforce intellectual property rights is that there is a market failure. That is, in the absence of governmental interventions, competitive markets would not provide private parties with sufficient inducements to undertake the socially optimal amount of investment in creating public goods in the form of new scientific and technological knowledge. This problem is likely to be especially acute in lines of research where the outcomes are more uncertain and where the implications of the findings are contingent on developments in other, possibly remote, domains of knowledge.

Although fundamental, or basic, science research has these characteristics, it also has consequences for those seeking knowledge with more immediate and

European) national systems the touchstone is priority of registration. Because novelty is not a requirement for the protection of property under copyright law, priority does not play quite the same role there, even though the absence of priority of registrations makes it more difficult to meet the requirement that the material registered for copyright must have originated with the author in question (Bender 1986).

17. This pertains to the awarding of patent rights in the case of rivalries among inventors. It does not imply that the original recipient of a patent automatically captures all, or even the lion's share, of the private economic gains deriving from a successful invention.

more predictable utilitarian implications, rather than being only an intellectual input into further research. One should think of the funding of fundamental scientific inquiries as providing maps to guide mission-oriented researchers, directing explorers on the applied science frontier to the more fruitful areas, and sparing them the wastage of time and resources in searching barren regions or trying to cross unbridgeable chasms (David, Mowery, and Steinmueller 1988, 1992). Basic research may, of course, yield unexpected discoveries that have immediate practical uses, some of which will be extremely valuable, as has been true in the cases of lasers and of enzyme restriction techniques for recombinant DNA in the field of biotechnology. These, however, are the rare exceptions. More typically, the important economic payoffs from basic research come in the form of higher rates of return on expenditures allocated for applied research. From the societal perspective, basic and applied research should be viewed as complementary activities. The problem is that by investing in applied R&D, firms and nations can reap much of the benefit of prior basic research to which they may have contributed nothing. The existence of this manifestation of the familiar appropriability problem, and the linkage between applied and basic R&D, however, do not imply that the best remedy is to institute some new and strongly protected private rights in intellectual creations that take the form of fundamental scientific discoveries.

It appears that for a society to arrange for just the right amount of investment in creating new knowledge of the right kinds is rather more difficult than economists first supposed. The incipient tendencies toward underinvestment caused by knowledge spillovers, free-riding behavior, and the difficulties that individuals and firms encounter when seeking to privately appropriate the benefits certainly can be overcome by instituting protections for intellectual property rights. But there will be costs of doing so, namely, the costs of creating payoff structures that channel R&D investment into activities where the expected social marginal rates of return are far below the expected private marginal rates of return. The consequence is what Dasgupta and Maskin (1987) refer to as the tendency toward "excess correlation" in research portfolios among firms engaged in racing one another for patents in areas where the profit potential of winners appears to be highest. These congested (and contested) territories will coexist with other, much less populated, fields of inquiry where researchers are comparatively starved for funding despite the potentially high social rates of return that incremental efforts could be expected to yield. The current distribution of expenditures on pharmaceutical R&D illustrates this.

Intellectual property rights generate concurrent areas of excess and deficient investment in R&D, thus falling rather short of allocative efficiency. Unfortunately the situation is not much improved by the possibilities of privately appropriating the benefits of inventions and discoveries in production methods that can be protected as trade secrets. The most reliable mode of private appropriation entails being first to exploit the newly acquired knowledge by establishing a cost advantage in actual production operations—racing potential competi-

tors to be farthest down the learning curve for the new activity.¹⁸ Thus there is likely to be a rent-dissipating scramble to come up with R&D results.

Furthermore, the public patronage of open scientific inquiry, carried on in a decentralized fashion by communities of largely autonomous researchers, also is not free of the tendency toward excess correlation of research strategies and racing behaviors among rival investigators and their research teams.¹⁹ Even the purest of pure scientists are human, driven by a mixture of personal ambition and socially reinforced patterns of behavior and responsive to the various reward systems that their environment provides. This is evident from the recent frequency with which academic scientists have participated as major equity-holders and executives in start-up ventures organized to exploit their knowledge, as well as from the longer history of consulting relationships between industrial firms and university-based researchers in the natural sciences, engineering, and medicine.

The tensions from conflicting norms of scientific conduct are likely to be relieved by artificially exaggerating the differences between the nature of the activities carried on in the different organizational or institutional milieus. Thus, it is easier to maintain conformity with the communitarian norms that call for full disclosure of findings and cooperation among academic scientists (and government scientists engaged in nonmilitary research) when the phenomena under investigation are perceived as being "fundamental," in the sense of not presenting immediately realizable opportunities for lucrative commercial application. In this way the differences in the reward structures of the nonprofit and private enterprise research communities induce a tendency for their research products to become more starkly differentiated along the basic-applied axis: the codified results disclosed in scientific papers aim at greater generality but tend to omit discussions of the specific, utilitarian applications of the finding under discussion. Such "practical" knowledge as may have been uncovered in the research process remains tacit when research findings are reported in scientific journals and at public conferences, even though the act of publication itself announces the possibility that arrangements could be made for the private transfer of commercially relevant information from the scientists involved.

Industrial enterprises may support open science by funding basic science research because they recognize the potential advantages of gaining access to the stores of codified and tacit knowledge that are made available to scientists through their membership in invisible colleges and specialized networks for

18. This is so because trade secrecy law does not convey an exclusive right; it provides no protection against competitors who acquire the same knowledge by socially conscionable means—which include not only independent research activities but also deliberate reverse engineering of one's product or decompiling the (machine language) object code of one's computer program to learn the source code (Bender 1986). Technological advances in such methods make it increasingly difficult to preserve industrial secrets for very long, so that an innovating firm can expect to protect only temporary lead-time under the law of trade secrecy.

19. See Dasgupta and David (1990) for fuller treatment of resource allocation problems in publicly supported science.

information exchange. Mowery (1983) and Cohen and Levinthal (1989) have called attention to firms' monitoring of external technological change through their conduct of in-house R&D, which is often overlooked in the emphasis usually placed on the internal generation of innovations through the performance of R&D. Rosenberg (1990) suggests that both the monitoring function and the development of capabilities for absorbing scientific knowledge may be important motives for company-financed "basic" research.

Scientifically sophisticated business corporations may in this way be better able to keep abreast of the advance of scientific knowledge in certain fields and to develop capabilities for rapidly integrating and utilizing new information about research methods and findings that are of commercial relevance in their branch of industry or in new branches that they may be enabled to enter.²⁰ Yet it is difficult for any individual firm to fund fundamental scientific research that is not profit oriented on a scale that is large in relation to its proprietary R&D activities.

One rather obvious reason for this lies in the uncertainties surrounding such long-term, option-like investments and in the necessity for such outlays to be financed internally from the firm's retained earnings. This is simply another aspect of the appropriability problem. Another limitation is the difficulty a profit-oriented organization will encounter in establishing a credible commitment to commercially disinterested patronage of scientific research. To gain full access to the club and its facilities for information exchange, the company must be accepted as being committed to having the scientists in its employ behave as if they were independent members of the open-science community. Doubts about this will intensify, quite reasonably (since the first-mentioned problem is widely recognized), with every increase in the scale on which the firm pursues such a program. Justifiably or not, scientists engaged in basic research under the auspices of private corporations often find themselves isolated by suspicions that the company which employs them may be trying to "free-ride" and is limiting their disclosure of proprietary information. Thus, profit-seeking motives for conducting basic research cannot be expected to reduce significantly the needs of modern industrial societies to support open-science research communities through various public patronage mechanisms.

II. DYNAMIC FEEDBACKS BETWEEN INNOVATION AND DIFFUSION

The dynamics of technology adoption (diffusion) processes are closely intertwined with the dynamics of technology development resulting from endogenous, incremental innovation—rather than belonging in the separate com-

20. Whereas these arguments concern the demands of firms for the products or byproducts of basic research activities, Dasgupta and David (1987) suggest the existence of a different set of considerations that motivate individual researchers to participate in open science (doing basic research for rapid publication), even though it is their aim, eventually, to find employment in industry (engaged in applied, proprietary research).

partments assigned to them by SLIM (David 1986; Stoneman and David 1987). This point is distinct from and, in the significance of its implications, goes well beyond the more widely acknowledged linkage between innovation and diffusion that was examined in the preceding section. There it was noted that although R&D performance and the generation of innovations might be stimulated by instituting strong protection for intellectual property, such a policy would adversely affect diffusion by raising the costs of access to the protected new technologies. Here it is important to emphasize that a variety of feedback effects link the processes of innovation and diffusion. Some of these connections generate *negative* feedback effects which tend to stabilize the rate of technological progress and therefore may cause policy interventions to have surprisingly perverse consequences. The dominance of *positive* feedback effects, however, would have the destabilizing consequence of producing "cross-catalytic" reinforcement of the dynamic behavior of the coupled innovation-diffusion subsystem. Thus, an impulse promoting a new technology's adoption leads to the quickening of the rate at which it undergoes further technical improvements; that, in turn, promotes further widening of the sphere of the technology's diffusion, and so forth, until it has displaced rivals and saturated the market. But the same structure that is capable of generating virtuous circles and upward spirals can also lead to the establishment of vicious circles: early failure of an innovation to penetrate the market can deprive it of opportunities to undergo subsequent improvements that would remedy its initial defects, thereby blocking its eventual diffusion.²¹

The New Microeconomics of Technology Diffusion

The view just sketched of the variety of outcomes that may emerge from the interdependence between the innovation and diffusion processes rests on the new perspective that has been gained by empirical and theoretical research on the microeconomics of technology adoption. The classic early studies of the diffusion of agricultural and industrial innovations were carried out by Griliches (1957, 1960) and Mansfield (1968; Mansfield and others 1971). They conceptualized the process as one involving a new technology with engineering and economic characteristics that were prespecified and unchanging and an equally unchanging population of potential users who had to be persuaded of the profitability, or utility, of using the innovation. The objective economic environment was one in which the only consequential change taking place was the gradual dissemination of information about the benefits of the new technology. This might occur as a "contagion"—through the contact between adopters and

21. Dynamics of this kind are exhibited in David and Olsen's (1986, 1991) formal models of rational decisions about when to adopt innovations, given learning by doing in the supply of a capital good embodying the new technology and under conditions in which there are heterogeneities among firms that are all fully informed and have consistent expectations which the dynamics of the system fulfills.

nonadopters in the population.²² Thus regarded, the gradual increase in the extent of an innovation's application across the firms and sectors of the economy has the appearance of an adjustment process, which eventually approaches the restoration of equilibrium.

An alternative conception, however, has emerged from consideration of the historical and contemporary evidence that many new technologies are initially introduced in forms, and under market conditions, that make them profitable in immediate applications for only some firms, and perhaps only in the operations of some industrial plants or departments within those firms. Subsequently, however, as the technology and its microeconomic environment evolve, the range of profitable applications is broadened (David 1969, 1975, 1984; Rosenberg 1972; Gold 1979, 1981; Sahal 1981).

The class of so-called equilibrium diffusion models constructed on these foundations emphasizes two points. The first is that even if information relevant for rational decisionmaking about the innovation were instantaneously disseminated without cost, there would remain many reasons why states of equilibrium would exist involving less than complete diffusion of the new technology within the industry. Heterogeneities among the potential adopters with regard to their objective circumstances that affect the choice of technique could lead some of them to prefer alternatives to the innovation in question.

Firms considering a new production technology embodied in fixed equipment may face different raw material costs, energy prices, and transport charges; they may differ in the makeup of technically related product arrays produced using joint facilities; they may operate in different labor markets and have different implicit or explicit contractual commitments to their employees; and they may encounter different terms for borrowing. Furthermore, such firms are likely already to possess some durable capital equipment of varying ages and vintages, which they would have to retire were they to adopt the new technology. Plant and equipment that is younger tends to have lower unit operating costs and is consequently more likely to continue generating some quasi-rents. Another differentiating circumstance arises where the innovation entails use of a technique of production that is fixed-factor-intensive in comparison with the preexisting alternatives: larger-scale producers will be in a better position to enjoy a reduction in unit costs because they can spread the greater fixed costs of the new technology over bigger production runs.²³

22. It is well known that if such contacts came about—one at each instant in time—through random mixing, and each had the same probability of generating a new (permanent) adoption, the expected time path of the proportion of a stationary population that already had adopted the innovation would follow the logistic.

23. David (1969) formalized threshold-size models of adoption and showed how the combination of a lognormal distribution of firm sizes and an exponential rate of decline in the break-even (or threshold) scale for adoption of a fixed capital-intensive technique would generate adoption paths with structural parameters that could be estimated econometrically by regression-probit analysis. Davies (1979) also proposed the estimation of probit models of diffusion, without that rationale.

Second, the extended duration of the upward course of the diffusion curve could be tracing a moving equilibrium in the distribution of the population of fully informed, rational agents—between those for whom adoption is advantageous and those for whom it is not. This follows directly from the first set of propositions, and it implies that a complete account must identify the forces driving the equilibrium point (corresponding to the point of indifference between adoption and nonadoption) through the population. An important set of forces that play this role consists of those that operate on the supply side to progressively lower the cost-performance ratio associated with the use of the new technology (David 1969; Stoneman and Ireland 1983; David and Olsen 1986).

The terms on which users can acquire effective access to new technologies will reflect one or more of three classes of cost. The first class contains the costs of securing and evaluating information about the technologies. The second is the class of costs of obtaining the specialized materials or equipment in which new technologies having particular performance characteristics are physically embodied by supplying firms. The third class includes the costs of specialized facilities, ancillary products, and services that are technically complementary to the innovation and thus will affect its performance. All three classes of technology-access costs share a common feature that is crucially important to the dynamics of diffusion processes: each is likely to undergo a decline in response to the widening use of the technology in question.

An example is coordinated technology transfer involving contractual arrangements between firms, or managed transfers between units of a large organization such as the multinational corporations studied by Teece (1976). This appears to be a decreasing-cost activity in the sense that its costs decline with each application of a given innovation or with each start-up undertaken by the transferring organization. In addition, uncoordinated transfers of technical information can and do occur through the accepted mobility of scientific and engineering personnel within, and sometimes between, industries. The larger the cadre of scientific and engineering workers who have become familiar with a new technology, the greater is the potential for the outward percolation of information through those channels.

With regard to the second category of access costs, it is essential to appreciate that technologies do not remain static. Typically they undergo a gradual evolutionary development along a trajectory that is bound closely to the course of their diffusion. Initial versions of new products or production systems often suffer from numerous flaws in design or manufacture. They are plagued by bugs or defects, the identification and correction of which in many instances depend on the accumulation of feedback information from the users. Rosenberg (1982) has labeled this process learning by using, by analogy with the widely documented phenomenon of learning by doing. Along both kinds of learning curves, the accumulation of experience that diffusion itself makes possible is what will govern and sustain a flow of incremental improvements that the supplying firms

may pass along as reductions in the cost-performance ratio that faces potential adopters.

Finally, as has recently become more widely appreciated among economists, network technologies—distributed, multicomponent systems within which there are strong technical complementarities—dramatically exhibit many of the same dynamic features in their development and elaboration as have here been recognized in freestanding product and process innovations. As a telephone network's coverage is extended by linking up additional subscribers, the cost of providing basic services to each subscriber will decline. The advantages of being able to access and be accessed by a widening circle of subscribers will raise the utility of joining the system for successive entrants. The special economic problem posed by such systems, however, is that integration requires some measure of technical compatibility or standardization and thus imparts to those standard "interfaces" the characteristics of public goods. The benefits derived (or the effective costs incurred) by any one prospective system user may be dependent on the willingness of others to incur the costs of achieving compatibility with the same network.²⁴

The existence of these sources of positive feedback brought about by the irreversible, dynamic, decreasing cost effects of the diffusion of a new technology implies that small initial advantages or disadvantages (and seemingly transient impulses, including policy shocks experienced at an early stage of a new technology's history) can cumulate readily into large advantages or disadvantages in comparison with alternative technologies. A particular product design, process technology, or organizational system thus can become "locked in," while rival technologies are "locked out" through the workings of decentralized competitive market processes.²⁵

From the policy standpoint the implications of this are significant indeed.²⁶ The dominance of positive feedbacks means that multiple outcomes are possible, some of which would be better than others from the standpoint of their ex post efficiency. It means that historical circumstances may matter vitally in the evolutionary selections made among alternative technologies. Furthermore, it means that there will be critical phases, typically rather brief in duration, during which modest policy interventions can exert disproportionately great leverage over the course of future developments. And there will be other, more protracted, phases when policymakers must recognize that very large commitments of resources would be required to successfully overcome or redirect the momentum that has been developed by the cumulative process of diffusion and incremental improvement along a particular technological trajectory.

24. For a survey of the recent and rapidly expanding literature on network externalities and compatibility standardization, see David and Greenstein (1990).

25. On this subject and the broader theme of path-dependence in economic processes, many references could be given to works of mine and of my Stanford colleague W. Brian Arthur and our students, but the following will suffice for the present occasion: Arthur (1989, 1990) and David (1985, 1988, 1992b).

26. Some of these have been explored more fully in David (1987, 1992c).

Feedback Effects and Some Technology Policy Paradoxes

Although a conceptual separation between diffusion phenomena and innovation phenomena may be useful for purposes of economic analysis, eventually their study must be reintegrated. The importance for intelligent economic and technology policymaking of taking a more integrated approach to the design of innovation and diffusion policies cannot be overstated. This point may be illustrated briefly by considering a few potential surprises or paradoxically perverse outcomes that might be produced by proceeding to set policies in ignorance of the actual interconnections that link the two areas of microeconomic behavior.

First, efforts to speed up the rate of innovation in industries supplying capital goods can create expectations of larger capital losses through obsolescence for firms that are considering adopting the new technology (Rosenberg 1976; Ireland and Stoneman 1986). If such expectations are entirely correct, a rational determination of the date of adoption, under conditions of perfect foresight about the future course of input and output prices, would involve deferring investment in the new equipment as long as the marginal cost of delay—which would be the loss of the instantaneous benefits (say, from savings in production costs)—is smaller than the marginal gain from delay. For an innovation embodied in a piece of durable capital, the gains from deferring purchase for another period are the sum of the averted rental costs and the averted capital losses (which would otherwise be incurred because of the falling reproduction costs of a piece of capital equipment with identical performance characteristics). Hence, a credible public policy commitment to accelerate the future rate of innovation in a specific area could retard the current pace of the diffusion of available technologies and thereby slow the rate of productivity growth.²⁷

Second, tax and other subsidies for R&D can reduce the costs of imitation and lead to expected wider diffusion of a new technology through an industry. But if it is expected that everyone else will adopt the technology quickly, the inducements to bear the costs of adopting it early are reduced. So, R&D subsidies may slow the initial speed of diffusion even though they do help disseminate information about the new technology more widely and thus increase the eventual extent of its adoption (Stoneman 1983; Stoneman and David 1987).

Third, providing public funds for disseminating information to users of innovative technologies when supply is monopolized under a regime of patents or trade secrets can induce the supplying firms to set initially higher prices for their wares than they might otherwise choose to do. In the absence of the subsidized information program, a monopolist would consider the demonstration value of exposing more potential adopters to information about the new product by increasing the cadre of those who were using it already. Because the information effect of initial adoptions positively shifts the market demand and marginal

27. For a mathematical simulation model in which the growth rates of labor productivity and total factor productivity growth are derived as functions of the rate of change along the diffusion path, see David (1991a, appendix A).

revenue schedules that the monopolist will be facing in the next period, there is a rationale for reducing the price to encourage more early adoptions—as long as the discount rate applied to future revenues is not too high.²⁸ Thus the combination of a policy measure meant to promote innovation with another measure that is supposed to encourage diffusion may work, perversely, to slow the actual course of adoption of a technology by the market.

The three feedback mechanisms that have just been identified involve the expectations that may be created by pro-innovation policies. The possibility of perverse impacts on the realized rate of technological progress arises because the expectational effects on pricing and adoption decisions are negative. However, these conclusions remain partial inasmuch as the analysis has abstracted from the possibility that positive feedbacks also might be at work. The important message is the need for caution and for careful prior empirical analysis of the entire decision system that may be affected by government policies motivated by the intention to promote technological progress in particular industries or sectors.

III. TECHNOLOGY, TRADE, AND GROWTH IN NORTH-SOUTH PERSPECTIVE

More than fifteen years ago, Harry Johnson (1975) called attention to the importance of technological innovation as the chronic disturber of existing patterns of comparative advantage. Although it has taken some time, the economics profession can now claim to have at last absorbed the point. One obvious manifestation of this is the large and rapidly growing literature associated with the industrial organization approach to international trade theory, and with technology-gap models of trade in particular.²⁹

This recent trend in economic analysis focuses on the relationship among technology, trade, and long-run growth rather than on the classical questions concerning the size and distribution of short-term (static efficiency) gains that could be achieved by moving to an open-economy resource allocation regime. International differences in production techniques, and in technologically based product characteristics, now occupy the center of the field in explanations of the structure of comparative advantage and the pattern of trade; differences in factor endowments (along with the Heckscher-Ohlin models in which they figured so prominently) have been pushed to the sidelines.

A related, distinctive feature of the new analytical outlook is the treatment of the sources of differences in the techniques employed by different countries and regions as endogenous to the dynamic process under examination, rather than as predetermined and place-specific. Economies of scale and externalities of locational agglomeration are realized through the expansion of production and the

28. See Stoneman and David (1987). The "second-best" policy of creating a monopolist who can expect to capture the spillover benefits (in the form of learning effects on the costs of production) by speeding up the initial diffusion process has been analyzed in another context by David and Olsen (1991).

29. This literature is carefully surveyed by Goglio (1991).

geographic redistribution of mobile resources, especially in the case of "foot-loose industries." Technological progress, as distinguished from the realization of static scale economies, is depicted as resting on the accumulation of knowledge obtained by investments in R&D and by the acquisition of production experience. Dynamic, irreversible scale economies, characterized as "learning," can be *localized* geographically, as well as production in technology space.

In the genre of technology-gap models of trade, the differential pace of technological progress is depicted as the force creating international divergences in comparative advantage and asymmetries in national patterns of foreign trade. It is supposed that additions to the stock of knowledge generated in one country will have positive externalities for firms located in that country but will not diffuse rapidly across national borders even though international spillovers and imitation by producers in other countries eventually may occur. General equilibrium models of North-South technological development and trade, such as those constructed recently by Grossman and Helpman (1989, 1990), suggest that the transfer (imitation) of technological innovations generated in the North and their application to the production of tradables in the initially lower-wage South must work to increase the specialization of the two regions. The volume of commodity production employment in the North is forced to contract by the entrance of Southern imitators, and (given the assumptions of full employment and labor malleability) the displaced workers are absorbed in the R&D activities of the North, thereby increasing the global rate of technological progress.

What should we take from this when considering the role of technology and the scope for technology policy measures in developing countries? Is it reasonable to identify those countries with the stylized, technologically laggard South of the "technology-gap" models? How informative is the foregoing characterization of the processes of technological innovation and diffusion (transfer) and of the conditions affecting them? These challenging questions certainly are worth considering at some length. But on this occasion I can best respond to them by briefly introducing two notes of skeptical caution. The first is a comment on the overly simplified conceptualization of the nature of technological knowledge in these models and the implications that has for the treatment of imitation and technology transfers. The second touches briefly on the potentially misleading conclusions that might be drawn from this literature in regard to areas of conflict and harmony between the interests of the industrial and the developing countries in strengthening the protection of intellectual property rights.

Tacit Knowledge and the Problems of "Learning to Borrow"

The basic conceptualization of technological knowledge in the technology-gap literature is that it is codified information that, because it is unambiguous and cheaply transmitted, is difficult to keep from leaking out for more than a brief interval of time. Competitors, it is supposed, soon will learn enough of a new technique to borrow it. A related view is that the essential technological knowledge is embodied in the very design of the product and, with some effort

and expense, can be retrieved by reverse engineering. These characterizations are tolerably accurate for some purposes, particularly when they are applied to the situation of innovating firms in the industrial countries. But because they convey only a part of the truth, the picture that emerges tends to understate the very considerable problems that may attend the transfer of complex modern technologies to developing countries.

As pointed out in section I, in addition to codified technological information there is uncodified, tacit knowledge which is complementary to it. Without the latter, production processes will not deliver output of the expected quality at the anticipated rate. Some of this tacit knowledge consists of the details and material specifications that have been omitted from the blueprints—intentionally, or simply because they are standard in the country where the design originated. Other necessary knowledge may reside in the technical expertise of production engineers or of the work force.

The more complex the process or the product design, the greater the dimensionality of the tacit knowledge problem. For example, as Steinmueller (1989) has pointed out, the knowledge involved in effective production of advanced information technology products entails far more than the transfer of a set of blueprints. Differences in the availability of individual components, subtle effects brought about by seemingly minor adaptations in design to accommodate local least-cost methods of manufacture, and the necessity of achieving interoperability with other information technology systems that are not standard, all may frustrate attempts to implement a patented process or product design. This is especially so if that attempt is made in a foreign industrial setting, without access to the tacit knowledge of agents who have experience getting the technology to work in its original setting.

The greater the technological distance between firms in terms of the degree of overlap or disjointness in the domains of their production experiences, the more serious is the problem posed by tacit knowledge. Tacitness itself implies that the agents will not have a conscious, focused perception of what it is that they know that is critical for successful operations. The smaller the set of analogous experiences, the more difficult it is likely to be for one firm to elicit from the other the list of subjects on which additional information is required. Of course, common training of skilled personnel and easy mobility of such personnel among firms will help greatly in overcoming these problems and hence will increase the likelihood of successful implementation of codified technologies that have been transferred through licensing of patents, reverse engineering, or imitative R&D programs.

It therefore seems unwarranted to assume that technology will leak automatically to developing countries in much the same way (albeit with a little longer lag) that it becomes diffused among competing firms in the technologically advanced region of its origin. A corollary of this, however, is that the successful transplanting of advanced technologies may have important externalities in building up stocks of tacit knowledge in the receiving country, which will make

it cheaper and less risky to transfer other analogous, or technically related, innovations. The development of technological competencies in a particular field, especially one such as information technology, can have pervasive applications across a range of industries and sectors. Development of information technology also will make it more feasible for potential borrowers to identify the most competent innovating firms from which they might seek to acquire access to the newest relevant techniques. The acquisition of competencies of these kinds might well be termed *learning to borrow*.

These self-reinforcing, positive feedback mechanisms imply that there may be marked divergences in the rate of imitation, or technology-absorbing capacity, across industries and even across broad sectors of developing economies. Those, however, have remained largely unrecognized in the technology-gap literature, which has focused instead on positive feedback mechanisms that operate to generate divergences in rates of innovation in the industrial countries. It would be useful to explore the properties of open-economy models that incorporate positive feedbacks affecting the rate of technology borrowing or imitation. Combining those feedbacks with positive feedbacks in the rate of innovation undoubtedly would uncover much richer dynamics, as well as policy implications rather different from those that have emerged to date.

Intellectual Property Protection, Innovation, and Diffusion: A Reprise

Many of the recent technology-gap models of international trade serve to reinforce the view that under conditions of free trade the economic interests of the developing countries are identified with conditions in which it is easy and inexpensive to borrow or copy new technologies (Grossman and Helpman 1990). In these models the interests of the North are seen as bound up almost exclusively with the generation of innovations. In other words, the industrial countries are pictured as operating in the first four compartments of the traditional linear model—they do fundamental and applied science, invention, and innovation, whereas the developing nations are occupied with imitation and technology diffusion.

This stylization, unfortunately, plays into traditional economic analysis of the tradeoffs involved in opting for the property solution of the appropriability problem. It reinforces the long-standing notion that there is an unavoidable conflict of interest between the technology-borrowing developing countries and the technology-generating industrial countries over issues of protection for intellectual property. In the 1960s and early 1970s the United Nations Conference on Trade and Development (UNCTAD) articulated the charges of many developing countries that the international intellectual property system was biased against them. They claimed that it gave monopoly rights to foreign holders of patents and copyrights at the expense of consumers in developing countries (Mody 1990; UNCTAD 1991, pp. 19–23). The explicit justification offered for not subscribing to international patent and copyright conventions was essentially a distributional,

or equity, argument: poor countries ought not be asked to pay for knowledge that would ultimately be part of the universal heritage of mankind.

An alternative rationale could have been given for regimes of weak intellectual property protection in the developing countries' markets, since in most cases the patentholders and copyright holders of the North had not conceived or developed their creations for the markets of these low-income countries. Although it might be debated as a matter of equity that being beneficiaries, the developing countries should pay, or that being poor, they should not be asked to pay, the important point remained that even though they were not expected to pay, it would not much matter from an efficiency standpoint. The world's supply of such innovations would not be much diminished, if at all, by these countries' adoption of a free-riding policy. Of course, this obvious point is obscured when one puts on the convenient blinders of that peculiar application of SLIM in which (Northern) firms undertaking R&D are imagined to be completely specialized in research activities and so not to have in mind any markets for their innovations in the industrial countries.

In the event, Argentina, Brazil, and India, among other countries, passed laws restricting the scope of intellectual property protection, while the international system of protection was somewhat strengthened (during the 1970s) by the expansion of the number of signatories to the Paris Convention for patents and the Berne Conventions for copyright protection. The 1980s witnessed a new departure as the United States, supported by other industrial countries, began pressing for a much greater strengthening of the machinery for enforcing compliance on the part of the signatories to these conventions and using the threat of trade retaliation to induce the holdouts to join. Increasingly, trade sanctions have been invoked by the United States in pursuit of a strengthened international regime of protection for intellectual property rights.

This has once again emerged as an arena of great contention into which I do not have to enter further on this occasion than will be needed to make the following simple observation: there is another face to the North-South conflict over intellectual property protection, and it is one that is too often overlooked. The complementary role of tacit knowledge creates difficulties in successfully transferring codified technological information from industrial to developing countries. Therefore, involving firms from industrial countries through cooperative ventures, technology support, and training contracts will be in the interests of firms in developing countries. (And one should also consider that there will be positive externalities in the form of learning to borrow, as has already been noted.)

Information asymmetries and monitoring difficulties make it virtually impossible to write efficient contracts specifying the transfer of tacit knowledge. It is nevertheless possible to design contracts for the successful implementation of technologies by bundling the provision of assistance (conveying tacit knowledge) together with the licensing of the use of codified information such as patents and copyrights (Arora 1991). But if protection for such property rights is weak in the borrowing country, and if transferees cannot be bound effectively to preserve trade secrets, the originating firm is unlikely to enter such contracts.

The implication is that, where the tacit knowledge components are vital and remain unavailable domestically, the would-be borrowers of technology have an interest in a regime of stronger protection for intellectual property. Such protection could take the form of statutory measures or judicial enforcement of trade secrecy rights.³⁰ This conclusion is distinct from, although not at odds with, the argument (recently advanced by Chin and Grossman 1990) that because the prospect of licensing patents to the South would stimulate a faster rate of innovation in the North, both regions stand to gain from an international regime providing strong protection of intellectual property. Because it focuses on the successful transfer of codified and tacit information regarding innovations *that already have been made* in the North, the foregoing analysis shows that the South would gain even when there was no incremental innovation-inducement effect of extending intellectual property protection into the South's markets.

To complete the suggested reversal of conventional thinking on this matter, it may also be pointed out that strong protection for intellectual property, especially the upholding of broad patent claims and judicial use of injunctions and criminal prosecutions in trade secrecy cases, can stifle innovation. The protection of one firm's property places constraints on the ability of other firms to profitably extend, improve, and integrate that piece of technology into larger, more productive systems. Technological progress is a cumulative and synergistic process carried on by a multiplicity of actors with heterogeneous competencies. As the experience of the West with the institutions of open science has amply shown, rapid disclosure of, and inexpensive access to, the latest research findings is a recipe for rapidly expanding the stock of knowledge. What is needed is to subsidize or otherwise provide incentives for those engaging in the risky business of research.

The familiar view is that instituting strong enforcement for patents, copyrights, and trade secrets involves accepting the tradeoff of static efficiency losses (because of monopoly restrictions on technology utilization) for faster productivity growth through technological innovation. We would do better to couple the familiar view with the obverse proposition. Tolerating weaker and narrower protection of intellectual property rights involves accepting the tradeoff of slower and less widespread diffusion of technological innovation to the developing countries in exchange for a more rapid pace of knowledge accumulation that may eventually become available to all. This would be true at least in the sectors of the industrial economies where R&D project costs are modest enough to permit recovery by firms that have only a brief period of technical advantage over their direct competitors.

30. Of course, one cannot suppose these arguments apply with the same force in situations in which a significant domestic capability for borrowing has been created already, or in branches of industry where, as seems to be the case for many pharmaceuticals, it is a relatively straightforward business to manufacture products with composition and properties that can be fully discovered outside the contexts in which they were designed and developed.

Such reorientations in policy perspectives may be counted among the benefits ascribed to escaping from the mental straitjacket of the SLIM approach to the dynamics of technological change.

REFERENCES

- Arora, Ashish. 1991. "The Transfer of Technological Know-How to Developing Countries: Technology Licensing, Tacit Knowledge, and the Acquisition of Technological Capability." Ph.D. diss. Stanford University, Stanford, Calif.
- Arrow, Kenneth J. 1962. "Economic Welfare and the Allocation of Resources for Inventions." In R. R. Nelson, ed., *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton, N.J.: Princeton University Press.
- . 1969. "Classificatory Notes on the Production and Transmission of Technological Knowledge." *American Economic Review* 59 (May):29–35.
- Arthur, W. B. 1989. "Competing Technologies, Increasing Returns, and Lock-in by Historical Events." *Economic Journal* 99 (March):116–31.
- . 1990. "Positive Feedbacks in Economics." *Scientific American* 262(2):92–99.
- Bender, David. 1986. "Protection of Computer Programs: The Copyright/Trade Secret Interface." *University of Pittsburgh Law Review* 47(906):907–58.
- Chin, Judith C., and Gene M. Grossman. 1990. "Intellectual Property Rights and North-South Trade." In Ronald W. Jones and Anne O. Krueger, eds., *The Political Economy of International Trade*. London: Blackwell.
- Cohen, W. M., and D. A. Levinthal. 1989. "Innovation and Learning: The Two Faces of R&D." *Economic Journal* 99(397):569–96.
- Dasgupta, Partha S., and Paul A. David. 1987. "Information Disclosure and the Economics of Science and Technology." In George Feiwel, ed., *Arrow and the Ascent of Modern Economic Theory*. New York: New York University Press.
- . 1988. "Priority, Secrecy, Patents, and the Socio-Economics of Science and Technology." Center for Economic Policy Research Publication 127. Stanford University, Stanford, Calif. (Forthcoming in the journal *Science in Context*)
- . 1990. "The New Economics of Science." Center for Economic Policy Research—High Technology Impact Program Working Paper. Stanford University, Stanford, Calif.
- Dasgupta, Partha S., and Eric Maskin. 1987. "The Simple Economics of Research Portfolios." *Economic Journal* 97:581–95.
- Dasgupta, Partha S., and Joseph E. Stiglitz. 1980. "Industrial Structure and the Nature of Innovative Activity." *Economic Journal* 90:266–93.
- David, Paul A. 1969. "A Contribution to the Theory of Diffusion." Center for Research in Economic Growth Memorandum 71. Stanford University, Stanford, Calif. (Forthcoming in Paul A. David, *Behind the Diffusion Curve*, Westview, Boulder, Colo.)
- . 1975. *Technical Choice, Innovation and Economic Growth: Essays on American and British Experience in the Nineteenth Century*. Cambridge, U.K.: Cambridge University Press.
- . 1984. "The Reaper and the Robot: The Diffusion of Labor-Saving Machinery in Historical Perspective." Center for Economic Policy Research Publication 2. Stanford University, Stanford, Calif. (Forthcoming in F. M. L. Thompson, ed., *Essays in Honor of Sir John Habakkuk*, Oxford University Press)

- . 1985. "Clio and the Economics of QWERTY." *American Economic Review* 75(2):332–37.
- . 1986. "Technology Diffusion, Public Policy, and Industrial Competitiveness." In Ralph Landau and Nathan Rosenberg, eds., *The Positive Sum Strategy: Harnessing Technology from Economic Growth*. Washington, D.C.: National Academy of Sciences Press.
- . 1987. "Some New Standards for the Economics of Standardization in the Information Age." In Partha Dasgupta and Paul L. Stoneman, eds., *Economic Policy and Technology Performance*. Cambridge, U.K.: Cambridge University Press.
- . 1988. "Path-Dependence: Putting the Past into the Future of Economics." Institute for Mathematical Studies in the Social Sciences Technical Report 533. Stanford University, Stanford, Calif.
- . 1991a. "Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror." In *Technology and Productivity: The Challenge for Economic Policy*. Paris: Organization for Economic Cooperation and Development.
- . 1991b. "Reputation and Agency in the Historical Emergence of the Institutions of 'Open Science.'" Center for Economic Policy Research Publication 261. Stanford University, Stanford, Calif.
- . 1992a. "Intellectual Property Institutions and the Panda's Thumb: Patents, Copyrights, and Trade Secrets in Economic Theory and History." Presented at the National Academy of Sciences Conference on Global Dimensions of Intellectual Property in Science and Technology, Washington D.C., January 8–10, 1992. (Revised as Center for Economic Policy Research Publication 287, Stanford University, Stanford, Calif.)
- . 1992b. "Path-Dependence and Predictability in Dynamic Systems with Local Network Externalities: A Paradigm for Historical Economics." In C. Freeman and D. Foray, eds., *Technology and the Wealth of Nations*. London: Frances Pinter.
- . 1992c. "Historical Economics in the Long-Run: Some Implications of Path-Dependence." In G. D. Snooks, ed., *Historical Analysis in Economics*. London: Routledge.
- David, Paul A., and Shane Greenstein. 1990. "The Economics of Compatibility Standards: An Introduction to Recent Research." *Economics of Innovation and New Technology* 1(1–2):1–32.
- David, Paul A., and Trond E. Olsen. 1986. "Equilibrium Dynamics of Diffusion When Incremental Technological Innovations are Foreseen." *Ricerche Economiche* 40(4): 73–77.
- . 1991. "Technology Adoption, Learning Spillovers, and the Optimal Duration of Patent-Based Monopolies." Center for Economic Policy Research—High Technology Impact Program Working Paper. Stanford University, Stanford, Calif. (Forthcoming in *International Journal of Industrial Organization*)
- David, Paul A., David Mowery, and W. Edward Steinmueller. 1988. "The Economic Analysis of Payoffs for Basic Research: An Examination of the Case of Particle Physics Research." Center for Economic Policy Research Publication 122. Stanford University, Stanford, Calif.
- . 1992. "Analysis of the Economic Payoffs from Basic Research." *Economics of Innovation and New Technology* 2(1):73–90.
- Davies, Stephen. 1979. *The Diffusion of Process Innovations*. London: Cambridge University Press.
- Eisenstein, Elizabeth L. 1980. *The Printing Press as an Agent of Change*. Cambridge, U.K.: Cambridge University Press.
- Fölster, Stephan. 1985. "Government Coordination of Competitive R&D Projects." Ph.D. thesis. Oxford University.

- Goglio, Alessandro. 1991. "'Technology Gap' Theory of International Trade: A Survey." ITP/TEC/28. Geneva: United Nations Conference on Trade and Development.
- Gold, Bela. 1979. *Productivity, Technology and Capital*. Boston, Mass.: Lexington Books.
- . 1981. "Technological Diffusion in Industry. Research Needs and Shortcomings." *Journal of Industrial Economics* (March):247–69.
- Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25 (October):501–22.
- . 1960. "Hybrid Corn and the Economics of Innovation." *Science* 132 (July):275–80.
- Grossman, Gene, and Elhanan Helpman. 1989. "Endogenous Product Cycles." NBER Working Paper 2913. National Bureau for Economic Research, Cambridge, Mass.
- . 1990. "Trade, Innovation, and Growth." *American Economic Review, Papers and Proceedings* 80 (May):86–91.
- Hempel, Carl G. 1966. *The Philosophy of Natural Science*. Englewood Cliffs, N.J.: Prentice-Hall.
- Ireland, Norman, and Paul L. Stoneman. 1986. "Technological Diffusion, Expectations and Welfare." *Oxford Economic Papers*.
- Jager, Milvin F. 1991. *Trade Secrets Law 1*. 1991 rev. New York: Clark Boardman.
- Johnson, Harry G. 1975. "Technological Change and Comparative Advantage: An Advanced Country's Perspective." *Journal of World Trade Law* 9(1):1–14.
- Kaufner, Erich. 1989. *The Economics of the Patent System*. Chur, Switzerland: Harwood Academic.
- Kline, Stephen J., and Nathan Rosenberg. 1986. "An Overview of Innovation." In Ralph Landau and Nathan Rosenberg, eds., *The Positive Sum Strategy: Harnessing Technology for Economic Growth*. Washington, D.C.: National Academy Press.
- Koch, Adrienne, and William Peden, eds. 1972. *The Life and Selected Writings of Thomas Jefferson*. New York: Modern Library.
- Lamb, David, and S. M. Easton. 1984. *Multiple Discovery: The Pattern of Scientific Progress*. Trowbridge, U.K.: Avebury.
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life*. Beverly Hills, Calif.: Sage.
- Machlup, Fritz. 1958. "An Economic Review of the Patent System." Study 15 of the Subcommittee on Patents, Trademarks, and Copyrights of the Senate Committee on the Judiciary. U.S. Senate, 85th Cong., 2d sess. Washington, D.C.: Government Printing Office.
- Mansfield, Edwin. 1968. *Industrial Research and Technological Innovation*. New York: Norton.
- Mansfield, Edwin, Michael Hamburger, John Rapoport, and others. 1971. *Research and Innovation in the Modern Corporation*. New York: Norton.
- Marschack, Jacob. 1971. "Economics of Information Systems." In M. D. Intrilligator, ed., *Frontiers of Quantitative Economics*. Amsterdam: North-Holland.
- Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. N. W. Storer, ed. Chicago, Ill.: University of Chicago Press.
- Mody, Ashoka. 1990. "The New International Environment for Intellectual Property Rights." In Francis W. Rushing and Carole Ganz Brown, eds., *Intellectual Property Rights in Science, Technology, and Economic Performance: International Comparisons*. Boulder, Colo.: Westview.
- Moulton, P. M., Joan Young, and J. J. Eberhardt. 1990. "Technological Advances in Instrumentation: Impacts on R&D Productivity." Battelle Institute Working Paper. Pacific Northwest Laboratory, Richmond, Wash.

- Mowery, David. 1983. "Economic Theory and Government Technology Policy." *Policy Sciences* 16:27-44.
- Nagel, Ernst. 1961. *The Structure of Science*. New York: Columbia University Press.
- National Academy of Sciences. 1989. *Information Technology and the Conduct of Research—The User's View*. Washington, D.C.: National Academy Press.
- Nelson, Richard R. 1959. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67:297-306.
- . 1990. "What Is Public and What Is Private About Technology?" Consortium on Competition and Cooperation Working Paper 90-9. University of California Center for Research in Management. Berkeley, Calif.
- Pavitt, Keith. 1987. "The Objectives of Technology Policy." *Science and Public Policy* 14:182-88.
- . 1992. "Internationalization of Technological Innovation." *Science and Public Policy* 19(2):119-23.
- Plant, Arnold. 1934. "The Economic Theory Concerning Patents for Inventions." *Economica* 1 (February):167-95.
- Polanyi, Michael. 1966. *The Tacit Dimension*. London: Routledge and Kegan Paul.
- Popper, Karl R. 1968. *The Logic of Scientific Discovery*. 2d ed. New York: Harper and Row.
- Rapoport, Antol. 1986. *General Systems Theory: Essential Concepts & Applications*. Cambridge, Mass.: Abacus.
- Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 97:71-102.
- Rosenberg, Nathan. 1972. "Factors Affecting the Diffusion of Technology." *Explorations in Economic History* 9 (Fall):3-34. Second series.
- . 1976. "On Technological Expectations." *Economic Journal* 86:523-35.
- . 1982. "Learning by Using." In Nathan Rosenberg, ed., *Inside the Black Box: Technology and Economics*. New York: Cambridge University Press.
- . 1990. "Why Do Companies Do Basic Research with Their Own Money?" *Research Policy* 19 (Winter):165-74.
- Sagasti, Francisco R. 1972. *A Systems Approach to Science and Technology Policy-Making and Planning*. Organization of American States Studies on Scientific and Technological Development 7. Washington, D.C.
- Sahal, Devendra. 1981. *Patterns of Technological Innovation*. Reading, Mass.: Addison-Wesley.
- Shannon, C. E., and Warren Weaver. 1949. *The Mathematical Theory of Communication*. Urbana, Ill.: Free Press.
- Smith, Adam. 1776/1976. *An Inquiry into the Wealth of Nations*. R. H. Campbell, A. S. Skinner, and W. B. Todd, eds. Oxford, U.K.: Clarendon.
- Steinmueller, W. Edward. 1989. "Four Observations on Intellectual Property and Information Technology." Technology and Productivity Workshop Paper, Stanford University, Department of Economics, Stanford, Calif.
- Stoneman, Paul L. 1983. *The Economic Analysis of Technological Change*. Oxford, U.K.: Oxford University Press.
- Stoneman, Paul L., and Paul A. David. 1987. "Adoption Subsidies vs. Information-Provision as Instruments of Technology Policy." *Economic Journal* 97:142-50.
- Stoneman, Paul, and Norman Ireland. 1983. "The Role of Supply Factors in the Diffusion of New Process Technology." *Economic Journal* 93 (supplement) (March):65-77.
- Teece, David. 1976. *The Multinational Corporation and the Resource Cost of International Technology Transfer*. Cambridge, Mass.: Ballinger.

UNCTAD (United Nations Conference on Trade and Development). 1991. *Transfer and Development of Technology in a Changing World Environment: The Challenge of the 1990s*. TD/B/C.6/153. Geneva.

Wright, Brian D. 1983. "The Economics of Invention Incentives: Patents, Prizes, and Research Contracts." *American Economic Review* 73(4):707-28.

COMMENT ON "KNOWLEDGE, PROPERTY, AND THE SYSTEM DYNAMICS
OF TECHNOLOGICAL CHANGE," BY DAVID

Ashok Desai

Because knowledge is inherently differentiated and nonquantifiable, the economics of science and technology offers limited scope for rigorous modeling and quantification. It has long provided a home for untidy minds. There was a danger that some basic order would be imposed on this area of indiscipline by a sequential model that looked on basic research, applied research, innovation, and diffusion as successive stages. Paul David has tried to demolish this emerging order and restore chaos. I, for one—having thrived on the disorderliness of technology studies—wholly applaud his enterprise.

The paper is wide ranging, and to save time I am going to concentrate on three areas that I think are central to David's analysis. These are the relationship between basic research and applied research, the relationship between innovation and technology diffusion, and the effect of policies and legal arrangements on diffusion, including technology transfers to developing countries.

I. BASIC AND APPLIED RESEARCH

David denies that basic research comes before applied research, or indeed that science is necessary to the development of technology. Rather, he points out, advances in science depend to a considerable extent on technology. Here he stresses advances in instrumentation, which increase the accuracy and productivity of scientific research and hence the efficiency of research and development.

To my mind, however, the basic point is not that applied research influences basic research but that the distinction between basic and applied research is an economic, not a scientific, one. Scientists would have us believe otherwise because basic and applied scientists are different tribes with different goals, ideals, rituals, and customs. But whatever the scientists may say, basic research is not necessarily more basic or deeper or better or prior. It is simply research that is not directed toward commercial gain, whereas applied research must have tangible expectations of commercial gain. What is basic research today may become applied research tomorrow. Conversely, what is applied research

Ashok Desai is chief consultant to the Indian Ministry of Finance.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

today may see its prospects for profits dwindle until only basic researchers are left pursuing it.

Both basic and applied research are subject to externalities, and both justify subsidies. But obviously, subsidies will have a much greater effect on the scale of basic research; applied research can survive without subsidies as long as corporations find it profitable.

Thus the question worth asking is not whether basic research should be subsidized but what the aim of subsidies should be and how they should be structured. There may be political and sentimental reasons for subsidizing basic research. But from an economic point of view, basic research is best looked on as a meadow in which the cows of applied research may feed. Thus the best candidates for subsidization would be the services provided by basic scientific institutions to the technology generators—for instance, the work of scientists or technically trained specialists, lumpy equipment that individual corporations may not be able to utilize fully, and libraries. Among the indicators of performance for basic research, communications criteria—or indicators of codification, in Paul David's terminology—must take precedence. The American custom of rewarding academics on the basis of their written output is imperfect, but it does reward them for the volume of their communication.

II. INNOVATION AND DIFFUSION

On the relationship between innovation and diffusion, David begins with the proposition that owing to difficulties of appropriation, there will be underinvestment in the production of knowledge if competition is unrestricted. He considers three administrative arrangements—patronage, procurement, and property. Patronage involves subsidies to research on the condition that the results will be fully disclosed. Procurement means contracting for certain results of scientific activity with or without insisting on public disclosure. Property is giving inventors exclusive rights of use and transfer over their inventions.

David notes that both patronage and procurement lead to races to establish priority. Actually, this is true of property as well and is indeed characteristic of the innovation process whatever the institutional arrangement. The first-comers earn higher rents irrespective of whether the law protects their property or not. Whatever their final effects, the intentions behind the patronage, procurement, and property systems are different.

- The patronage system seeks to reward research that would not be commercially justified and hence to increase the output of knowledge.
- The procurement system seeks to turn technology known to be possible but uneconomic into products and thereby to expand the realm of the tried and tested.
- The property system seeks to create a market for technology that is distinct from the market for products made with the technology and thereby to

reward the inventor who does not have the resources or inclination to go into production.

The interrelationships between innovation and diffusion are complex. David brings out their complexity and warns against an assumption that faster or easier diffusion is better for innovation. An innovator will not normally transfer his innovation unless he finds transfer more profitable than nontransfer. Hence any legal modification must be based either on a legal concept of ownership that has little to do with economics—and I would suggest that the general position of industrial countries on intellectual property rights is such a legal definition—or on an assumed divergence between the private and the social utility of knowledge, which is a rather woolly difference. I expect many papers but no firm conclusions in this area.

III. TECHNOLOGY TRANSFERS

In his discussion of transfers of knowledge between the North and the South, David makes the important point that much of the knowledge that is formally transferred to developing countries is codified knowledge and that it would be more efficiently used if it were accompanied by tacit knowledge, which is more often transferred by teaching, example, and imitation. I think he is absolutely right. But what is even more important than the transfer of tacit knowledge is the stock of tacit knowledge in developing countries. I note that technologically dynamic economies are characterized by close links between industry and the repositories of tacit knowledge—universities, laboratories, research institutes, and so on. I also note that the most successful developing economies, such as the Republic of Korea and Taiwan (China), have one or two cities where institutions of learning and industries are closely intertwined. This synergy of science and technology within developing economies may be more important for their industrial performance than the links they forge with industrial countries.

COMMENT ON "KNOWLEDGE, PROPERTY, AND THE SYSTEM DYNAMICS
OF TECHNOLOGICAL CHANGE," BY DAVID

Morris Teubal

I would like to comment on Paul David's paper from the perspective of its relevance to a structuralist framework of economic growth—a framework that is increasingly important to developing as well as industrial countries (see Justman and Teubal 1990, 1991). This policy focuses on structural change (interpreted broadly to include new industries, new generic technologies, and a restructuring of important sectors and regions) as a condition for growth. Socially desirable structural change may be difficult to achieve because market forces by themselves may not be adequate to provide the requisite infrastructure and capabilities (see Halperin and Teubal 1991; Justman and Teubal 1992). Governments must assist the process by adopting policies to encourage the development of new capabilities.

Countries face enormous challenges in developing the wide spectrum of capabilities required for conventional and high-technology industries. Cutting-edge capabilities, which are generated by precompetitive, collaborative research and development in such areas as microelectronics, biotechnology, and new materials, differ from routine research and development. The latter are short term, oriented to specific products and processes, and executed by individual firms.

Similarly, technical capacity in conventional industries, which largely relates to the adoption, assimilation, and diffusion of foreign technology, may also require "collective action" over and beyond the efforts of individual firms. For example, successful absorption of a product design capability by a technology center serving the plastics industry may be a precondition for the successful adoption of this capability (or of product design services) by individual firms. Thus a complementary technology policy may be required to effect the structural changes dictated by the liberalization of trade, technology, and foreign investment.

What does Paul David's paper tell us in this context? I would like to discuss four issues reviewed in the paper: the codification of tacit knowledge, standardization in scientific research and in instrumentation, diffusion policies, and the relation of intellectual property to generic research.

Morris Teubal is professor of economics at the Hebrew University of Jerusalem and head of the Industrial Development Policy Group at the Jerusalem Institute.

- *Tacit knowledge.* David suggests that codified tacit knowledge may be an important component in the adoption of foreign technology and is probably a precondition for diffusion. Choices will have to be made, given that codification is largely a fixed cost. His discussion breaks some new ground and is extremely suggestive. Further research should analyze this issue in concrete settings.

- *Research and instrumentation standards.* David points to the existing trend toward standardizing methods of measurement and scientific procedures. This important trend has implications for the transmission of cutting-edge technologies. It tends to promote the diffusion of scientific research and support such new technologies as optronics and lasers, microelectronics, and biotechnology. It also has implications for the configuration of a particular infrastructure—for example, whether it should be centrally developed and housed in a single institution or whether a more decentralized setting is desirable.

- *Diffusion policies.* I am less sure that David's technology policy paradoxes are relevant for developing countries. His analysis showing that stimulation of innovation through research and development may delay diffusion—and with it, productivity growth—seems to be severely limited, for two sets of reasons. The first relates to an open economy that exports goods and imports technology. In this case the innovator resides abroad, beyond the scope of policy in the technology-adopting country. Moreover, research and development within the country may be directed to developing new products for export. This is the case in Israel and, I presume, in other countries as well.

The second set of reasons concerns the mechanisms through which certain kinds of innovations are diffused to small and medium-size enterprises in developing countries. As noted earlier, that may have to begin with the absorption of technology within, say, a technology center. Successful absorption depends as well on markets for technological services, in connection, for example, with product design, quality control, testing, choice of production equipment, and so on. Consultants are important here in bridging the gap between supply and demand (Justman and Teubal 1992). This process is quite different from that implied by David's framework, although his assumptions may be more appropriate to radical innovations in advanced countries.

- *Intellectual property.* David points out the possibility of overinvestment in research and development that results from the "all-or-nothing" payoff structure implied by the patent system and, more generally, the inadequate profile of the research and development projects undertaken. Given the importance of precompetitive development of technology, it is important to point out that some of the technological effort that is important for developing countries may not receive patent protection, specifically that associated with establishing standards and methods of measurement, the properties of materials involved in the new technologies, and other engineering knowledge. This has been pointed out by Nelson, Pack, and Kalachek (1967) and Tassej (1982), among others.

The implication is that society should emphasize alternative methods of promoting these activities, including temporary catalytic subsidies and assistance in coordinating the various actors.

REFERENCES

- Halperin, Ariel, and Morris Teubal. 1991. "Government Policy and Capability Creating Resources in Economic Growth." *Journal of Development Economics* 35 (April):219-41.
- Justman, Moshe, and Morris Teubal. 1990. "The Structuralist Perspective on Economic Growth and Development: Conceptual Foundations and Policy Implications." In Robert Evenson and Gustav Ranis, eds., *Science and Technology: Lessons for Development Policy*. Boulder, Colo.: Westview.
- . 1991. "A Structuralist Perspective on the Role of Technology in Economic Growth and Development." *World Development* 19(9):1167-83.
- . 1992. "Strategic Technology Policy for New Industrial Infrastructure: Creating Capabilities and Building New Markets." Report of the Industrial Development Policy Group. Jerusalem Institute. Israel.
- Nelson, Richard R., Howard Pack, and Edward D. Kalachek. 1967. *Technology, Economic Growth and Public Policy*. Washington, D.C.: Brookings Institution.
- Tassey, Gregory. 1982. "Infra Technologies and the Role of Government." *Technological Forecasting and Social Change* 11:163-180.

Accumulating Technological Capability in Developing Countries

Martin Bell and Keith Pavitt

In the early stages of development the accumulation of technology is influenced by factor endowments and intersectoral linkages. In later stages the level of technological knowledge itself can become a source of comparative advantage, reflected in production know-how, the design of capital goods, and a capacity for reverse engineering and imitative research and development. Evidence shows that firms play a central role in this process and that, contrary to received wisdom, production capacity does not lead automatically to technological capacity in developing countries. Market-related institutions tend to undervalue technological accumulation; it is essential that the development banks do not follow suit in the projects and programs they finance.

In the past few years there has been renewed interest in activities that generate technical change. They have become a central feature in the new trade and growth theories (see, for example, Krugman 1986; Grossman and Helpman 1990; Romer 1990) and have emerged in empirical studies as one of the major factors explaining differences among the industrial countries in growth and trade performance (Fagerberg 1987, 1988; Cantwell 1989). The comfortable assumption that best-practice techniques diffuse quickly and cheaply among countries is no longer so widely held, as is evident in the contrast between the economic performance of Japan and the United States (and as explored, for example, by Abramovitz 1986). "Forging ahead" and "falling behind" are now receiving as much analytical and policy attention as "catching up."

Similarly, a growing body of evidence from the developing countries not only shows considerable international variation in the static efficiency with which apparently "given" technologies are used (see Pack 1987) but also shows considerable differences in two types of dynamic efficiency: first, the intensity with which industrial technologies already used by firms are changed by continuing adaptation, improvement, and development (see Katz 1984; Bell, Ross-Larson, and Westphal 1984; Enos and Park 1988; and Meyer-Stamer and others 1991); and second, the efficiency with which new bases of comparative advantage are created in increasingly technology-intensive industries.

Martin Bell and Keith Pavitt are with the Science Policy Research Unit at the University of Sussex.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

None of this variability in the technological dynamism of economies and firms is likely to come as much of a surprise either to economic historians or to development economists. It shows that, although centrally planned socialism may be dead, history most certainly is not. There remains ample scope for experimentation and debate within market economies about the theoretical framework within which to analyze technical change and about the policies and institutions most likely to promote such change and improve dynamic efficiency.

The theoretical debate is being pursued elsewhere. In particular, the characteristics of technical change in market economies that we describe later in this paper have led some analysts to adopt a heretical evolutionary approach to technical change, emphasizing the central importance of dynamic competition through continuous innovation and imitation, together with disequilibria, uncertainty, learning, and interfirm and intercountry differences in competencies and behavior (see, for example, Freeman 1982; Nelson and Winter 1982; Dosi 1988; Dosi and others 1988; Dosi, Pavitt, and Soete 1992).

In this paper we attempt to contribute to the policy debate on industrial technology in developing countries. Sections I and II deal with the framework for our analysis. We reject the clear distinction between innovation and diffusion, with its implicit assumption that technical change can easily and quickly be promoted through investment in new production capacity. Instead, we distinguish such capacity from the resources and institutions that make up a country's technological capabilities, emphasizing the evidence that the accumulation of technology does not necessarily follow from policies to achieve other objectives; it must become a policy objective in its own right.

Section III summarizes what we know about the nature and determinants of technological accumulation and examines their implications for policies and institutions. We draw on the experiences of the industrial and developing countries as well as of the centrally planned economies to try to distinguish between the characteristics of successful and unsuccessful countries.

Section IV turns to the accumulation of technology and how it relates to technical change, given that technical competence requires the accumulation of skills and institutions that support continuous changes in products, processes, and procedures. In sections V and VI we identify the links between technological capacity and competitive advantage and examine the reasons for the uneven pace of industrial technology in the developing countries, stressing that technology is no longer an automatic by-product of production capacity. We also review the policies and institutions that support the development of indigenous capabilities, with special emphasis on the importance of firms and the role of government policy. While the "market failure" approach is useful in defining government's role in funding education, training, and basic research, it has been unable to grapple with the international differences in dynamic efficiency now emerging in both industrial and developing countries. In some countries market-related institutions appear to undervalue processes and projects of cumulative

learning that are technically complex, increasingly separated from production, and long-term and diffuse in their effects.

The final section concludes that successful technological accumulation depends on (a) the acquisition of foreign technology; (b) investment in education, training, and research; (c) economic incentives for innovation and imitation; (d) continuous growth of demand; and (e) institutions and policies designed to encourage firms to accumulate technology.

I. THE MISLEADING DISTINCTION BETWEEN "INNOVATION" AND "DIFFUSION"

Conventionally, technical change in industry has been seen as a two-step process: first, the development and initial commercialization of significant innovations; and second, the wider application—or diffusion—of those innovations. The former activity is heavily concentrated in the industrial countries and is significant in developing economies only as they approach the international technological frontier. Before that stage, since developing countries are seen as involved primarily in the diffusion of technology or in the choice and adoption of existing technologies, creative innovation and technical change seem irrelevant. From this perspective, the accumulation of technology in industrializing countries is seen as technology that is embodied in production capacity: in other words, in the stock of capital goods and operating know-how required to produce existing goods at the relevant production efficiency frontier.

In fact, diffusion involves more than the acquisition of machinery or product designs and related know-how. It also involves continuing, often incremental, technical change to fit specific situations and to attain higher performance standards. In technologically dynamic situations, these forms of technical change typically involve two stages. First, technology is adopted for incorporation in new production facilities, at which time the original technology may be improved on or adapted for the specific situation, as described by Amsalem (1983). Second, there is a postadoption phase that both raises initial efficiency and modifies the technology to conform to changes in input and product markets. The analysis of learning curves in industrial production shows the economic gains from this continuing improvement, but it has typically obscured the underlying processes (Bell and Scott-Kemmis 1990). These learning curves are generated by continuing paths of creative technical change as described in Hollander's (1965) analysis of cost reductions in Du Pont's rayon plants after the technology was imported from Europe, or in more general analyses of the role of continuous improvement in the competitive success of Japanese firms (see Imai 1986). The significance of this incremental technological dynamism has also been highlighted in a handful of studies of firms in developing countries—for example, in the steel industry in Brazil (Dahlman and Fonseca 1987) and in the petrochemical industry in the Republic of Korea (Enos and Park 1988). More recent studies have emphasized the importance of continuing change in the organizational dimension of production technology (see Hoffman 1989; Meyer-Stamer and others 1991).

Alongside these two stages of technical change there is the continuous accumulation of knowledge and skills in the technology-adopting firms and countries. Initially, firms must accumulate the skills and know-how for operating the new processes at their expected performance standards and for producing the new products to existing specifications. In a second stage firms accumulate the deeper forms of knowledge, skill, and experience required to generate continuing paths of incremental change that both improve on the original performance standards of the technology in use and modify its inputs, outputs, and processes in response to changing input and product markets. At the same time they may also strengthen their capabilities for seeking out and acquiring technology from other firms and economies.

At a third stage firms can build on these capabilities to introduce more substantial technical changes—perhaps incorporating significant improvements into processes already used or into process technology acquired from elsewhere—to modify existing products, produce substitutes, diversify into the production of input materials or equipment, or improve the technologies used by supplier industries. This stage may blur into a fourth in which firms produce the kinds of technical change that are usually thought of as “innovations.”

Thus there is considerable scope for variation in the gains from adopting internationally transferred technology. In particular, the intensity with which the developing countries pursue these interacting processes is likely to influence a range of important performance variables: the efficiency of investment in new production capacity (both the efficiency of input combinations and the level of technical efficiency initially attained); the subsequent rate of total factor productivity growth in existing firms and industries; the competitiveness of their product specifications and designs; the strength of backward and forward linkages to suppliers and customers; the structural change toward more technology-intensive lines of production; and, as the international technology frontier is approached, the ability to successfully enter new product markets.

Because the shortcomings of conventional concepts and terminology make it difficult to see what affects technological development, the literature has begun to focus on the resources required for acquiring, using, and changing technology in industrial production and the processes involved in accumulating those resources. See, for example, Teitel (1982); Katz (1984); Bell, Ross-Larson, and Westphal (1984); Dahlman, Ross-Larson, and Westphal (1987); Enos (1991); and Zahlan (1991); and Lall's (1987, 1990) more elaborate taxonomies of technological capabilities.

II. THE DISTINCTION BETWEEN PRODUCTION CAPACITY AND TECHNOLOGICAL CAPABILITIES

Our approach draws a distinction between two stocks of resources: production capacity and technological capability. The former incorporates the resources used to produce industrial goods at given levels of efficiency and given input combinations: equipment (capital-embodied technology), labor skills

(operating and managerial know-how and experience), product and input specifications, and organizational systems. Technological capability incorporates the additional and distinct resources needed to generate and manage technical change, including skills, knowledge and experience, and institutional structures and linkages. This distinction is important because we are interested in the dynamics of industrialization, and hence in the resources necessary to generate and manage that dynamism.

We also identify two processes: technical change and technological learning (or accumulation). The former encompasses any way in which new technology is incorporated into the production capacity of firms and economies: through the incorporation of new technology in relatively large lumps (such as investment in new or additional production facilities) or through the incorporation of incremental technical changes. Inputs for large technical change projects (capital goods, engineering services, project management services, and so forth) can usually be acquired through market mechanisms, and the technology-using firm can, in principle, play a passive role. But such a "turnkey" approach cannot be used to generate continuing incremental changes in existing facilities; the user of the technology must play an active role and must therefore have the relevant technological capabilities.

Technological learning refers to any process that strengthens these capabilities for generating and managing technical change. These processes are the central focus of the paper. One reason for this emphasis is that the intangible resources required to generate and manage technical change can no longer be considered a marginal adjunct to the resources constituting industry's production capacity. They are becoming quantitatively much more significant, reflecting the rising knowledge-intensity of industrial production. In the industrial countries expenditures by leading industrial companies on research and development are now often larger than the investment in fixed capital (for Japan, see Kodama 1991). At the same time, firms have increased expenditures on human resources (see, for example, Wiggenhorn 1990 or the more general review in Eurich and Boyer 1985).

Alongside this quantitative trend is a less clearly definable qualitative shift toward increased differentiation and specialization in the knowledge resources used by industrial firms. A growing distinction has emerged between the kind of knowledge and skill required to operate given production systems and the kind of knowledge required to change them. As a consequence, explicit investment in acquiring and accumulating knowledge has become a necessary basis for building industry's capacity for generating change. As we note later, there is good reason to expect firms in certain market conditions to underinvest in these technological capabilities.

A second reason for emphasizing the distinction between production capacity and technological capabilities is that it helps to focus discussion about policy. For two hundred years the central policy debates (particularly on trade) about late industrialization have focused on alternative measures for stimulating the accu-

mulation of production capacity. To the extent that the technological dynamism of industry has been considered at all in these debates, it has usually been treated as a more or less automatic by-product of trade policy and investment in production capacity. For those advocating protectionist policies, technological dynamism was often seen as an inherent property of the protected sector; it would emerge more or less automatically given the existence of a manufacturing sector or (later) a capital goods sector. Similarly, but from the opposite policy position, technological dynamism was seen as an almost inevitable consequence of an export orientation or of the competitive pressures generated by liberal trade regimes. Obviously, trade policy matters, but as industry's change-generating resources have become increasingly specialized and complex, it has become less and less likely that trade or other policies concerned with optimizing investment in production capacity will also ensure the most efficient levels of investment in the capacity to generate technological dynamism. The policy debate must explicitly address the problems of accumulating these capabilities.

III. MAJOR FEATURES OF TECHNOLOGICAL ACCUMULATION

Many factors influence the ability to acquire and use technology.

- *Resource inputs.* Even in advanced countries, research is rarely the core activity in accumulating technology. Because the central feature of technology is its complexity, trial and error are central to its operation and improvement. Thus even major innovations require the design, construction, and testing of prototype products and pilot process plants, and expenditure on these development activities far outweighs the funding for research programs. Product design, process, and production engineering without any direct links with research and development are important sources of technical change and are likely to be even more important in developing countries. As Imai (1986) points out, Japan's experience shows that knowledge and skill can be harnessed to drive continuous improvements in production and thereby sustain industrial competitiveness.

- *The role of firms.* Firms are the most important actors in accumulating technology because they learn from developing and operating specific production systems. By concentrating on other institutions, science and technology policy in most developing countries has been misplaced; the objective should be to encourage the development of technological capabilities within firms because market mechanisms alone may be insufficient.

- *Tacit knowledge.* A large part of technology involves tacit—not codified—knowledge: rules of thumb that are acquired only with experience and are embodied in people and institutions. Transfers of such knowledge are neither costless nor quick (see, for example, Rosenberg and Frischtak 1985; Scott-Kemmis and Bell 1988). This has major implications for policymakers when they design national systems of science and technology.

- *Learning by doing and by other methods.* Although the importance of acquiring tacit knowledge through experience highlights the role of learning by

doing, it is far from the whole story. Learning should not be seen as an inevitable by-product of some other activity; it may have to be undertaken as an activity in its own right. One kind of activity is seldom adequate preparation for undertaking a qualitatively different kind of activity. Routine production skills, for instance, may contribute very little to developing the kinds of knowledge and experience that are required to generate and manage technical change (Bell, Scott-Kemmis, and Satyarakwit 1982; Bell 1984). It is therefore important to define what activities contribute most effectively to the accumulation of technology. Given the increasing division of labor within firms in the operation and management of technical change, this will include research and development laboratories, design offices, production engineering, and so forth.

- *Cumulative change.* Technological accumulation tends to be incremental. Given the importance of specific and tacit knowledge, individuals and firms are not capable of learning simultaneously across diverse technological and organizational dimensions. Nor do they simply jump into totally new areas of technology. Instead, they tend to move along particular trajectories in which past learning contributes to particular directions of technical change and experience reinforces the existing stock of knowledge and expertise. These cumulative properties of technological accumulation have three implications. First, differences in technical efficiency between firms and countries arise not only from different factor endowments or from barriers to entry but also from differences in accumulated technological competence (as reflected in the models reported by Soete 1981; Fagerberg 1987, 1988). Second, national competence cannot be changed rapidly (see Pavitt 1988). And third, the rate and composition of technological accumulation influences not only short-term competitive efficiency but also longer-term comparative advantage.

- *Externalities, appropriability, and uncertainty.* There are significant externalities in the accumulation of technology, in the sense that the full benefits are not necessarily appropriated by the firms investing in the technology. Thus in the Republic of Korea the effectiveness of one firm's acquisition of technology for the petrochemical industry was enhanced by drawing on the engineering capabilities of a firm in the refinery industry (Enos and Park 1988). Elsewhere, the engineering and project management capabilities accumulated by a power-generating utility enhanced the efficiency of firms entering the power-engineering and equipment industries (UNCTAD 1985). In these and countless other ways, the contribution made by firms to an economy's overall pool of technology may in principle be little different from that of other institutions more explicitly concerned with education and training. Nevertheless, the two kinds of institutions may not be effective substitutes: the relevant skills and knowledge can only be acquired in firms and through their investments in learning. Precisely because it is impossible to appropriate the full returns to these investments and because the private returns are likely to be uncertain, there is likely to be significant underinvestment by firms.

Table 1. A Technology-Based Classification of Business Firms

Characteristic	Category of firm			
	Supplier-dominated	Scale-intensive	Science-based	Specialized supplier
Typical core sector	Traditional manufacturing (for example, textiles, garments, leather and footwear, wood products)	Bulk materials (steel, glass) Consumer durables Automobiles	Electrical and electronics Chemicals	Capital goods Instruments Software
Size of firm	Small	Large	Large	Small/medium
Type of user	Price sensitive	Price sensitive	Price and performance sensitive	Performance sensitive
Main focus of technological activities	Cost reduction Design for niche markets	Cost reduction and product improvement	Cost reduction and product innovation and improvement	Product innovation and improvement
Main immediate sources of technical change	Suppliers	Process, production and design engineering	Corporate R&D	Design and development
Main channels of imitation and technology transfer	Purchase of equipment and related services	Equipment purchase; know-how licensing; reverse engineering	Reverse engineering; R&D Hiring qualified scientists and engineers	Reverse engineering Learning from advanced users
Main sources of technologi- cal accumulation	Production learning Advisory services	Production learning Suppliers Design	Basic research Production engineering Design	Advanced users
Main directions of techno- logical accumulation	Processes and related equipment Product design Production Organization [Variable]	Process technology and related equipment	Technology-related products	Product development
Main methods of protection against imitation	Nontechnical (marketing, trademarks)	[Upstream] Process secrecy Imitation and learning curve lags	[Concentric] R&D know-how Patents Imitation and learning curve lags	[Concentric] Design know-how Patents Knowledge of users' needs

• *Sectoral differences.* Table 1 examines the acquisition of technology by firms in four categories of technological development, based on studies of the industrial economies. (For more detail see Pavitt 1984; for a recently developed alternative see Malerba 1992.) Each category is reflected in a model of technical change found in economics. It is worth describing them in a little more detail because each category has its distinctive method of acquiring technology, with particular implications for policy.

In *supplier-dominated firms* technical change comes from suppliers of machinery and other production inputs. Technical choices reflect relative factor costs, and opportunities for technological accumulation are mainly improvements and modifications of production methods and associated inputs and occasionally of product design. Most technology is transferred in the form of capital goods and other inputs. Supplier-dominated firms bear some resemblance to those found in the conventional production function.

In *scale-intensive firms* technical change is generated by the design and operation of complex production systems or products, or both. Given the economic advantages of increased scale combined with the complexity of products and production systems, the risks of failure associated with radical change are potentially very costly. Process and product technologies therefore develop incrementally on the basis of operating experience and improvements in components, machinery, and subsystems. Technology accumulates as the capacity to design and build components, machinery, and subsystems and eventually as the capacity to design and build large-scale, complex systems. Technology transfers require the firm to license production and design know-how and related training, in addition to equipment and other inputs. (See Schmookler's 1966 analysis of investment-induced technical change and Arrow's 1962 analysis of learning by doing.)

In *science-based firms* technology emerges mainly from corporate research and development and is heavily dependent on academic research. Fundamental discoveries open major new product markets and invite a wide range of potential applications, including a horizontal search for new and related product markets. Technology imports require more than the purchase of production inputs and the licensing of production know-how; they also require a strong capacity for reverse engineering (that is, analyzing and copying competitors' products), which itself requires research, development, and design activities and trained research scientists and engineers with foreign contacts. This style of technical change is best exemplified by the large innovating firm, as described in Schumpeter (1943).

Specialized supplier firms provide high-performance equipment in the form of components, instruments, or software to advanced user firms (Stigler 1956; Rosenberg 1976). The accumulation of technology takes place through the design, construction, and use of these production inputs. Such firms, which benefit from the operating experience of advanced users—in the form of information, skills, and the identification of possible modifications—continuously

accumulate the skills to match advances in design with user requirements. Typically, given the complexity and interdependence of production processes, the purchaser puts a premium on reliability and performance rather than on price.

We explore next the extent to which these patterns of technological accumulation in the industrial countries help us understand the dynamics of competitive advantage in the developing countries.

IV. TECHNOLOGICAL ACCUMULATION AND COMPETITIVENESS

The mechanisms through which a country's competitiveness changes over time proceed from these patterns. Our framework envisages two extremes. At one, in supplier-dominated sectors (like textiles), the Heckscher-Ohlin assumptions about comparative advantage may hold reasonably well: technology is universally available, and technical choices are made largely on the basis of factor endowments. In such sectors low-wage countries can exploit their comparative advantage, provided their firms are effective in acquiring technology. At the other economic extreme, comparative advantage in high-wage countries is dominated by technological leads and lags in science-based, scale-intensive, and specialized supplier sectors (Soete 1981). What happens in between these extremes? Below we sketch the historical experience of the industrial countries and discuss whether we can anticipate a similar evolution in the developing countries.

Technological Trajectories in Industrial Countries

David (1975) and, in a different tradition, Porter (1990), who have explored the paths of national technological development, note that the accumulation of technology has involved the cumulative acquisition of (largely country-specific) "intangible capital" in the form of personal, organizational, and institutional skills. Changes in international competitiveness evolved along with—and increasingly as a result of—these national technological trajectories.

Three sets of inducement mechanisms seem to have been particularly important: factor endowments, persistent investments with strong intersectoral linkages, and the cumulative mastery of core technologies. Their relative significance has changed over time. In the early stages the directions of technical change in a country or region were strongly influenced by local factor endowments and investment opportunities. At higher levels of development, the accumulation of specific technological skills accounted for the paths of technical change.

- *Factor endowments.* The most obvious local inducement mechanism has been the search to alleviate a relative factor scarcity. Its historical importance in the development of labor-saving techniques is well documented, as is its importance in generating technical responses to differing natural resource endowments. Consider, for example, the effects of fuel prices on the development of automobiles and related technologies in the United States, Europe, and East

Asia. A search for space-saving technologies is said to have been important in early postwar improvements in Japan's methods of mass production. Environmentally related inducement mechanisms may become more important in the future.

- *Investment-led inducements and intersectoral linkages.* Another analytical tradition stresses the importance of investment-induced technical change (Schmookler 1966), and technical linkages and imbalances among firms and sectors (see Carlsson and Henriksson 1991; Justman and Teubal 1991). The exploitation of abundant natural resources is one variant, creating opportunities for local technical change, technological accumulation, and competitiveness in upstream extraction and downstream processing. Witness the effects of abundant natural resources in Canada, the United States, and Scandinavia, where, for example, the abundance of wood has contributed to the competitiveness of wood-processing machinery and, more generally, the development of natural resources has contributed to competitiveness in the capital goods used in these sectors (Patel and Pavitt 1991b). Other important linkages have included those from investment in mass-produced automobiles (and some other consumer durables) to technologies for manufacturing associated capital equipment, and from government investment programs to shipbuilding, railway, and communications sectors in Japan (Nakaoka 1987).

- *Mastery of core technologies.* Cost- and investment-induced mechanisms, however, cannot explain all technology-based competitiveness. For example, Switzerland's competitiveness in marine diesels has little obvious link with any national endowment in maritime resources but can be explained by the engineering competence originally developed in textile machinery. Thus it reflects the cumulative mastery and exploitation of a core technology with multiple potential applications. The core technologies are science-based, and the linkage is horizontal rather than vertical; for example, Switzerland's strength in pharmaceuticals grew out of earlier strength in dyestuffs.

Two features are common to countries that have followed such technology-based trajectories. First, they tend to make technological choices not simply on the basis of expected financial rewards but also with a view to the expected value of learning for the exploitation of future technological opportunities (Pavitt and Patel 1988). Second, in science-based and scale-intensive sectors, country-specific inducement mechanisms and technological advantage are reflected in the technological activities of a number of large domestic firms (Patel and Pavitt 1991a).

Technological Accumulation in Developing Countries

Over the past thirty years developing countries have increased and diversified their production capacity, moving from sectors that are mainly supplier-dominated (such as textiles) into products in which scale-intensive, specialized supplier firms predominate (such as metals and metal products, capital goods, bulk chemicals, and consumer durables) and even into science-based sectors (semiconductors and telecommunications equipment). This experience confirms

that between the "price-sensitive" goods produced with simple technology and low wages and the "performance-sensitive" goods produced with world-frontier technology and high wages, there is an intermediate range of goods in which firms and countries can be competitive in world markets. This is likely to be the case in product groups where, on the supply side, progress toward the technological "best practice" is relatively cheap and where, on the demand side, lower input costs compensate for any distance from such best practice.

The rate of technological accumulation, however, has been very uneven among the world's catching-up countries. In the early 1980s Poznanski (1984) pointed to the poor performance of the (then) centrally planned economies compared with the so-called newly industrialized economies (see also the more recent comparisons by Ray 1991). And within this category East Asian economies have performed better than Latin American countries, particularly in electronics (Riedel 1988; Freeman 1991).

The sectoral trajectories of growth and diversification may be similar to those found in earlier patterns of industrialization: from supplier-dominated to scale-intensive sectors, and from scale-intensive to science-based and specialized-supplier sectors. And abundant natural resources and infrastructural investment programs will continue to be important inducement mechanisms, even though labor scarcity is likely to be less important for technological accumulation than energy and environmental scarcities.

Furthermore, recent historical research on Japan shows at least a superficial resemblance to contemporary experience in Korea (Nakaoka 1987; Odagiri and Goto 1992). Japan's modernization was built on a combination of imported equipment and indigenous capability in metals and machinery. Ties with foreign firms were necessary in the automotive and electrical industries, and reverse engineering was important in railway rolling stock. Nakaoka (1987) and Kim (1984, 1985) argue, however, that it is more difficult for developing countries today to absorb fast-moving technologies that depend on specialized activities (such as corporate research and development, design, and production engineering) rather than on line production experience.

These observations both clarify the significant differences between today's and earlier industrializing countries and explain why some countries have performed better than others.

• *Links between production capacity and technological capabilities.* The central feature in technological accumulation of the industrial countries was the parallel and interacting accumulation of production capacity and technological capability. This was perhaps most evident in the industries where technical change is now supplier-dominated. For example, in the nineteenth century textile mills in the northeast United States developed and produced much of their own machinery. As a consequence, the basis for new areas of competitiveness was present in, and emerged from, old areas of competitiveness.

With growing specialization and vertical disintegration in these increasingly supplier-dominated industries, that pattern is no longer characteristic. Amsden

(1989) has emphasized the small role of Korea's early leading sector—cotton spinning and weaving—as the foundation for more complex industrial activity. Thus automotive production has probably replaced textiles as the major stimulus to developing local technological capacity in capital goods, and telecommunications investments have become a major stimulus to upstream developments in electronics (see Hobday 1990 on Brazil's experience).

- *Access to science-based and scale-intensive technologies.* A distinguishing feature of technology today is that most of the fast-moving fields are in science-based and scale-intensive sectors, making the process of acquiring technology more difficult and demanding. While sources of capital goods technologies—machines, processes, and instruments—are relatively dispersed, with multiple suppliers (Patel and Pavitt 1991a) and every incentive to sell to firms in developing countries (Cooper 1991), the acquisition of these technologies is very complex.

The large firms that dominate these sectors develop and control significant proportions of both their product and process technologies. They are reluctant to lose their competitive advantage and prefer direct investment to licensing when they move abroad (Patel and Pavitt 1991a; Contractor 1985). Contrary to a widely held assumption, technology does not spread among the industrial countries mainly through these channels but rather through reverse engineering and imitative research and development (de Melto, McMullen, and Wills 1980; Levin and others 1987). And in a study of forty-eight product innovations and their U.S. imitations, Mansfield, Schwartz, and Wagner (1981) estimated that the ratio of imitation to innovation times was on average 0.70 and of imitation to innovation costs, 0.65. Only 11 percent of imitation costs were incurred as a result of patent protection (although the share rose to about 30 percent in pharmaceuticals, one sector in which property rights make up a significant proportion of the total cost of the technology).

These characteristics help explain why the accumulation of production capacity in today's industrializing countries has led to a more limited and uneven accumulation of technological capacities than in earlier historical periods: production capacity is a less direct and automatic inducement to the capacity to generate and manage change. Obviously, there has been some accumulation of technological capabilities because the growth and diversification of industrial output could not have been achieved without some competence in acquiring technology from foreign sources (see Lall 1980). But the depth of this competence has been limited when technology is incorporated in new production capacity through turnkey projects and direct foreign investment.

One common response to these circumstances is to argue that it does not matter because industrial latecomers have in the past—and can today—draw on a wealth of available technology and therefore need not undertake the costly investment in capabilities for generating technical change. But this argument is unconvincing. First, it maintains the simplistic distinction between innovation and diffusion that we dismissed earlier. Second, it ignores historical experience.

In the past, latecomers drew very heavily on technology imports, but they also developed their own technological capabilities in association with their expanding production capacities. And third, it ignores the experience of the most successful developing economies. In particular, across a wide range of sectors in Korea and Taiwan (China) firms have accumulated substantial capabilities for generating continuous changes in acquired technologies, for synthesizing diverse elements of technology into new processes and products to replicate imported technologies, and for developing innovations to improve performance (Westphal, Kim, and Dahlman 1985; Enos and Park 1988). These accomplishments have been reflected in the growth of business-financed research and development (Ozawa 1974, 1985; Fukasaku 1986; Amsden 1989; Cohen and Levinthal 1989). Korea's experience in electronics also shows the growing importance of postgraduate education (and subsequent work experience abroad) for local scientists and engineers, both for training and to provide access to the informal international networks that are so important in advanced technologies. In Latin America, India, and China the accumulation of these kinds of technological capabilities has been much more limited, or narrowly focused, and in Africa it has been virtually absent.

V. NATIONAL SYSTEMS FOR ACCUMULATING TECHNOLOGY

Although policy prescriptions based on national systems for technological accumulation are considered unfashionable in an age of globalization, we must point to the local, person- or institution-embodied nature of technological accumulation, the strong influence of national institutions, and the overriding importance of the home country to the technological activities of even the largest and most multinational of firms (Porter 1990; Patel and Pavitt 1991a). While other units of analysis (such as the region) may sometimes be more appropriate, the nation-state is a good starting point. Some progress has been made during the past ten years in describing the essential characteristics of "best-practice" national systems for the accumulation of technology (see, for example, Nelson 1992). The main conclusions can be summarized as follows.

The Importance of Firms

Given the specific, cumulative, and partly tacit nature of technology, most technological learning takes place in firms. Even in the industrially advanced countries, research and development activities are only the tip of the iceberg, since they are a form of accumulation typical of large firms in science-based technologies. In smaller firms technological activities are sometimes part-time and come under the rubric of "design" and "production engineering" departments. Both historical studies and contemporary research on developing countries show that research and development, design, and production engineering often emerge within firms that have been involved in prior activities in quality control and production organization (Katz 1987; Amsden 1989; Mowery and Rosenberg 1989).

There has been a vigorous and productive tradition of applied research to identify the conditions for the successful management of technical change in firms in industrial countries (see, for example, Rothwell 1977; Harris and Mowery 1990). The literature shows that successful implementation at the project level depends on the effective integration of specialized disciplines, functions, and divisions within the firm and on outside links with sources of expertise and with the needs of customers. We know far less about what factors affect a firm's strategies for technological accumulation or about the management of technology in developing countries.

The failure to recognize the firm as the central player in the accumulation of technology has been a major shortcoming of technology policy. The most flagrant examples have been the former U.S.S.R. and other centrally planned economies, where the design and research and development functions were separated (geographically and organizationally) from production units (Hanson and Pavitt 1987). Similar policies have been followed in some other developing countries where government-funded research and development laboratories were established in the mistaken expectation that—without matching technological capabilities in firms themselves—they would produce practically useful results (Desai 1980; Bell 1984). These policies were sometimes buttressed by the theoretically respectable—but empirically inaccurate—assumption that technology emerges from research and development as easily transmissible and applicable information.

Linkages between firms—especially when they involve (often untraded) flows of knowledge and skills (see Lundvall 1988)—are also important sources of technology transfer between science-based and technology-intensive firms and their (often small and specialized) suppliers. The centrally planned economies tended to discourage the emergence of specialized suppliers, with negative effects on the diffusion of technology and the efficiency of the capital stock (Hanson and Pavitt 1987).

Market Structure and Competitive Pressures

Conventional cross-sector studies of the influence of market structure on technological performance in industrial countries are not very revealing because, as recent research has shown, industrial structure is endogenous and is determined by technological opportunity and appropriability (Levin, Cohen, and Mowery 1985; Pavitt, Robson, and Townsend 1987). When both are high, concentration tends to be high, and innovating firms tend to be in such industries as chemicals, electrical equipment, and electronics. When opportunity is high and appropriability is low, innovating firms tend to be small, as in capital goods.

The importance of competition as an incentive for the accumulation of advanced technology emerges from case studies of textile production in Africa (Mytelka 1985), from national studies of competitiveness (Porter 1990), and from statistical studies of large firms (Patel and Pavitt 1992). An almost com-

plete lack of competitive pressure was one reason the centrally planned economies failed to adopt more efficient techniques.

Government and Market Failure

Most governments in market economies have stepped in to correct market failure and influence the rate and direction of technological change through policies that include the adoption of standards for interfaces and networks and the imposition of penalties or restrictions on technologies that have adverse effects on health, safety, and the environment. They might also be persuaded to include policies to improve technology and information flows to small firms in traditional industries and—with greater success—to the agricultural sector, where technology is generated by suppliers and information for users is imperfect.

A major government contribution to technological accumulation is its investment in education and training. It is widely recognized that education policy has a strong influence on the effectiveness with which technologies are assimilated and improved. Thus, literacy is advantageous in supplier-dominated technologies, and higher technical and graduate engineering skills are necessary in scale-intensive and specialized-supplier technologies. Comparative studies traditionally tended to concentrate on differences in educational achievement between industrial and developing countries. More recently, however, significant differences in achievement have been identified that are said to have contributed to differences in the rate of technological accumulation within each of the two categories.

Thus among the industrial countries, differences have become apparent in the levels of education and skills of the two-thirds of the working population who do not receive a qualification in higher education; in particular, workers in Germany, Japan, and their neighboring countries have higher qualifications and skills than those in the Anglo-Saxon countries, and this has measurable effects on productivity and product quality (see, for example, Prais 1981). Similarly, among the developing countries workers from East Asia are better qualified than those from Latin America—in terms of literacy rates, technical secondary education, and graduate engineers—with similar results (Westphal, Kim, and Dahlman 1985).

Although various forms of “doing” are central to the learning processes, so are more explicit kinds of training. In industrializing countries in particular, the potential role of explicit technological training within the firm is especially important, and although there are good grounds for expecting firms to underinvest in this activity, the issue has received remarkably little policy attention. This seems to be largely because it falls between two areas of policy and analysis. Education and training typically focus on the infrastructural institutions that generate human capital for industry (and other areas of the economy), while industrial policy usually concentrates on the accumulation of production capacity through investment in physical capital and on-the-job training. Effective policy intervention to induce firms to invest more substantially in training to create change-generating human resources seems rare.

The Economic Significance of Basic Research

Most analyses of the contribution of academic research to the development of science-based technologies are not very enlightening. Many economists conclude that developing countries do not need academic research to further economic growth since they assume that its benefits come in the form of published information that can easily be obtained elsewhere (see, for example, Vernon 1987). Moreover, sociologists often assume that most basic research in developing countries is unsatisfactory because only a very small proportion is published and cited in the international literature (see, for example, *Scientometrics* 1992).

Both sets of assumptions overlook studies showing that the main economic benefit of basic research is not published information but a supply of scientists and engineers with problem-solving skills, familiarity with research methodologies and instrumentation, and membership in informal and international networks of professional peers (Gibbons and Johnston 1974; Nelson and Levin 1986; Senker and Faulkner 1991). For this reason, the capacity for academic research and postgraduate training is an important component of national policies to promote advanced technologies, even if they are not initially at world levels of academic excellence; imitative learning activities in science are a necessary part of the process of development and should be linked to policies for international academic exchanges.

VI. GOVERNMENTS AND DYNAMIC EFFICIENCY: DO INSTITUTIONS EVALUATE THE FULL ECONOMIC BENEFITS OF LEARNING?

Dealing with market failure far from exhausts the public policy agenda related to technological accumulation—if only for the very good reason that so-called perfect market conditions create no financial incentive for innovation. Governments have therefore established systems for protecting intellectual property rights to reinforce the potential temporary monopoly rents afforded would-be innovators by the natural time lags and costs of imitation.

More relevant to the concerns of this paper have been the persistent differences in technological accumulation among the industrial countries that cannot be explained by different rates of market failure or defense of intellectual property rights. For some analysts, these differences can be attributed entirely to differences in macroeconomic conditions, reflected in rates of inflation, interest, and growth. For others, they reflect deep-seated differences in management, industrial finance, and education and training. Thus there is a view that Germany and Japan, with their strong emphasis on “insider” methods of corporate control, have more effective systems for ensuring their commitment to long-term corporate goals, including technology (Corbett and Mayer 1991). There have also been strong criticisms of systems of corporate management that emphasize hierarchical organization and short-term financial targets rather than research and development, production, and marketing skills (Abernathy and Hayes 1980; Chandler 1989).

At the heart of this debate is the need to devise a system to ensure that market processes fully evaluate and encourage cumulative learning—or technological accumulation (Pavitt and Patel 1988). In addition to training externalities, the very processes of specialization and professionalization are likely to result in underinvestment in learning for two reasons. First, the potential benefits of learning are no longer immediate and obvious (as is, say, a cost-reducing improvement in a process) but rather are in the future and diffuse (as with, for example, an improved capacity to design a specific product). Second, the knowledge required to evaluate accurately the potential value of learning has become increasingly complex and beyond the comprehension of anyone not continuously active in the field.

In this context, conventional project appraisal techniques used by managers and public policymakers are inadequate because they completely neglect the “option value” of path-dependent, irreversible learning (Myers 1984). One possible solution is to treat potential technological investments explicitly as an option value (Mitchell and Hamilton 1988). Another is to ensure that corporate and financial institutions have enough engineers to counteract the dangers of myopic analysis by accountants and business school graduates. Amsden (1989) notes this is true of Korea; Morita (1992) advised that it would be useful in Great Britain.

As we have noted, technological capacity does not follow automatically from the establishment of production capacity. Consequently, policies for technological accumulation cannot be treated as a derivative of, say, industrial or trade policy but must be dealt with explicitly. Deliberate firm-specific investment in technological learning involves short-term cost and risk. In developing countries, where imitative innovation dominates, the main risk is technical failure associated with learning. Nakaoka (1987) points out that the degree of technical risk is closely correlated with the expected degree of learning: modest technical objectives are safe but allow little learning; ambitious technical objectives are risky but can lead to great strides forward. One of the critical skills is identifying the appropriate mix—at each stage of development—of imported and indigenous technology in the following areas: components and parts, machinery, skills, production engineering and management, and product development and commercialization. In Japan, Nakaoka reports, the government provided risk finance, training funds, and a market for the products that were developed. This suggests two areas of government policy beyond those concerned with the provision of skills and technologically competent financial and industrial institutions: providing advantageous terms for funding high-risk technologies, and temporary and focused import substitution to stimulate local demand in the early stages of learning.

VII. CONCLUSION

What we think we know is that these features are associated with the successful accumulation of technology:

- *A substantial inward flow of foreign technology, closely coupled with the rapid development of indigenous capabilities in business firms.* In Japan's modernization, there has been a close association between the volume of technology imports and the rate of indigenous technological accumulation (Ozawa 1974; Oshima 1984). At the other extreme, the formerly centrally planned economies imported foreign technology less frequently, and in smaller magnitudes, than either industrial or developing market economies (Hanson and Pavitt 1987), and the same may have been true for India (Lall 1985). Among industrial countries Germany and Japan have been more successful than the United Kingdom and the United States (Patel and Pavitt 1991c).

- *Investment in education and training.* The work forces in the United States and the United Kingdom are deficient in vocational and technical skills in comparison with those in Germany, Japan, and their neighbors. The formerly centrally planned economies in Central Europe still have a well-educated labor force in the German tradition. In science and mathematics education East Asian countries are now moving ahead of the Anglo-Saxon and some continental European countries (see, for example, Prais 1981; *The Independent* 1992).

- *Incentives for innovation and imitation.* Such arrangements were weak or nonexistent in centrally planned economies. In industrial countries the incentives for innovation include intellectual property rights, in addition to natural imitation lags. The pressure of competition and a favorable macroeconomic climate are also of crucial importance everywhere. On these grounds, East Asia has performed better than Latin America (Katz 1991). The formerly centrally planned economies face particular difficulties.

- *Favorable market conditions.* In addition to the pressures of competition, the prospect of a large and continually expanding product market is likely to encourage technological accumulation, not only in producer firms but also through vertical links to suppliers. Except in the largest countries, such market growth is more likely to be achieved under policies of export promotion than import substitution. In addition, a more equal distribution of income is likely to stimulate domestic demand for capital and durable consumer goods in the early stages of learning (see Nakaoka 1987).

There is still disagreement on the question of whether economic agents and institutions, under certain circumstances, systematically undervalue the economic benefits of path-dependent technological learning in firms (see, for example, Stiglitz 1987). Some analysts argue, for example, that this is the case for bankers and managers in the United States and the United Kingdom, in comparison with their counterparts in Germany and Japan (see, for example, Pavitt and Patel 1988).

It is also argued that effective learning in developing countries requires sector-specific intervention involving preferential finance and various measures of protection (Wade 1990; Kim 1980). Investment in infrastructure is also essential to meet the needs of different categories of technology (see Justman and Teubal 1991).

We still need to know how to establish the conditions for successful learning. With a few notable exceptions (such as Westphal, Kim, and Dahlman 1985; Katz 1987; Amsden 1989), we have few empirical studies on what institutions and policies, including government policy, can be used in developing countries to encourage learning in firms or industries. For this reason we propose a research agenda to answer the following questions:

- What are the characteristics of successful firms in developing countries in terms of technology and dynamic efficiency? How are these characteristics similar to—and different from—those found in industrial countries?
- How does the long-term accumulation of technological capability evolve in successful firms in developing countries? Does it follow the sequence described in this paper? Do imports of capital goods lead to the acquisition of production know-how, then to incremental modifications to products and processes, and finally to the design of capital goods and to imitative research and development?
- How important are interfirm and intersectoral linkages in developing countries? And how important are linkages between scale-intensive firms and their specialized suppliers of capital goods?
- What is the role of government-funded infrastructure (for instance, education and training, basic research, and government-sponsored research)?
- How important for technology are academic research and international training for scientists and engineers?
- How does the behavior of financial institutions (and of senior managers) influence the evaluation of the economic benefits of technology?
- What should the development banks do to ensure that the potential costs and benefits of technological accumulation are properly evaluated in investment projects and programs?

REFERENCES

- Abernathy, William, and Robert Hayes. 1980. "Managing Our Way to Economic Decline." *Harvard Business Review* (July–August):67–77.
- Abramovitz, Moses. 1986. "Catching Up, Forging Ahead and Falling Behind." *Journal of Economic History* 46:385–406.
- Amsalem, Michael A. 1983. *Technology Choice in Developing Countries: The Textile and Pulp and Paper Industries*. Cambridge, Mass.: MIT Press.
- Amsden, Alice. 1989. *Asia's Next Giant: South Korea and Late Industrialization*. Oxford, U.K.: Oxford University Press.
- Arrow, Kenneth. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29:155–73.
- Bell, Martin. 1984. "Learning and the Accumulation of Technological Capacity in Developing Countries." In Martin Fransman and Kenneth King, eds., *Technological Capability in the Third World*. London: Macmillan.
- Bell, Martin, and Donald Scott-Kemmis. 1990. "The Mythology of Learning-by-Doing in World War II Airframe and Ship Production." University of Sussex, Science Policy Research Unit, Brighton, U.K.

- Bell, Martin, Bruce Ross-Larson, and Larry Westphal. 1984. "Assessing the Performance of Infant Industries." *Journal of Development Economics* 16:101-28.
- Bell, Martin, Donald Scott-Kemmis, and Wit Satyarakwit. 1982. "Limited Learning in Infant Industry: A Case Study." In Frances Stewart and Jeffrey James, eds, *The Economics of New Technology in Developing Countries*. London: Frances Pinter.
- Cantwell, John. 1989. *Technological Innovation and Multinational Corporations*. Oxford, U.K.: Blackwell.
- Carlsson, Bo, and Robert Henriksson. 1991. *Development Blocks and Industrial Transformation: The Dahmenian Approach to Economic Development*. Stockholm: Industrial Institute for Economic and Social Research.
- Chandler, Alfred. 1989. *Scale and Scope*. Cambridge, Mass.: Harvard University Press.
- Cohen, Wesley, and Daniel Levinthal. 1989. "Innovation and Learning: The Two Faces of R&D." *Economic Journal* 99:569-96.
- Contractor, Farok. 1985. "Licensing versus Foreign Direct Investment in U.S. Corporate Strategy." In Nathan Rosenberg and Claudio Frischtak, eds., *International Technology Transfer*. New York: Praeger.
- Cooper, Charles. 1991. "Are Innovation Studies on Industrialized Economies Relevant to Technology Policy for Developing Countries?" UNU/INTECH, Maastricht, Netherlands.
- Corbett, Jenny, and Colin Mayer. 1991. *Financial Reform in Eastern Europe*. Discussion Paper 603. London: Centre for Economic Policy Research.
- Dahlman, Carl, and Ferrando Fonseca. 1987. "From Technological Dependence to Technological Development: The Case of the USIMINAS Steel Plant in Brazil." In Jorge Katz, ed., *Technology Generation in Latin American Manufacturing Industries*. London: Macmillan.
- Dahlman, Carl, Bruce Ross-Larson, and Larry E. Westphal. 1987. "Managing Technological Development: Lessons from Newly Industrializing Countries." *World Development* 15:759-75.
- David, Paul. 1975. *Technical Choice, Innovation and Economic Growth*. Cambridge, U.K.: Cambridge University Press.
- de Melto, Dennis, Kathryn McMullen, and Russel Wills. 1980. *Innovation and Technological Change in Five Canadian Industries*. Discussion Paper 176. Ottawa: Economic Council of Canada.
- Desai, Ashok. 1980. "The Origin and Direction of Industrial R&D in India." *Research Policy* 9:74-96.
- Dosi, Giovanni. 1988. "Sources, Procedures, and Microeconomic Effects of Innovation." *Journal of Economic Literature* 36:1120-71.
- Dosi, Giovanni, Christopher Freeman, Richard Nelson, Gerry Silverberg, and Luc Soete, eds. 1988. *Technical Change and Economic Theory*. London: Frances Pinter.
- Dosi, Giovanni, Keith Pavitt, and Luc Soete. 1992. *The Economics of Technical Change and International Trade*. London: Harvester-Wheatsheaf.
- Enos, John. 1991. *The Creation of Technological Capability in Developing Countries*. London and New York: Frances Pinter.
- Enos, John, and Woo-Hee Park. 1988. *The Adoption and Diffusion of Imported Technology: The Case of Korea*. London: Croom Helm.
- Eurich, Nell, and Ernest Boyer. 1985. *Corporate Classrooms: The Learning Business*. Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching and Princeton University Press.
- Fagerberg, Jan. 1987. "A Technology Gap Approach to Why Growth Rates Differ." *Research Policy* 16:87-99.

- . 1988. "International Competitiveness." *Economic Journal* 98:355-74.
- Freeman, Christopher. 1982. *The Economics of Industrial Innovation*. London: Frances Pinter.
- . 1991. "Catching Up in World Growth and World Trade." University of Sussex, Science Policy Research Unit, Brighton, U.K.
- Fukasaku, Yukiko. 1986. "Technology Imports and R&D at Mitsubishi Nagasaki Shipyard in the Pre-War Period." *Bonner Zeitschrift für Japanologie* 8:77-90.
- Gibbons, Michael, and Ron Johnston. 1974. "The Roles of Science in Technological Innovation." *Research Policy* 3:220-42.
- Grossman, Gene, and Elhanan Helpman. 1990. "Trade, Innovation and Growth." *American Economic Review* 80:796-815.
- Hanson, Philip, and Keith Pavitt. 1987. *The Comparative Economics of Research, Development and Innovation in East and West: A Survey*. Chur, Switzerland: Harwood Academic.
- Harris, Robert, and David Mowery. 1990. "Strategies for Innovation: An Overview." *California Management Review* 32:7-16.
- Hobday, Michael. 1990. *Telecommunications in Developing Countries: The Challenge from Brazil*. London: Routledge.
- Hoffman, Kurt. 1989. "Technological Advance and Organizational Innovation in the Engineering Industries." Industry Series Paper 4. World Bank, Industry and Energy Department, Washington, D.C.
- Hollander, Samuel. 1965. *The Sources of Increased Efficiency: A Study of Du Pont Rayon Plants*. Cambridge, Mass.: MIT Press.
- Imai, Masaaki. 1986. *Kaizen: The Key to Japan's Competitive Success*. New York: McGraw-Hill.
- The Independent*. 1992. "World's Best Leave British Children Trailing in Science." March 13:3.
- Justman, Moshe, and Morris Teubal. 1991. "A Structuralist Perspective on the Role of Technology in Economic Growth and Development." *World Development* 19:1167-83.
- Katz, Jorge. 1984. "Domestic Technological Innovation and Dynamic Comparative Advantage." *Journal of Development Economics* 16:13-37.
- . 1991. "Industrial Restructuring, Public Expenditure and Social Equity: The Experience of Argentina." *Science and Public Policy* 18:375-78.
- Katz, Jorge, ed. 1987. *Technology Generation in Latin American Manufacturing Industries*. London: Macmillan.
- Kim, Linsu. 1980. "Stages of Development of Industrial Technology in a Developing Country: A Model." *Research Policy* 9:254-77.
- Kim, Yonug-Ho. 1984. "Kankoku no Yushutsushiko-gata Kogyoka no Junkan-makanizumu" [The mechanism of business cycles in Korea's export-oriented economy]. *Kikan Keizai-kenkyu* 7:86-22.
- . 1985. "Shuhenbu no Gijutsu-nijugyappa to Gijutsu-tsuseikikeiro" [The double technological gap and the process of technological catching-up in the periphery]. *Kikan Keizai-kenkyu* 8:44-66.
- Kodama, Fumio. 1991. *Analyzing Japanese High Technologies*. London: Frances Pinter.
- Krugman, Paul. 1986. *Strategic Trade Policy and the New International Economics*. Cambridge, Mass.: MIT Press.
- Lall, Sanjaya. 1980. "Developing Countries as Exporters of Industrial Technology." *Research Policy* 9:24-51.

- . 1985. "Trade in Technology by a Slowly Industrializing Country." In Nathan Rosenberg and Claudio Frischtak, eds., *International Technology Transfer*. New York: Praeger.
- . 1987. *Learning to Industrialize: The Acquisition of Technological Capability by India*. London: Macmillan.
- . 1990. *Building Industrial Competitiveness in Developing Countries*. Paris: Organization for Economic Cooperation and Development, Development Centre.
- Levin, Richard, Wesley Cohen, and David Mowery. 1985. "R&D, Appropriability, Opportunity and Market Structure: New Evidence on the Schumpeterian Hypothesis." *American Economic Review* 75:20–24.
- Levin, Richard, Alvin Klevorick, Richard Nelson, and Sidney Winter. 1987. "Appropriating the Returns from Industrial Research and Development." *Brookings Papers on Economic Activity* 3:783–820.
- Lundvall, Bengt-Ake. 1988. "Innovation as an Interactive Process: From User-Producer Interaction to the National System of Innovation." In Giovanni Dosi, Christopher Freeman, Richard Nelson, Gerry Silverberg, and Luc Soete, eds., *Technical Change and Economic Theory*. London: Frances Pinter.
- Malerba, Franco. 1992. "Learning by Firms and Incremental Technical Change" *Economic Journal* 102:845–59.
- Mansfield, Edwin, Mark Schwartz, and Samuel Wagner. 1981. "Imitation Costs and Patents: an Empirical Study." *Economic Journal* 91:907–18.
- Meyer-Stamer, Jong, Christopher Rauh, Hady Riad, Sabine Schmitt, and Tilman Welte. 1991. *Comprehensive Modernization on the Shop Floor: A Case Study on the Brazilian Machinery Industry*. Berlin: German Development Institute.
- Mitchell, G., and W. Hamilton. 1988. "Managing R&D as a Strategic Option." *Research-Technology Management* 31:15–22.
- Morita, Akio. 1992. "S Does Not Equal T, and T Does Not Equal I." Innovation Lecture, Royal Society, London, February 6.
- Mowery, David, and Nathan Rosenberg. 1989. *Technology and the Pursuit of Economic Growth*. Cambridge, U.K.: Cambridge University Press.
- Myers, Stewart. 1984. "Finance Theory and Finance Strategy." *Interfaces* 14:126–37.
- Mytelka, Lynn. 1985. "Stimulating Effective Technology Transfer: The Case of Textiles in Africa." In Nathan Rosenberg and Claudio Frischtak, eds., *International Technology Transfer*. New York: Praeger.
- Nakaoka, Tetsuro. 1987. "On Technological Leaps of Japan as a Developing Country." *Osaka City University Economic Review* 22:1–25.
- Nelson, Richard, ed. 1992. *National Systems Supporting Technical Advance in Industry*. Oxford, U.K.: Oxford University Press.
- Nelson, Richard, and Richard Levin. 1986. "The Influence of Science, University Research and Technical Societies on Industrial R&D and Technical Advance." Policy Discussion Paper Series 3. Yale University, Research Program on Technology Change, New Haven, Conn.
- Nelson, Richard, and Sidney Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Odagiri, Hiroyuki, and Akira Goto. 1992. "Technology and Industrial Development in Japan." Tsukuba University, Japan.
- Oshima, Koichi. 1984. "Technological Innovation and Industrial Research in Japan." *Research Policy* 13:285–301.
- Ozawa, Terutomo. 1974. *Japan's Technological Challenge to the West, 1950–1974: Motivation and Accomplishment*. Cambridge, Mass.: MIT Press.

- . 1985. "Macroeconomic Factors Affecting Japan's Technology Inflows and Outflows: The Postwar Experience." In Nathan Rosenberg and Claudio Frischtak, eds., *International Technology Transfer*. New York: Praeger.
- Pack, Howard. 1987. *Productivity, Technology, and Industrial Development: A Case Study in Textiles*. New York: Oxford University Press.
- Patel, Parimal, and Keith Pavitt. 1991a. "Large Firms in the Production of the World's Technology: An Important Case of 'Non-globalization.'" *Journal of International Business Studies* 22:1–21.
- . 1991b. "The Limited Importance of Large Firms in Canadian Technological Activities." *Foreign Investment, Technology and Economic Growth*. Calgary, Canada: University of Calgary Press.
- . 1991c. "Europe's Technological Performance." In Christopher Freeman, Margaret Sharp, and William B. Walker, eds., *Technology and the Future of Europe: Global Competition and the Environment in the 1990s*. London: Frances Pinter.
- . 1992. "The Innovative Performance of the World's Largest Firms: Some New Evidence." *Economics of Innovation and New Technology* 2:91–102.
- Pavitt, Keith. 1984. "Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory." *Research Policy* 13:343–73.
- . 1988. "International Patterns of Technological Accumulation." In Neil Hood and Jan-Erik Vahlne, eds., *Strategies for Global Competition*. London: Croom Helm.
- Pavitt, Keith, and Parimal Patel. 1988. "The International Distribution and Determinants of Technological Activities." *Oxford Review of Economic Policy* 4:35–55.
- Pavitt, Keith, Michael Robson, and Joe Townsend. 1987. "The Size Distribution of Innovating Firms in the UK: 1945–1983." *Journal of Industrial Economics* 35:297–316.
- Porter, Michael. 1990. *The Competitive Advantage of Nations*. London: Macmillan.
- Poznanski, Kazimierz. 1984. "Competition between Eastern Europe and Developing Countries in the Western Market for Manufactured Goods." Evidence prepared for the Joint Economic Committee of the U.S. Congress, Washington, D.C.
- Prais, Sig. 1981. "Vocational Qualifications of the Labor Force in Britain and Germany." *National Institute Economic Review* 98:47–59.
- Ray, George. 1991. *Innovation and Technical Change in Eastern Europe*. London: National Institute for Economic and Social Research.
- Riedel, James. 1988. "Economic Development in East Asia: Doing What Comes Naturally?" In Helen Hughes, ed., *Achieving Industrialization in East Asia*. Cambridge, U.K.: Cambridge University Press.
- Romer, Paul. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98:71–102.
- Rosenberg, Nathan. 1976. *Perspectives on Technology*. Cambridge, U.K.: Cambridge University Press.
- Rosenberg, Nathan, and Claudio Frischtak, eds. 1985. *International Technology Transfer*. New York: Praeger.
- Rothwell, Roy. 1977. "The Characteristics of Successful Innovators and Technically Progressive Firms." *R&D Management* 7:191–206.
- Schmookler, Jacob. 1966. *Invention and Economic Growth*. Cambridge, Mass.: Harvard University Press.
- Schumpeter, Joseph. 1943. *Capitalism, Socialism and Democracy*. New York: Harper and Row.
- Scientometrics*. 1992. "International Conference on Science Indicators for Developing Countries." 23 (1) (special issue).

- Scott-Kemmis, Donald, and Martin Bell. 1988. "Technological Dynamism and the Technological Content of Collaboration: Are Indian Firms Missing Opportunities?" In Ashok Desai, ed., *Technology Absorption in Indian Industry*. New Delhi: Wiley Eastern.
- Senker, Jackie, and Wendy Faulkner. 1991. "Industrial Use of Public Sector Research in Advanced Technologies." University of Sussex, Science Policy Research Unit, Brighton, U.K.
- Soete, Luc. 1981. "A General Test of Technological Gap Trade Theory." *Weltwirtschaftliches Archiv* 117:638-66.
- Stigler, George. 1956. "Industrial Organization and Economic Progress." In Lawrence White, ed., *The State of the Social Sciences*. Chicago, Ill.: University of Chicago Press.
- Stiglitz, Joseph. 1987. "Learning to Learn, Localized Learning and Technological Progress." In Partha Dasgupta and Paul Stoneman, eds., *Economic Policy and Technological Performance*. Cambridge, U.K.: Cambridge University Press.
- Teitel, Simon. 1982. "Skills and Information Requirements of Industrial Technologies: On the Use of Engineers as a Proxy." In Moshe Syrquin and Simon Teitel, eds., *Trade, Stability, Technology and Equity in Latin America*. New York: Academic.
- UNCTAD (United Nations Conference on Trade and Development). 1985. *Technology Issues in the Energy Sector of Developing Countries: Technological Impact of the Public Procurement Policy—The Experience of the Power Plant Sector in the Republic of Korea*. TT/60. Geneva.
- Vernon, Raymond. 1987. "Key Factors in the Application of Industrial Technology in Developing Countries." Harvard University, Cambridge, Mass.
- Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton, N.J.: Princeton University Press.
- Westphal, Larry, Linsu Kim, and Carl Dahlman. 1985. "Reflections on the Republic of Korea's Acquisition of Technological Capability." In Nathan Rosenberg and Claudio Frischtak, eds., *International Technology Transfer*. New York: Praeger.
- Wiggenhorn, William. 1990. "Motorola U: When Training Becomes an Education." *Harvard Business Review* (July-August):71-83.
- Zahlan, Anthony. 1991. *Acquiring Technological Capacity: A Study of Arab Consulting and Contracting Firms*. London: Macmillan.

Technology Gaps between Industrial and Developing Countries: Are There Dividends for Latecomers?

Howard Pack

Do poorer developing countries benefit from relative backwardness? Was their growth in the years 1960–85 greater than would be predicted on the basis of the accumulation of capital and education? The cross-sectional evidence for convergence is not consistent with time-series estimates of rates of growth of total factor productivity. Moreover, some of the existing growth of productivity in developing countries is attributable to the reallocation of labor from agriculture to manufacturing and services, which have higher marginal productivity. Nevertheless, a few economies in Asia have benefited from their ability to shift to higher industrial production functions. Their historical experience, including the interaction between general economic policy and acquisition of technology, sheds light on the lessons for realizing benefits from industrial backwardness.

There has recently been a resurgence of interest in the sources of differences in levels of per capita income across countries and in whether such differences are decreasing over time. The concept that poorer countries could take advantage of the benefits of relative backwardness was put forward by Gerschenkron (1962).

The measurement and explanation of the varying levels of manufacturing productivity in developing countries has received considerable attention (Clague 1970, 1991; Hirschman 1957; Pack 1984, 1987; Teitel 1981). Concern with economywide productivity levels increased following a series of papers on the convergence of per capita income levels (Barro 1991; Baumol 1986; De Long 1988; Dollar 1992; Mankiw, Romer, and Weil 1990). Krueger (1969) analyzed the extent to which human capital could explain differences in income levels and was a precursor of many of these studies.

Part of the resurgence in interest stems from the research of endogenous growth theorists (Lucas 1988; Romer 1986), who interpret the neoclassical Solow-Swan model as implying convergence. In the view of the endogenous

Howard Pack is in the departments of City and Regional Planning, Economics, and Public Policy and Management at the University of Pennsylvania. The author has benefited from comments by Carl Dahlman, Mark Gersovitz, Gregory Ingram, Janet Rothenberg Pack, Gustav Ranis, Shekhar Shah, and Lawrence Summers.

growth theory, economies at different levels of per capita income will not converge to the same income levels. Nevertheless, the failure of economies to converge could be attributable to many factors other than the external economies suggested by early versions of endogenous growth theory.¹

The recent literature on convergence finds that countries which were among the poorest in 1960 have had more rapid growth in per capita income than would have been expected on the basis of standard factors such as capital deepening. This research employs a one-sector economywide model and hence suppresses important aspects of the development process. In particular, for the lowest-income countries, agriculture is the major sector and as late as 1960 employed roughly 77 percent of the labor force. Historically, a key aspect of the development process has been the sectoral transformation of economies in which agriculture's share of the labor force declined from perhaps 90 percent to 6 percent or less. A potential source of growth is the transfer of labor away from agriculture to the industrial and services sectors, which exhibit greater marginal productivity.² Such differences in marginal productivity are the core of the dualism literature, which emphasizes market imperfections that result in different factor intensities across sectors.

The large absolute size of agriculture implies that reallocation of labor away from it, as well as the growth of its total factor productivity (TFP), could be significant sources of growth of per capita income in developing countries. Yet the most recent evidence suggests that in the period 1960–80 growth of agricultural TFP in the developing countries was not as rapid as in the industrial countries (Lau and Yotopoulos 1989). This fact casts doubt on the case for convergence.

Once the productivity gains from reallocation are exhausted, continuing rapid growth in per capita income requires fast productivity growth in the industrial sector. Only a few developing economies, mainly in Asia, have achieved this. Their rapid growth and convergence toward the per capita income of more advanced countries was catalyzed by their ability to assimilate technology and shift the production function along which they operate. This change was facilitated by higher levels of human capital, but many other factors were involved.

In section I of this paper, I examine the evidence from cross-country studies of convergence and compare it with the results of time-series studies of TFP growth in individual sectors. I then derive the implications for the cross-sectional evidence offered to support the convergence hypothesis of the observed reallocation of labor over time from agriculture to industry and services in developing countries. A related issue, considered in section II, is the usefulness of the implicit

1. Mankiw, Romer, and Weil (1990) demonstrate that much of the variation in per capita income levels can be explained by explicitly recognizing differences in saving rate, population growth rates, and the rate of technological change.

2. Sectoral reallocation has been an important source of growth in industrial countries as well. See Denison (1985).

view that developing countries will *automatically* benefit from their ability to borrow the technology of industrial countries and that real per capita income in developing and industrial countries will converge. A question arises concerning the mechanisms for achieving continuing rapid growth after the benefits of reallocation have been largely exhausted; this requires the realization of high growth rates of TFP *within* sectors. In section III, I concentrate on the manufacturing sector, although the evidence on agriculture is covered briefly.

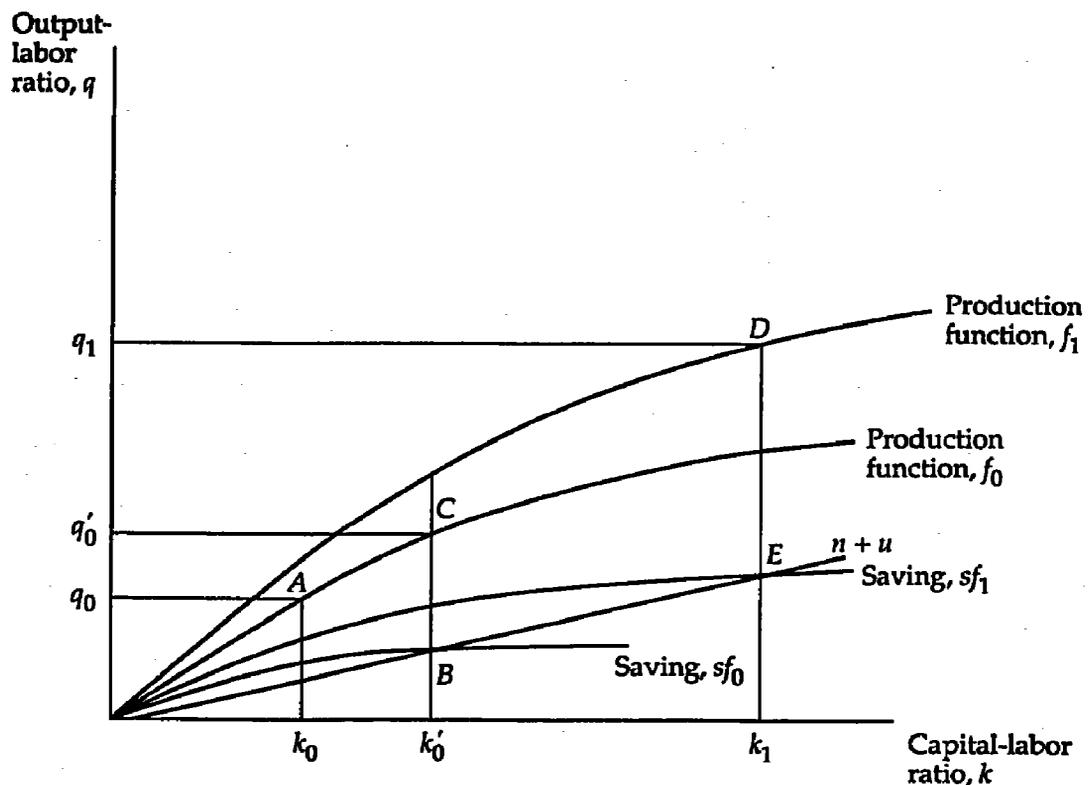
Countries appear to have two choices once they rely on increasing intrasectoral TFP for further per capita income growth. The first choice is based simply on getting the prices right. It may be compared with an escalator going to the fifth floor of a department store—a steady ascension in which the passing scene changes slowly. The second route may be envisioned as an elevator that whisks one to an unrecognizable display of new products on the fifteenth floor in a short time. The first implies a rate of growth of perhaps 1.5 percent per capita a year; the second, experienced mainly in a few Asian economies, implies a growth rate of 5 percent. The second mode of development is analyzed in a brief discussion of the Republic of Korea and Taiwan (China) in section IV. Section V presents conclusions.

I. TECHNOLOGY GAPS AND CONVERGENCE WITHIN A NEOCLASSICAL GROWTH FRAMEWORK

Much of the recent literature on convergence can conveniently be recast in terms of the Solow model shown in figure 1.³ The view of the convergence argument is that nations initially, at point *A* in figure 1, with capital-labor and output-labor ratios k_0 and q_0 , will move along the production function f_0 as they increase their capital-labor ratios. They take advantage of the readily available new techniques to the right of k_0 —"technique" being used here to denote a specific capital-labor ratio. One interpretation of the convergence argument is that if investment rates are equal across countries, as a result of domestic saving or international capital movements, all countries will move toward *C*. Equilibrium is determined at *B*, where saving, sf_0 , equals the requirement for new investment shown by $n + u$, where n is the rate of labor force growth and u is the rate of labor-augmenting technical change.

Countries starting to the left of point *B* should grow more rapidly than those already at it. They do not have to incur expenses for research or the commercialization of the techniques to the right of k_0 . These more advanced techniques are costlessly available to initially poor economies. The assumption is that there exists an international best-practice production function along which nations (and the industries and firms constituting them) can move. Output per effective

3. Mankiw, Romer, and Weil (1990) also employ the Solow model. Proponents of endogenous growth theory whose analysis predicts no diminishing returns to physical and human capital would not accept the depiction of the economy's production function. Since convergence may be interpreted to imply diminishing returns, the production function $Q = AK$ used in endogenous growth theories is not employed.

Figure 1. *The Neoclassical Growth Model*

Note: The investment required to maintain a constant capital-labor ratio is $n + u$.

worker will increase as a result of capital accumulation and will converge toward q'_0 . As long as labor-augmenting technical change is taking place at the rate of u per year, income per unit of labor increases at this rate. The Gerschenkron view, as interpreted by the convergence literature, does not promise any added fillip to the growth rate in per capita income above that attributable to capital accumulation and labor-augmenting technical change as a nation moves toward the steady state at point C. If the advanced techniques did not exist, expenses would have to be incurred to invent and commercialize them, and the move to the more capital-intensive technologies would yield smaller net benefits.

An alternate interpretation of Gerschenkron would be that industrial countries operate along the production function f_1 and that the developing countries operate initially on f_0 but have the opportunity to move toward f_1 and to shift from technologies that have lower TFP at any given capital-labor ratio to ones that have greater TFP. As a result of such a shift, the intermediate-term growth in per capita income would include the impact of capital-deepening and technical change. Technical change would now include the effect of introducing the more productive technology, f_1 , as well as the existing rate of labor-augmenting technical change. Convergence is the move from A to D. This second interpretation is empirically relevant given the large number of studies that find different levels

of TFP among countries at the sectoral and firm level.⁴ A rapid shift in the production function provides the economic engine for the rapid elevator experienced in the Asian economies.

Some Problems

Barro and others who have recently tested convergence theories employ the first model, which involves movements along f_0 , to explain cross-country growth patterns. There are two problems with these studies. First, the cross-country results are widely at variance with existing estimates of TFP growth for the developing countries based on time series.⁵ Second, the use of economywide production functions ignores potential sectoral reallocation effects.

Economywide time-series TFP growth rates and convergence. Dollar (1992) presents the most careful study of convergence and the only one to address explicitly questions that arise in the context of development analysis. I will utilize his specification, which is similar to but more complete than others. The cross-country estimating equation he employs is

$$(1) \quad GDPG = f(ED, INV, DIS, VAR, GDP60)$$

where *GDPG* is the average rate of real per capita growth of gross domestic product (GDP), using Summers-Heston (1988) measures, from 1960 to 1985; *ED* is primary school enrollment in 1960; *INV* is the average share of investment in GDP over the 1960–85 period; *DIS* is distortion in the real exchange rate (an attempt to measure outward orientation); *VAR* is the coefficient of variation of *DIS*;⁶ and *GDP60* is per capita income in 1960 using the Summers-Heston measures. Dollar finds that after allowing for the first four determinants of growth, countries that were poor in 1960 are able to close the gap in per capita income by more than would be predicted—the coefficient of $\log(GDP60)$ is negative and significant.

The underlying assumption of the model is that an internationally accessible cross-country production function exists in which growth in income per labor-year is explained by investment rates and a measure of the services provided by

4. For estimates at the sectoral level, see Arrow and others (1961); Daniels (1969); Pack (1984); and Clague (1991). For firm-level figures, see Pack (1987). Analysis of best-practice frontiers also indicate that industries in many developing countries are off the international frontier (see, for example, Nishimizu and Page 1987).

5. For a review of estimates of TFP for manufacturing see Pack (1988). Chenery, Robinson, and Syrquin (1986) review economywide estimates based on time series through the late 1970s. Elias (1990) provides the most recent comprehensive set of economywide TFP estimates using the Summers-Heston data.

6. Barro (1991), employing a similar model, finds he cannot explain the slow growth of Latin American and African countries and employs a dummy. Dollar, by using *DIS* and *VAR*, attempts to introduce a variable that captures the well-documented policy distortions which have contributed to such weak performance.

Table 1. *The Impact of Reallocation of Labor across Sectors*

Country group	Share of labor force in agriculture		Marginal product of labor in agriculture relative to economywide marginal product of labor, 1960 ^a	Growth in output caused by sectoral reallocation of labor ^b	Economywide TFP growth (A*), 1950-85 (5)	Net TFP growth, (5)-(4) (6)
	1960 (1)	1980 (2)				
Low-income	0.77	0.72	0.44	0.04	0.30	0.26
Lower-middle-income	0.49	0.30	0.54	0.07	1.30	1.23
Upper-middle-income	0.71	0.56	0.49	0.04	1.70	1.66
Industrial	0.18	0.06	0.42	-0.01	1.80	1.81

a. The average product of labor was converted to the marginal product of labor using a Cobb-Douglas production function.

b. This is the growth in TFP attributable to the shift in labor among sectors, calculated using the second part of equation 2.

Source: For columns 1 and 2, World Bank (1984); column 3 calculated on the basis of data from United Nations (various years) and estimated agricultural output elasticity for labor from Kawagoe, Hayami, and Ruttan (1985); for column 5, Elias (1990).

educated labor.⁷ In Dollar's analysis these conventional production function variables are supplemented by two variables: exports, as a proxy for the knowledge obtained from entering into international trade (which shifts the production function from f_0 to f_1), and GDP in 1960, as a measure of the techniques (to the right of k_0 in figure 1) to be borrowed from abroad. The latter may be interpreted as similar to the gains from embodied foreign technological progress.

The coefficient of $GDP60$ in equation 1 is a measure of the ability to move quickly along f_0 as a result of initial backwardness. This shift should appear as higher rates of TFP growth in time-series studies of individual countries. Countries that had low incomes in 1960 should exhibit higher rates of TFP growth than richer nations. Elias (1990, table 3) calculates the economywide TFP growth rates for four groups of countries for 1950–85. These growth rates (shown in table 1, column 5) do not confirm a pattern of convergence. To the contrary, they show that poorer countries grew more slowly over this period. Moreover, these figures overstate the extent of the shift in the production function because they do not consider the impact of sectoral reallocation.

Sectoral reallocation. A well-documented empirical regularity in cross-country studies is the structural transformation of economies as per capita income increases; agriculture declines in relative importance compared with industry and services (Kuznets 1959; Chenery 1960; Syrquin 1984). In addition, Kuznets emphasized the large intersectoral discrepancies in the marginal product of labor, reflecting various factor market imperfections. Substantial evidence exists that capital-labor ratios are greater in industry and services than in the agricultural sector, implying the probability of a greater marginal product of labor in the two urban-based sectors than in agriculture.⁸ At the same time, the marginal product of capital in agriculture may be greater than that in industry and services. Thus the standard sources-of-growth calculation includes within the estimate of TFP growth the effect of the reallocation of factors of production across sectors.

In a time-series analysis of a given country, TFP growth consists of two components (Massell 1961): the weighted intrasector growth rates of TFP and the impact of factor reallocation across sectors. The economywide growth rate of TFP, A^* , is

$$(2) \quad A^* = \sum_i \frac{V_i}{V} A_i^* + (1 - \alpha) \sum_i \frac{MP_i^L}{MP^L} \left(\frac{L_i}{L} \right)^* + \alpha \sum_i \frac{MP_i^K}{MP^K} \left(\frac{K_i}{K} \right)^*$$

where V_i is value added originating in each sector; V is economywide value

7. Barro (1991), following Lucas (1988) and Romer (1989), argues that higher levels of human capital are conducive to more efficient physical investment. Thus, investment may be viewed as endogenous. But his most comprehensive estimates treat investment as exogenous and not part of a simultaneous equation system.

8. If the production functions differ sufficiently among sectors, differences in capital-labor ratios may not result in a larger marginal product of labor in the industrial sector, but this is an empirically implausible case.

added; A_i^* is the sectoral growth rate of TFP; MP_i^L and MP_i^K are the marginal products of labor and capital, respectively, in sector i ; MP^L and MP^K are the economywide marginal products of labor and capital, respectively; L_i/L and K_i/K are the shares of labor and capital in sector i as a share of economywide labor and capital, respectively; and α is the elasticity of output with respect to capital.

The figures in column 4 of table 1 show the increases in output attributable to the shift of labor from agriculture to industry and services over time. The figures are calculated using the relative sectoral marginal product of labor for economies in four income groups: low-income economies, lower-middle-income economies, upper-middle-income economies, and industrial market economies.⁹ It is assumed that the underlying production function is Cobb-Douglas, and the marginal product of each factor can be calculated by multiplying the elasticity of output with respect to the factor by its average product.

The shift of labor from agriculture to industry and services accounted for an increase of 0.04 percentage points per year in output for low-income countries during 1960–80. Comparable figures for lower-middle-income, upper-middle-income, and industrial countries are 0.07, 0.04, and -0.01 (table 1, column 4). Although these are group averages and some countries changed their position among the four income groups between 1960 and 1980, the result would not be much different if averages of the changes in individual countries were used.

The productivity gain from reallocation of the labor force overstates the net gain. If the marginal product of labor is lower in the agricultural sector than in industry and services, the marginal product of capital is greater in the agricultural sector than in industry and services unless TFP levels are sufficiently greater in the industrial and services sectors. Thus the reallocation of capital toward industry and services will reduce the gross gain in TFP from the reallocation of labor. The comparable calculation for capital cannot be made because data on capital stock by sector are not available. Given the mobility of capital, however, the intersectoral divergence in the marginal product of capital is likely to be much less than that in labor. Therefore the gross effect of the reallocation of labor, shown in column 4 of table 1, would be reduced by a relatively small figure.

The reallocation of labor (column 4) accounts for one-seventh of the growth in economywide TFP for the low-income countries. It accounts for a small percentage of the economywide TFP growth for the two middle-income groups and for roughly zero for the industrial countries. Economywide TFP growth net of labor reallocation effects is shown in column 6. The figures reveal no benefit of backwardness. Economywide TFP growth, net of reallocation effects, is greatest for the industrial countries and is very low for the low-income group. The middle-income groups also exhibit lower TFP growth than the industrial countries.

9. These are the main income groups for economies, as defined in World Bank 1984.

Table 2. *Annual Rates of Growth in Total Factor Productivity in Agriculture*

Country group	1960-70	1970-80
Developing	0.42	-0.25
Industrial	0.98	0.70

Source: Lau and Yotopoulos (1989, table 12).

These figures are consistent with the low time-series measures of the intrasectoral growth rates of TFP for individual sectors in developing countries (see Lau and Yotopoulos 1989 for agriculture and Pack 1988 for manufacturing). The benefits of relative backwardness cannot be demonstrated through grouped country data. The negative coefficient found for *GDP60* in regressions implies that TFP growth should be greater in countries that had lower income levels in 1960. This is not confirmed by time-series studies of individual countries. *GDP60* may be capturing the potential for sectoral reallocation in the succeeding period but not the realization of higher rates of intrasectoral productivity growth.

While functions such as equation 1 have not been estimated for agriculture and manufacturing, there are suggestive cross-country agricultural production functions. For example, a pooled time-series cross-country agricultural production function estimated by Lau and Yotopoulos (1989) finds that rates of growth of TFP in the agricultural sector were *higher* in industrial countries in both 1960-70 and 1970-80 (table 2). Moreover, their sample of developing countries is weighted toward middle-income countries. A substantial body of evidence suggests that the poorest developing countries had a weaker TFP performance. If faster intrasectoral TFP growth is to be observed, it must be in manufacturing or services.

Using data from seventeen countries, Nishimizu and Page (1987) find a negative relationship between overall TFP growth rates in industry and the initial level of GDP per capita, but only after omitting countries with conspicuously weak TFP performance. Within the developing countries as a group there is no significant relation between TFP growth and initial per capita income.

Are Countries on Their Production Functions?

Existing studies of convergence utilize the period 1960-85. Toward the end of this period, however, many developing countries suffered serious macroeconomic shocks, including the 1979 oil price increase; a deterioration in the current account (partly reflecting a large drop in the commodity terms of trade) as a result of the world recession; and a growing foreign exchange constraint as short-term borrowing rates increased the cost of servicing existing loans and banks became reluctant to lend additional funds. Industrial countries, too, suffered from the most serious recession since the 1930s. As a consequence, the regressions testing convergence reflect these short-term demand-side phenomena as well as the supply-side phenomena, which they are intended to test. The

estimates thus contain an unknown bias and a confounding of supply and demand effects. There is also an unknown bias in the growth-accounting estimates of TFP, which should, in principle, consider only points on the production function. Thus Elias's (1990) estimates may have a downward bias because of the inclusion of the 1980–85 period. But time-series studies confirm the low TFP growth in developing countries for earlier years, not including the 1980s (see the references cited in Pack 1988).

II. PITFALLS POSED BY TECHNOLOGY GAPS

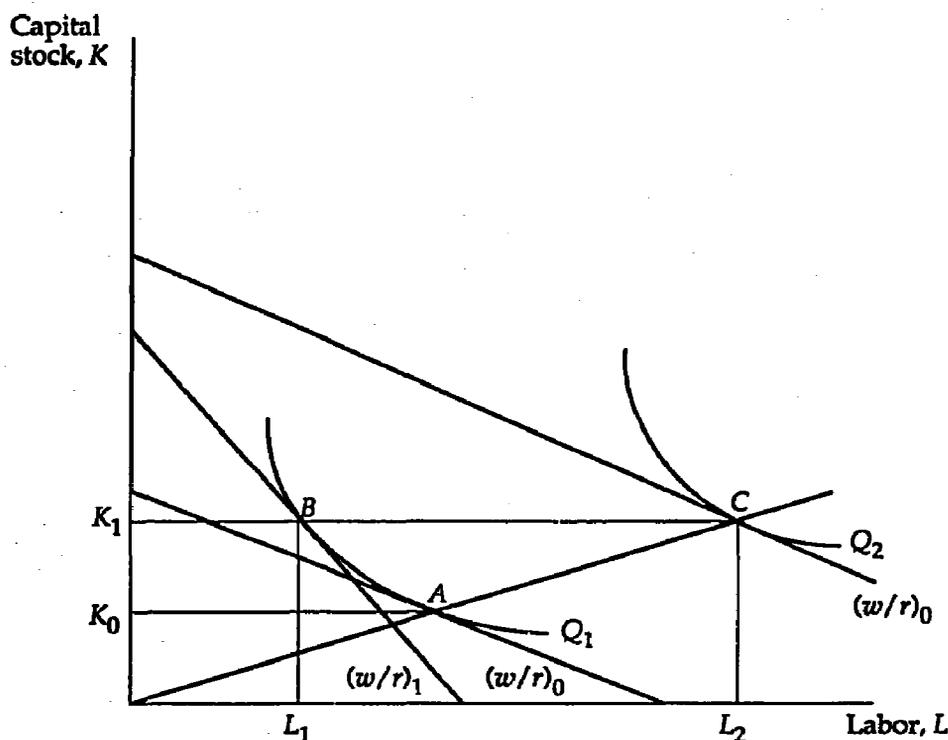
The usefulness of assuming that benefits will automatically accrue to lagging countries ignores an earlier literature that views the existence of a set of more capital-intensive techniques in industrial countries as posing a risk for developing countries. ("Technique" is used here to denote a specific capital-labor ratio or method of production available within the set of techniques constituting an isoquant map.)

The difficulty most widely cited was that capital-intensive production methods—such as automation of the movement of material within factories and the use of large tractors in agriculture—were developed in industrial countries in response to a different set of relative factor prices than that prevailing in the developing countries. The existence of advanced equipment facilitated the substitution of capital for labor along isoquant Q_1 in figure 2 when developing countries with a high ratio of unskilled labor to capital and a low shadow price for the wage-rental ratio, $(w/r)_0$, distorted domestic factor prices to $(w/r)_1$. Because of their prior development in industrial countries, firms in developing countries could more easily purchase advanced technologies and choose point B , whereas social optimality requires the choice at point A . Assuming that aggregate investment allows an increase in the capital stock from K_0 to K_1 and total labor supply L_2 , the economy could potentially reach Q_2 if technique A were chosen. With the actual choice of technique B , L_1L_2 workers are consigned to employment in the informal sector, producing an output, Q_i , that is less than the difference between Q_2 and Q_1 .

A small enclave of the labor force benefits from a high capital-labor ratio, K_1/L_1 , and a correspondingly high output-labor ratio, Q_1/L_1 . For the entire economy, however, the output-labor ratio, $(Q_1 + Q_i)/L_2$, is below that achievable with the adoption of a more appropriate technique along the ray from the origin through A , $Q_2/L_2 = q'_0$ in figure 1. Letting $K_1/L_2 = k'_0$ in figure 1, the move from k_0 to k'_0 does not increase output to q'_0 but to a level between q_0 and q'_0 .¹⁰ Since the output lost from the incorrect choice of techniques was empirically important in many countries (White 1978; Pack 1982), the benefits of backwardness would have been dissipated by this phenomenon during much of the period between 1960 and 1980.

10. The value of k_0 is the ratio of capital to labor at A in figure 2.

Figure 2. Choice of Technique: Substitution of Capital for Labor in Developing Countries



Although the responsibility for this misallocation lies with government intervention that artificially raises the wage-rental ratio, the impact is magnified by the existence of efficient capital-intensive equipment designed in industrial countries. There may be potential losses from backwardness.

The income-reducing impact of allocative inefficiency is worsened by the technical inefficiency that has frequently characterized the use of modern equipment. Developing countries adopting the technique at B have often not been able to use the technique to achieve output Q_1 (figure 2). Output has been significantly lower as a result of inadequate mastery of production engineering and the absence of competitive incentives to remain efficient. In contrast, firms choosing the simpler, labor-intensive technique A have often been able to realize Q_1 (Pack 1987; Rhee and Westphal 1977). A more advanced technique is introduced, but the potential increase in output along f_0 in figure 1 is not realized because of both allocative and technical inefficiency.

III. GROWTH AND TECHNOLOGY ABSORPTION

Moving from A to C along f_0 in figure 1, all gains in per capita income are the result of capital accumulation and thus of forgone consumption (ignoring foreign-financed investment). The more interesting implication of the Gerschenkron hypothesis is the potential benefit from moving to D on the higher production function f_1 , thus obtaining a free dividend from being a latecomer.

Although it is undoubtedly helpful not to have to reinvent the automatic loom along f_0 as a country's capital-labor ratio increases, it is potentially very rewarding to be able to move into a technology regime that is more productive at every capital-labor ratio.¹¹ Empirical evidence from the early constant-elasticity-of-substitution studies onward indicate that developing and industrial countries show very different TFP values (see footnote 4). The movement along the production function f_0 as depicted in figure 1 is likely to yield relatively slow growth rates of output per capita even if countries avoid the pitfalls discussed in the last section. These countries are on the escalator mentioned above, whereas countries that can achieve the transition to D may double their per capita income in a dozen years, experiencing an industrial elevator that transforms their economies.

In principle the transition to a new production function can be achieved simply by importing the required inputs: seeds, fertilizer, pesticides, and equipment for agriculture, and machinery for manufacturing. In practice, although it is possible to use such imports to move above f_0 , it is not easy to attain f_1 . In agriculture, research and development to adapt international agricultural technology to local conditions and dissemination of the new technology by extension agents require sustained public sector action because of the lack of appropriability of the benefits (Birkhaeuser, Evenson, and Feder 1991). In industry, imported equipment could potentially be operated at international TFP levels, thus closing the gap, but purchasers may not achieve international best-practice norms.¹²

Calculations that use a variety of production functions estimated from industrywide data show differences in TFP across manufacturing sectors. These gaps exist in the earliest studies of Arrow and others (1961) and Daniels (1969) and the more recent studies of Pack (1984) and Clague (1991). Thus differences in TFP do not appear to be transitional phenomena along a path from f_0 to f_1 . Such sustained differences may be accounted for by dissimilarity in the vintage or type of machinery employed, the knowledge of production engineering accompanying it, or inefficiencies induced by the incentive regime, even though the same equipment is employed.¹³

A few Asian economies, however, including Korea and Taiwan (China), appear to have narrowed productivity gaps considerably. Their TFP growth rates in individual industrial sectors exceed those in the comparable sectors in industrial countries by a considerable margin. Having exhausted the productivity

11. Romer (in this volume) emphasizes the role of the transmission of ideas.

12. For an extensive discussion of these problems and estimates of their effect in the textile industry in Kenya and the Philippines see Pack (1987).

13. Although industrywide data on TFP indicate persistent differences, it is possible that individual firms are in a state of transition from f_0 to f_1 . Recent panel-data analyses, particularly Liu (1991) for Chile, provide evidence of slowly improving firm-level productivity but do not compare the firms in the sample to world best-practice levels. Liu's estimated rates of TFP growth are too slow to close the gap with respect to best practice.

enhancement possible from reallocation, they were able to move to much higher production functions to sustain their growth. Few other countries have exhibited large values of TFP growth either for the national economy or for individual sectors (Pack 1988; Elias 1990). The key question is how these economies have been able to shift from f_0 to f_1 and achieve a major transformation of their economies in a few decades.

Closing Productivity Gaps in Manufacturing

In analyzing the routes available for achieving the upward shift in the production function toward world best-practice frontiers, it is important to disaggregate the analysis to at least manufacturing and agriculture. Services and construction are left out of the purview because of the limited amount of research on these sectors. In light of the extensive analysis of the sources and nature of productivity growth in agriculture (Kawagoe, Hayami, and Ruttan 1985; Binswanger, Yang, and Bowers 1987; Lau and Yotopoulos 1989), I concentrate here on manufacturing.

Alternate Modes of Technology Transfer in Manufacturing

Firms have several alternatives for obtaining new technology that, if mastered, yield a higher level of TFP for any given capital-labor ratio. These alternatives include: (a) the purchase of new equipment;¹⁴ (b) direct foreign investment; (c) the purchase of technology licenses for domestic production of new products or the use of new processes;¹⁵ (d) the use of nonproprietary technology, including that obtained from purchasers of exports; (e) acquisition of knowledge from returning nationals who have been educated or have worked in industrial countries and from nationals who remain in industrial countries; and (f) domestic research and development and efforts in reverse engineering.

All these possibilities, except for the research and development efforts, represent an attempt to move toward international best practice by transferring technologies available abroad. The research and development alternative may have an element of aiding the identification, modification, and absorption of foreign technology rather than generating a completely indigenous technology.

IV. THE EXPERIENCE OF TWO SUCCESSFUL ASIAN ECONOMIES

This section describes some of the means by which two of the fastest-growing newly industrializing economies—Korea and Taiwan (China)—were able to shift toward an international production function.¹⁶

14. De Long and Summers (1991) find that investment in new equipment as a percentage of GDP is an important explanatory variable in cross-country growth performance. For calculations of the impact of embodiment of new technology in equipment on growth rates in Taiwan (China), see Pack (1992).

15. See, for example, the detailed analysis in Nagaoka (1989) of the role of licensing in Japan.

16. The evidence for the interpretations in this section is set forth in Dahlman and Sanada-Kone (1990), Westphal, Rhee, and Pursell (1981), Pack and Westphal (1986), and Pack (1992).

Korea and Taiwan (China)

Until the mid- to late 1970s neither Korea nor Taiwan (China) employed explicit technology policies. The main exceptions were the restrictions placed on direct foreign investment and a fairly perfunctory review of technology licensing agreements in Korea. The ability of the two countries to close the initial productivity gaps was a result of firms' responses to the incentives contained in national economic policies. Among these policies were: (a) the relative neutrality of the foreign trade regime with respect to profitability between domestic and foreign sales and the relatively low variance in protection across sectors; (b) export targeting in Korea and undervaluation of the real exchange rate in Taiwan (China) to encourage exports to a greater extent than would have been the case given the protection afforded to new industries in the domestic market; (c) a relatively undistorted labor market that, along with some movement toward market rates of interest (particularly in Taiwan, China), kept the wage-rental ratio closer to its scarcity value than in other developing countries.

The responses to these incentives led to a set of favorable but unintended technological consequences. For example, as a result of the rapid rates of export growth that these policies encouraged, there was a substantial inflow of non-proprietary technology, embodied in equipment and in the knowledge provided by customers (Westphal, Rhee, and Pursell 1981). This inflow was greater because exports and production increased most in older labor-intensive sectors in which technology from industrial countries was less protected. Technology and knowledge were relatively easy to acquire and absorb in these sectors even without a large stock of highly educated engineers. Much of the relevant information was based on mechanical knowledge rather than on electronic, biological, or chemical principles that would have required more formal education of employees.

Moreover, the machinery that was employed to manufacture the increased output was quite labor-intensive, in response to the low wage-rental ratio (Ranis 1979; Rhee and Westphal 1977). The simple equipment and the absence of continuous processing were conducive to minor innovations for increasing productivity, which were often suggested by blue-collar workers. Thus, the trade and factor price regimes were complementary and were conducive both to obtaining static gains in output and to fostering the move toward best practice. In this period, until the late 1970s, it is likely that much of the growth in productivity was the unplanned consequence of getting the prices right. Dollar and Sokoloff (1990) find that TFP growth in labor-intensive sectors in Korea exceeded that in the capital-intensive sectors. Technology policy was implicit in the standard economic policies, and technological learning complemented the conventional economic responses, stimulating further growth in production and exports as a consequence of reduced production costs.

In the 1970s a more explicit policy toward technology acquisition appeared. This policy differed in the two economies. In Korea the growth of large local

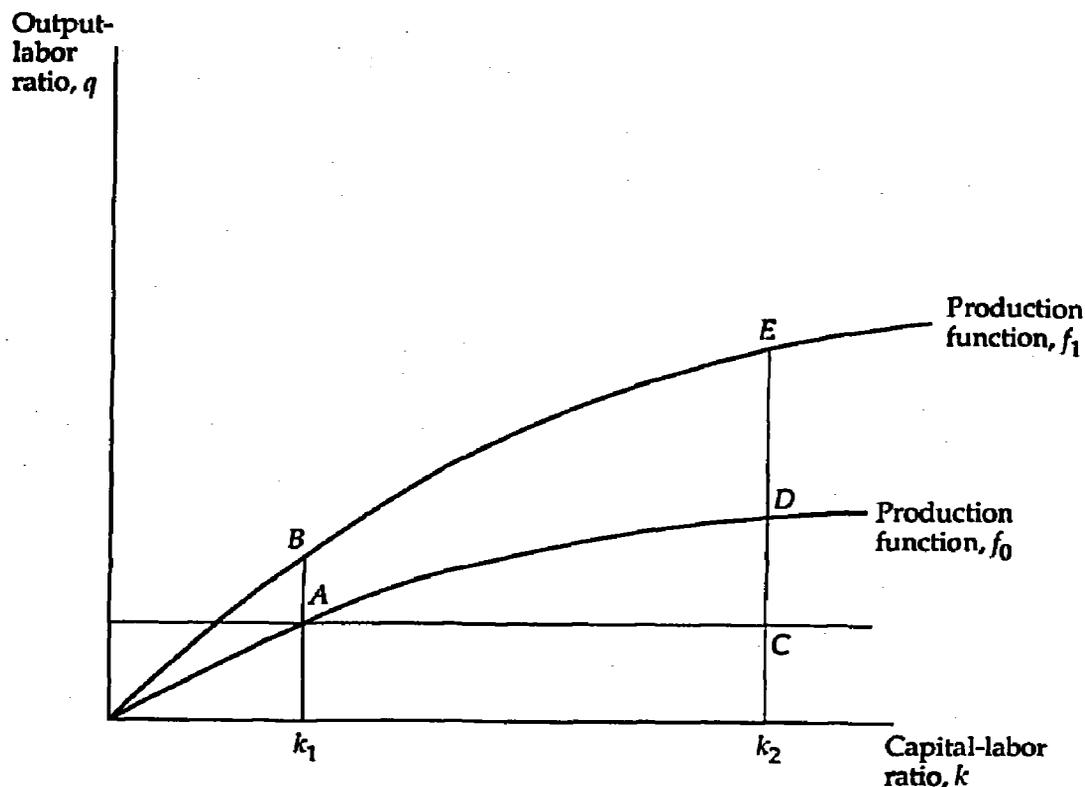
firms was encouraged by the use of selected credit and other instruments. As domestic real wages increased and newer lower-wage competitors entered the international market, large Korean firms were encouraged to acquire the technological capacity to enter sectors that were more capital- and technology-intensive and to achieve best-practice productivity (Pack and Westphal 1986). Information about production technology in these more complex producer goods sectors was likely to be more closely guarded than in the consumer goods industries, and importers in industrial countries were less likely to transfer such technology. The Korean government encouraged firms to obtain technology licenses, acquire advanced equipment, and engage in their own research and development.

In Taiwan (China) the transfer of knowledge in the consumer industries, in which the early export drive was concentrated, was similar to that in Korea (Pack 1992). As Taiwan entered newer areas, however, it did not encourage the growth of large-scale firms capable of substantial research and development. The industrial structure was characterized by many small firms, reflecting the prevalence of high interest rates and the limited use of selected credit directed to larger firms. Therefore Taiwan utilized central institutions such as the Industrial Technology Research Institute, as well as technology diffusion institutions such as the China Productivity Center, to introduce new technologies, develop new products and processes, diffuse knowledge of them, and scan international markets for both products and processes (see Dahlman and Sananikone 1990). Moreover, in the newest sectors the ability to attract back Taiwanese nationals or to utilize the knowledge of those who remain abroad has been critical (Pack 1992).

In both economies, people (and the knowledge they embody) who have been educated abroad return because of the high wages made possible by growing exports. Purely domestically oriented firms with smaller sales bases could not have offered sufficiently high wages to attract them. The newly acquired international knowledge was embedded in a framework conducive to efficiency. Competitive pressures led to a search within plants for better productivity performance. As a result, imported practices were improved, and purely domestic efforts were made to increase productivity.

The Interaction of Knowledge Acquisition, Investment, and Human Capital

Both Korea and Taiwan (China) invested extensively in education and in the accumulation of substantial physical capital. The ratio of investment to GDP increased from relatively low levels to more than 30 percent in the 1980s. Figure 3 elucidates the process. Initially the economy is at point *A* on production function f_0 . As physical and human capital accumulation proceed, it moves to point *E* on production function f_1 . The shift to the higher production function is realized because of the growing utilization of international best practice. Note, however, that the benefit from this accumulation of knowledge would have been less— $AB < DE$ —if capital per worker had not grown (Nelson 1973). Thus the size of the benefit from the growing import of knowledge and from local efforts

Figure 3. *Interaction of Technical Change and Factor Accumulation*

to increase productivity depends on the stock of physical investment and skills complementing local unskilled labor.¹⁷

Efforts to obtain international knowledge will have lower payoffs if they are not accompanied by a growth in the stock of capital per worker.¹⁸ Capital, both physical and human, can be partly supplied by other countries in the form of direct foreign investment. The remarkable development of Singapore, for example, demonstrates the potency of externally provided capital and skills in facilitating a rapid movement to international best practice (Lim and Fong 1991). In the initial period of rapid growth of industrial productivity, Korea and Taiwan (China) benefited from both capital accumulation and the move toward international best practice.

It may be conjectured that the extent of the shift in the production function would have been less if the sectors in which exports grew had been those in which these countries were close to world best practice. In Chile, a more recent example of improved policies, TFP growth has been much slower. Part of the

17. This has also been a major conclusion of some endogenous growth models, although they conclude that the production function in figure 3 should not indicate diminishing returns in the accumulable factors, physical and human capital. There is no empirical evidence to support this view.

18. Dollar (1992) also emphasizes this complementarity, deriving it from the interaction term in his cross-country regressions.

explanation for this may lie in its emphasis on primary exports, minerals, and agricultural products. It is likely that these sectors in Chile were much closer to international best practice than were the industrial growth sectors in Korea and Taiwan (China). Moreover, in some of the expanding export sectors, such as electronics, the best-practice frontier was itself shifting rapidly, and the two Asian economies were able to take advantage of this. If these conjectures are correct, early proponents of import-substituting industrialization such as Singer and Prebisch may have been correct in their intuition of the dynamic (TFP growth) benefits of industrialization. They were mistaken, however, in their emphasis on import substitution rather than export growth as the process for realizing these benefits.

V. CONCLUSIONS

Models that attempt to test whether developing countries can reap the benefits of their initial backwardness and thus grow faster have led to a welcome revival of interest in the sources of growth of potential output in developing countries. A number of these studies find that, after adjusting for relevant variables, countries that had lower per capita income levels at the beginning of the 1960–85 period experienced faster growth than would have been predicted. This is attributed to the ability to take advantage of the large existing stock of knowledge of the more advanced countries. In contrast, analyses of the sources of growth over time in these countries do not confirm that poorer countries have had higher rates of growth of TFP than richer countries, at either the national or the sectoral level. Moreover, adjusting for the impact of shifts of resources away from sectors of low marginal productivity, the gaps between the TFP growth of the industrial countries compared with both middle-income and low-income developing countries is even larger.

As often occurs in a new area of research, concern with establishing the stylized facts has dominated the interest in assessing the mechanisms through which developing countries may achieve greater growth. The source of growth in a few Asian economies was their ability to extract relevant technological knowledge from industrial economies and utilize it productively within the domestic economy. Not only was the technology acquired, but it led to rapid growth in productivity because it was embedded in economies that offered strong inducements to export and in which firms attempted to decrease cost and improve quality.

The lessons for those developing countries that have experienced slow growth is not that they will *automatically* reap the benefits of their current low-income status, as implied by the convergence argument. Rather, intensive efforts will have to be made to acquire technology and to do so within a framework in which successful firms do not use it to extract rents. Some of the knowledge may be acquired as a by-product or an externality from exporting. Hence, a competitive initial environment with a neutral or slightly proexport incentive structure may be a precondition for the acquisition of nonproprietary knowledge. At the

same time, practices that discourage direct foreign investment, as well as the microscopic inspection and slow approval of licensing agreements, need to be relaxed. Establishing relative factor prices in line with endowments is also important. Simpler technologies and more labor-intensive sectors allow increases in domestic productivity and the exploitation of nonproprietary knowledge that is more readily available in these sectors.

In the current international milieu of rapidly changing macroeconomic policies, the need for liberalization in these microeconomic dimensions should not be disregarded. A strong case can be made that getting most macroeconomic policies correct without giving explicit attention to technology acquisition will lead to moderate growth, but at rates that will not approach those of the Asian "miracle" economies.

REFERENCES

- Arrow, K. J., H. B. Chenery, B. S. Minhas, and R. M. Solow. 1961. "Capital-Labor Substitution and Economic Efficiency." *Review of Economics and Statistics* 43:225-50.
- Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics* 105:407-43.
- Baumol, William J. 1986. "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show." *American Economic Review* 76:1072-85.
- Binswanger, Hans, M. Yang, and A. Bowers. 1987. "On the Determinants of Cross-Country Aggregate Agricultural Supply." *Journal of Econometrics* 36:111-31.
- Birkhaeuser, Dean, Robert E. Evenson, and Gershon Feder. 1991. "The Economic Impact of Agricultural Extension: A Review." *Economic Development and Cultural Change* 39:607-50.
- Chenery, Hollis B. 1960. "Patterns of Industrial Growth." *American Economic Review* 50:624-54.
- Chenery, Hollis B., Sherman Robinson, and Moshe Syrquin. 1986. *Industrialization and Growth: A Comparative Study*. New York: Oxford University Press.
- Clague, Christopher. 1970. "The Determinants of Efficiency in Manufacturing Industries in an Underdeveloped Country." *Economic Development and Cultural Change* 18:188-205.
- . 1991. "Relative Efficiency, Self-Containment, and Comparative Costs of Less Developed Countries." *Economic Development and Cultural Change* 39:507-30.
- Dahlman, Carl J., and Ousa Sananikone. 1990. "Technology Strategy in the Economy of Taiwan (China): Exploiting Foreign Linkages and Investing in Local Capability." World Bank, Industry and Energy Department, Washington, D.C.
- Daniels, Mark. 1969. "Differences in Efficiency among Industries in Developing Countries." *American Economic Review* 49:159-71.
- De Long, J. Bradford. 1988. "Productivity Growth, Convergence, and Welfare: Comment." *American Economic Review* 78:1138-54.
- De Long, J. Bradford, and Lawrence H. Summers. 1991. "Equipment Investment and Economic Growth." *Quarterly Journal of Economics* 106:445-502.
- Denison, Edward F. 1985. *Trends in American Economic Growth*. Washington D.C.: Brookings Institution.

- Dollar, David. 1992. "Exploiting the Advantages of Backwardness: The Importance of Education and Outward Orientation." World Bank, East Asia and Pacific Department, Washington D.C.
- Dollar, David, and Kenneth Sokoloff. 1990. "Patterns of Productivity Growth in South Korean Manufacturing Industries, 1963-1979." *Journal of Development Economics* 33:309-27.
- Elias, Victor. 1990. "The Role of Total Factor Productivity on Economic Growth." Background paper for *World Development Report 1991*. World Bank, World Development Report office, Washington, D.C.
- Gerschenkron, Alexander. 1962. *Economic Backwardness in Historical Perspective*. Cambridge, Mass.: Harvard University Press.
- Hirschman, Albert. 1957. *The Strategy of Economic Development*. New Haven, Conn.: Yale University Press.
- Kawagoe, Toshihiko, Yujiro Hayami, and Vernon Ruttan. 1985. "The Inter-Country Agricultural Production Function and Productivity Differences among Countries." *Journal of Development Economics* 19:113-32.
- Krueger, Anne O. 1969. "Factor Endowments and Per Capita Income Differences." *Economic Journal* 78:641-59.
- Kuznets, Simon. 1959. *Six Lectures on Economic Growth*. Glencoe, Ill.: Free Press.
- Lau, Laurence, and Pan Yotopoulos. 1989. "The Meta-Production Function Approach to Technological Change in World Agriculture." *Journal of Development Economics* 31:241-70.
- Lim, Linda Y. C., and Pang Eng Fong. 1951. *Foreign Direct Investment and Industrialization in Malaysia, Singapore, Taiwan, and Thailand*. Development Centre Studies. Paris: Organization for Economic Cooperation and Development.
- Liu, Lili. 1991. "Entry-Exit, Learning, and Productivity Change: Evidence from Chile." Policy Research Working Paper 769. World Bank, Country Economics Department, Trade Policy Division, Washington, D.C.
- Lucas, Robert E. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22:3-42.
- Mankiw, Gregory, David Romer, and David N. Weil. 1990. "A Contribution to the Empirics of Economic Growth." NBER Working Paper 3541. National Bureau of Economic Research, Cambridge, Mass.
- Massell, Benton. 1961. "A Disaggregated View of Technical Change." *Journal of Political Economy* 69:547-57.
- Nagaoka, Sadao. 1989. "Overview of Japanese Industrial Technology Development." Industry Series Paper 6. World Bank, Industry and Energy Department, Washington D.C.
- Nelson, Richard R. 1973. "Recent Exercises in Growth Accounting: New Understanding or Dead End?" *American Economic Review* 73:162-68.
- Nishimizu, Mieko, and John M. Page, Jr. 1987. "Economic Policies and Productivity Change in Industry: An International Comparison." World Bank, Country Economics Department, Washington D.C.
- Pack, Howard. 1982. "Aggregate Implications of Factor Substitution in Industrial Processes." *Journal of Development Economics* 11:1-38.
- . 1984. "Total Factor Productivity and Its Determinants." In Gustav Ranis, Robert L. West, Mark W. Leiserson, and Cynthia Taft Morris, eds., *Comparative Development Perspectives: Essays in Honor of Lloyd G. Reynolds*. Boulder, Colo.: Westview.
- . 1987. *Productivity, Technology, and Industrial Development*. New York: Oxford University Press.

- . 1988. "Industrialization and Trade." In Hollis B. Chenery and T. N. Srinivasan, eds., *The Handbook of Development Economics*. 2 vols. Amsterdam: North-Holland.
- . 1992. "New Perspectives on Industrial Growth in Taiwan." In Gustav Ranis, ed., *Taiwan: From Developing to Mature Economy*. Boulder, Colo.: Westview.
- Pack, Howard, and Larry E. Westphal. 1986. "Industrial Strategy and Technological Change: Theory versus Reality." *Journal of Development Economics* 22:87-128.
- Ranis, Gustav. 1979. "Industrial Development." In Walter Galenson, ed., *Economic Growth and Structural Change in Taiwan*. Ithaca, N.Y.: Cornell University Press.
- Rhee, Yung W., and Larry E. Westphal. 1977. "A Microeconomic Investigation of Choice of Technique." *Journal of Development Economics* 4:205-38.
- Romer, Paul. 1986. "Increasing Returns and Long-Run Growth," *Journal of Political Economy*. 94:1002-37.
- . 1989. "Human Capital and Growth: Theory and Evidence." NBER Working Paper 3173. National Bureau of Economic Research, Cambridge, Mass.
- Summers, Robert, and Alan Heston. 1988. "A New Set of International Comparisons of Real Product and Price Level Estimates for 130 Countries, 1950-85." *Review of Income and Wealth*. 34:1-25.
- Syrquin, Moshe. 1984. "Resource Reallocation and Productivity Growth." In Moshe Syrquin, Lance Taylor, and Larry E. Westphal, eds., *Economic Structure and Performance: Essays in Honor of Hollis B. Chenery*. Orlando, Fla.: Academic Press.
- Teitel, Simon. 1981. "Productivity, Mechanisation and Skills: A Test of the Hirschman Hypothesis for Latin American Industry." *World Development* 9:855-71.
- United Nations. Various years. *Yearbook of National Income Accounts*. New York.
- Westphal, Larry E., Yung W. Rhee, and Garry G. Pursell. 1981. *Korean Industrial Competence: Where it Came From*. World Bank Staff Working Paper 469. Washington, D.C.
- White, Lawrence J. 1978. "The Evidence on Appropriate Factor Proportions for Manufacturing in Less Developed Countries: A Survey." *Economic Development and Cultural Change* 27:27-60.
- World Bank. 1984. *World Development Report 1984*. New York: Oxford University Press.

COMMENT ON "TECHNOLOGY GAPS BETWEEN INDUSTRIAL
AND DEVELOPING COUNTRIES," BY PACK

John Page

Howard Pack's paper is a useful critique and reinterpretation of some of the literature on the "new economics of growth," especially the growing number of cross-country regressions that explain the sources of intercountry variations in per capita income growth.

Perhaps the most striking aspect of this new literature is the focus on "convergence"—that is, the capacity of poor countries to catch up to per capita income in the industrial countries. According to both neoclassical and endogenous growth theories, poorer economies can accelerate their rate of per capita income growth during the transition through improvements in such fundamentals as the rate of investment and the level of human capital. Where they diverge is in interpreting the potential for growth that does not derive from such fundamental processes of accumulation.

In neoclassical growth theory (and in its empirical counterpart, growth accounting) the component of growth that cannot be explained by accumulation is ascribed to change in total factor productivity (TFP) or to technical progress. In many endogenous growth models, external economies explain different rates of growth in per capita output. The theories differ in their interpretation of the steady-state properties of long-term growth. Roughly speaking, neoclassical models converge to a uniform rate of per capita income growth across countries that is determined by the rate of technical progress; endogenous growth models do not converge to a uniform rate of per capita income growth because of the presence of externalities related to the process of accumulation. Empirical tests of convergence are therefore viewed as a means of testing the validity of these two competing theoretical views of growth.

Pack's paper looks at these empirical tests and asks what, if anything, we learn from this burgeoning literature. My reading of his results and arguments leads me to conclude that we learn very little. Cross-country growth regressions find significant coefficients on variables intended to describe "relative backwardness," but the interpretation of the coefficient as evidence of catching up is subject to doubt. Pack advances three arguments to explain why this is so. First,

John Page is senior adviser in the Country Economics Department of the World Bank.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

it is not consistent with time-series evidence on TFP change across countries, since between 1950 and 1985 those growth rates were lower, on average, in poorer countries than in richer countries.

Second, the one-sector models employed in cross-country regressions ignore a major source of TFP change: intersectoral reallocation of factors from low-productivity to high-productivity sectors. To one who was trained in economic development in the early 1970s, this is an astonishing omission in the recent theoretical and empirical literature.

As Pack demonstrates, reallocation effects can be both empirically significant and contrary to the interpretation of the convergence literature. What is perhaps more disturbing in Pack's results—and deserving of more emphasis—is that, net of sectoral reallocation, TFP change in the poorest countries is negative, a phenomenon that neither the neoclassical nor endogenous growth models can explain and to which I shall return later.

Finally, Pack questions the interpretation of a number of the variables introduced into cross-country regressions by various authors. He argues that a misinterpretation of the variables can lead to a misinterpretation of the results. His most intriguing speculations concern the ratio of investment to gross domestic product. While the traditional interpretation of this variable is that its coefficient reflects the effect of capital deepening, it is also plausible that, as Pack argues, higher rates of investment are associated with higher rates of structural transformation.

It is also possible that high investment ratios act as a proxy for the overall adequacy of the economic environment and that growth is positively associated with "good" economic environments. If this is so, a share of the positive coefficient on investment may not reflect the results of accumulation but rather the productivity-enhancing effects of good economic policy. This is a point with which I suspect Pack would agree.

Pack has done an admirable job of convincing us that whatever the coefficient on "relative backwardness" measures, it is probably not convergence in the sense of endogenous growth theories. The empirical literature is therefore not yet well enough developed to allow a definitive test of the relevance of the competing neoclassical and endogenous interpretations of growth.

But a major question that arises in a forum such as the World Bank is whether this debate is of more than academic interest to development practitioners and policymakers. I would answer, yes. As Pack points out in his introduction, the fundamental question is to what extent are sources of growth that cannot be attributed directly to accumulation amenable to policy action? If growth rates can never converge because of externalities related to accumulation, or if TFP change is limited by an exogenous rate of technical progress, public policy even during the transition is probably best confined to getting the fundamentals right. But if, as the experiences of a number of high-performing Asian economies suggest, rates of TFP change are themselves determined by policy choices, the possible areas for intervention are considerably expanded.

I would therefore like to interpret Pack's evidence in this context rather than in terms of the debate between competing interpretations of economic growth. To do this, I need to introduce yet another view of the component of growth that cannot be attributed to accumulation. Nishimizu and Page (1982) offer an interpretation of TFP change that brings together the concepts of technical efficiency and technical progress.¹ Our motive was the realization that intrasectoral (industry) rates of TFP change were negative in many developing countries over extended periods of time. Negative TFP change is equivalent to an increase in real unit costs of production and is contrary to traditional interpretations of the residual in growth-accounting exercises as technical progress.

Our framework is very similar to Pack's. Technical progress is associated with movements in best practice. It is zero or positive but never negative. Changes in technical efficiency, however, are changes in the average efficiency with which best practice is applied. These can be either positive or negative, depending on whether an economy (or firm) is catching up to or falling behind best practice. To use the variables in figure 1 in Pack's paper, movements of the production frontier, f_1 , are technical progress. Movements from points below f_1 (say, along f_0) toward f_1 are changes in technical efficiency. TFP change, the measured residual in growth accounting, contains both elements, but the techniques of growth accounting cannot disentangle the two.

Why is this distinction relevant? To use Pack's analogy, the process of catching up by means of productivity change is similar to riding on an escalator. Technical progress is positive and common to all countries, like standing on a rising escalator. Change in technical efficiency is the process of running up the escalator to catch an economy on a higher stair. If change in technical efficiency is amenable to policy action, running up the escalator is possible.

Viewed in this light, Pack's data paint a pessimistic picture of the prospects for catching up, albeit with some room for optimism. If rates of TFP growth in high-income countries can be interpreted as the rate of technical progress, they lie in the range of 1 to 2 percent a year (on a value added basis). Low- and middle-income countries have lower average rates of TFP change. Thus change in technical efficiency is, on average, negative in the developing world. Countries are falling back on the escalator.

A subset of economies, however, is running ahead. Pack discusses the experience of the Republic of Korea, and Taiwan (China); I would add Japan and Hong Kong. Each of these economies recorded extended periods of rapid TFP growth (more than 1 to 2 percent a year) and therefore a positive change in technical efficiency. Pack makes a persuasive case that this positive change is associated with public policy interventions to promote exports, combined with a

1. Mieko Nishimizu and John M. Page, Jr., "Total Factor Productivity Growth, Technical Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia 1965-1978," *Economic Journal* 92 (1982):920-36.

high degree of openness to international technology and knowledge and strong incentives to attract skilled nationals back to their home country.

Pack's evidence persuades me that change in technical efficiency is indeed amenable to policy action. Good policies can engender positive change, while bad policies can have the opposite effect. I would like to conclude, therefore, by speculating on what might constitute good policies.

First, proexport policies may have substantial benefits. Exports introduce competition, international standards of best practice and product quality, and information about international trends in productivity change, both directly through information links and indirectly through relative price movements. It is significant, however, that for much of the period of rapid growth in technical efficiency the proexport policies of Japan, Korea, and Taiwan (China) were not associated with neutral incentives and a liberal economic environment. (This was not true of Hong Kong.) Thus we may need to extend our research and thinking about the role of trade and trade policy beyond the standard prescriptions of neutrality of incentives and outward orientation.

Second, changes in technical efficiency are not costless. The accumulation of human capital and the mastery of technological skills are necessary for accelerating productivity growth. Both require investment and patience. It is not accidental that those countries which registered high rates of change in technical efficiency were also among the world leaders in the share of investment in national income and in school (especially primary school) enrollment rates.

Third, technology policies as practiced in most low- and middle-income countries may be misguided. Basic research and university-based applied research may be far less relevant to the rapid acquisition of technical efficiency than firm-based innovations and general education. As Pack points out, neither Korea nor Taiwan (China) employed explicit technology policies until quite late in their growth process.

The lessons for low- and middle-income countries ultimately reflect both traditional concerns with investment and accumulation and the more recent recognition of the scope for increases in productivity derived from improvements in technical efficiency. On the traditional side, the accumulation of physical investment and human capital is paramount, although the mechanisms by which these variables act on growth may be more subtle than the cross-country growth literature suggests. On the other side, change in technical efficiency remains an area of great unrealized potential for developing economies other than those in East Asia. Policies matter for both traditional and nontraditional sources of growth.

COMMENT ON "ACCUMULATING TECHNOLOGICAL CAPABILITY IN DEVELOPING COUNTRIES," BY BELL AND PAVITT, AND "TECHNOLOGY GAPS BETWEEN INDUSTRIAL AND DEVELOPING COUNTRIES," BY PACK

Jorge M. Katz

The two papers presented in this session provide a vivid picture of the fragmentary and unsatisfactory state of received theory concerning the sources and nature of technical change and its long-term effects on productivity growth and productivity differentials across firms, industries, and nations. It is unfortunate that without further progress we can scarcely claim to have a conceptually valid and empirically useful set of principles governing international competitiveness and the sources of dynamic comparative advantages across countries—certainly one of the central topics for the economics profession.

Pack examines the conventional neoclassical account of intercountry differences in total factor productivity growth and the likelihood that the gap between industrial and developing countries will become smaller as poorer countries catch up, moving upward along an equilibrium growth path. This approach to development economics is based on two propositions. First, a library of completely specified and freely accessible production functions exists somewhere in the world, and developing countries are free to make use of it. Second, growth in per capita income is to be explained solely on the basis of the expansion of the ratio of capital to labor, the degree of initial backwardness (which reflects how much any given country can benefit from tapping the common pool of available technologies), and the ability of a country to get its prices right in order to compete on the world market.

To deal with questions of technical change and productivity growth at this level of abstraction, the neoclassical model needs to make a number of stringent assumptions concerning industrial organization and the behavior of economic agents, markets, and institutions. It has to assume that we are looking at steady-state equilibrium situations in which factors are paid their marginal productivity and entrepreneurs maximize profits under conditions of perfect information. It also has to postulate a well-defined intertemporal relationship between generations to make *dynamic* equilibrium at all possible. The likelihood of disequilibrium is eliminated from neoclassical growth models through an elegant—but, in Solow's words, "ultimately unacceptable"—simplification.

Jorge M. Katz is professor of industrial economics at the University of Buenos Aires in Argentina.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

The idea is to imagine that the economy is populated by a single immortal consumer, or by a number of identical immortal consumers. The immortality itself is not a problem. Each consumer could be replaced by a dynasty each member of which treats his or her successors as extensions of himself or herself. But no shortsightedness can be allowed. He or she is supposed to solve an infinite time utility function. The next step is harder to swallow in conjunction with the first. For this consumer every firm is just a transparent instrumentality, an intermediary, a device for carrying out intertemporal optimization subject only to technological constraints and initial endowments. Thus, any kind of market failure is ruled out from the beginning. There are no strategic complementarities, no coordination failures, no Prisoner's Dilemmas. The result is a construction in which the whole economy is assumed to be solving a Ramsay optimal growth problem through time, disturbed only by stationary stochastic shocks to taste and technology. To these the economy adapts optimally. (Solow 1988, p. 310)

The neoclassical account of development assumes that the developing countries today are exact replicas of the industrial countries at some point in the past and that if everything goes right they should, in due time, move upward along an equilibrium growth path of steady-state configurations. Pack's elegant and attractive presentation of the neoclassical story ends on a discouraging note, though. He tells us that the conventional model fails to provide a sound explanation of the economic successes of the Republic of Korea and Taiwan (China) and that we need a much more complex institutional and structural model if we are to account for the rich behavioral scenarios that underlie these cases. Although we knew that all along, equilibrium growth models still carry a lot of intellectual power within the profession.

Bell and Pavitt present an unorthodox view of technical change more concerned with the dynamics of technological accumulation and with the forces that account for the development of domestic technological capabilities in peripheral societies. Their views derive from a quite different microeconomic and institutional model in which imperfect information allows for behavioral differences among firms, as well as for innovative lags and leads, quasi-rents, and endogenous changes in market structure (Nelson 1981). Technological learning can be different from company to company—or from country to country—depending on how much a given firm or society spends on research and development, the quality of its human resources, and the particular way in which domestic institutions operate to support (or discourage) the generation and diffusion of technical knowledge. It now becomes possible to postulate models of "adaptive" behavior in individual firms that do not have to assume the firm has complete prior knowledge of all its future technological possibilities or that its only objective function is to maximize profits.

Once the idea of regular and predictable behavior inherent in the neoclassical logic is abandoned, the notion of evolutionary performance can be introduced.

Current behavior is strongly influenced by the recent past. There is a certain "biological" flavor in these models that is imparted by the evolutionary mechanism which underlies the dynamics of firm behavior and of market structure.

After nearly two decades of involvement in empirical studies of technical change in Latin American countries (Katz 1986, 1987), I agree with most of what Bell and Pavitt have to say. Yet I recognize that we still lack a comprehensive overall theoretical model as elegant and mathematically operational as the neoclassical one, and for some people, this lack constitutes a major shortcoming. Substantial progress in this direction has been attained in the path-breaking work by Nelson and Winter (1982).

The paper presents many small and theoretically attractive issues as side questions to the central argument. For example, the authors argue that there is a fundamental difference between setting up production capacity and gradually developing in-house technological capabilities and skills. They correctly point out that the latter capacity is crucial and note that firms need an explicit strategy to develop these skills.

Many of our case studies do in fact show that companies have an explicit strategy for encouraging technological learning and that such strategies proceed from simple to complex technological missions and achievements. Firms were observed to move from the development of in-house product design to production engineering skills and only later—perhaps as much as ten years later—to questions of production planning and organization that involve issues of layout balancing, time-and-motion studies, and so forth. Such skills demand quite different human resources than product design activities and are incorporated at a much later stage.

Although this is a minor issue in the paper, the topic is by no means trivial because it goes to the heart of the discussion of infant industry protection and public policy in the industrial field. On the basis of very little empirical evidence, economists have argued that infant industry protection should be short-term and should compensate for no more than 20 or 30 percent of production costs. But the experiences of Korea and Taiwan (China) seem to indicate that we need to go far beyond such levels and that new institutions and policy instruments need to be developed for such purposes.

Let me take issue with a more fundamental point. The paper describes the process of microlevel learning and observes that "there are good reasons to expect firms in *certain market conditions* to underinvest in [the creation of] these capabilities . . ." I think that we could go much further and admit that it is not just individual market conditions that affect a firm's technological strategy: macroeconomic forces play a major role. By admitting this, we can begin to develop a microeconomic and macroeconomic analytical framework that will be needed if we are to understand the idiosyncratic nature of the industrial organization environment of developing countries and eventually develop a theoretical construction that reflects their uniqueness rather than assuming that they are just lagging versions of mature industrial societies.

Throughout the 1980s structural adjustment has required many developing countries to deregulate and open their markets to foreign competition, to privatize public enterprises, and to undertake new patterns of industrial relations. A whole new industrial organization scenario has been developing about which we know little. Consider, for example, the following issues. First, the degree of economic concentration has increased significantly, with a small group of large domestic conglomerates now exercising a high degree of control over the production structure. Second, the share of multinational corporations in total industrial output has declined in the past decade. Third, a massive industrial restructuring has occurred in which industries that are relatively intensive in domestic value added, such as the metalworking sector or the production of capital goods, have contracted, while industries that process raw materials—such as steel, petrochemicals, food processing, and pulp and paper—and that are much less intensive in domestic engineering content and value added, have expanded.

In a general way we can consider the present evolution as a reaction to past overemphasis on import substitution industrialization. It simply happens that with a much lower rate of external protection, domestic value added at world competitive standards cannot be sustained in a good number of the mechanical engineering industries that many developing countries attempted to enter during the 1960s and 1970s.

Moreover, it is not just that the (currently) more dynamic sectors (those producing industrial commodities) involve less domestic value added and engineering content but also that in those metalworking and capital goods industries that have managed to survive, success has been attained after a considerable restructuring effort. In most cases firms have increased the import content of their production activities and have moved “backward” toward final assembly of imported components. This indicates that in the new industrial organization scenario, based on the need to be competitive at a global level, firms have opted for less in-house technological effort, minimal product design engineering, and so on.

The point is that the macroeconomic environment and the globalization of the economy provide much less room for the development of the sort of in-house capabilities that Bell and Pavitt are talking about, and that their paper could be enriched by exploring the weight of these forces.

REFERENCES

- Katz, Jorge. 1986. *Desarrollo y Crisis de la Capacidad Tecnológica Latinoamericana*. Buenos Aires: Inter-American Development Bank.
- . 1987. *Technology Generation in Latin American Countries*. London: Macmillan.
- Nelson, Richard R. 1981. “Research on Productivity Growth and Productivity Differences: Dead Ends and New Departures.” *Journal of Economic Literature* 19(3): 1029–64.

- Nelson, Richard R., and Sidney Winter. 1982. *An Evolutionary Theory of Economic Changes*. Belknap Press of Harvard University Press.
- Solow, Robert M. 1988. "Growth Theory and After." *American Economic Review* 78:307-17.

COMMENT ON "ACCUMULATING TECHNOLOGICAL CAPABILITY IN DEVELOPING
COUNTRIES," BY BELL AND PAVITT, AND "TECHNOLOGY GAPS BETWEEN
INDUSTRIAL AND DEVELOPING COUNTRIES," BY PACK

Carl J. Dahlman

I would like to discuss a theme that has not received sufficient discussion here: that taking advantage of underdevelopment is not automatic but requires investment. I have eight points to make in the brief time allotted.

1. A key priority for developing countries is to find effective ways to tap into the large and expanding stock of technological knowledge. This strategy has important implications because it focuses on using knowledge rather than on producing knowledge.

2. There are many ways to tap into foreign knowledge. Howard Pack lists six modes of technology transfer in manufacturing: purchases of new equipment, foreign direct investment, technology licensing, use of nonproprietary technology (including that obtained from purchasers of exports), acquisition of knowledge from returning skilled nationals, and research and development and reverse engineering. Several others could be added, such as copying, foreign publications, trade fairs, data bases, foreign experts, informal linkages with nationals abroad, local education, and training.

None of these methods, however, is costless, even if the source does not charge for the knowledge. They all require investments of time, effort, and resources. Developing economies have used them to different extents and have recorded different returns. To illustrate briefly, both Taiwan (China) and the Republic of Korea have gotten a tremendous amount of technological information from foreign buyers and suppliers. Because foreign buyers are a rich source of technology, the very process of participating in world trade is an important way of gaining access to technology. A significant difference between the two economies is their industrial structure. Korea has large firms that have started to invest in research and development, whereas Taiwan (China) has many small and medium firms that do not invest much in research. Therefore a large, publicly funded infrastructure for supporting research and development plays a more important role in Taiwan. Singapore has followed still another strategy,

Carl J. Dahlman is chief of the Industry Development Division in the World Bank's Industry and Energy Department.

relying heavily on foreign investment, and it has had a very rapid rate of GNP per capita growth and rising wages. In the beginning its attraction to foreign investors—in addition to its location—was its low wage scale. As foreign investment increased, however, wages rose rapidly. Singapore then invested in physical infrastructure and technical training, and as a result it has been able to continue to attract foreign investment in spite of the rapid increase in wages.

The benefits from foreign investment depend very much on the overall incentive regime. Singapore has been well served by foreign investment because its orientation required world-class technology. Many countries in Latin America also relied heavily on foreign investment, but because they had protected, inwardly oriented policy regimes, they often purchased outmoded or obsolete technologies that retarded their industrial competitiveness.

3. Foreign technology is not the end of the story but the beginning. Technology has to be adapted to local circumstances, diffused through the economy, and continually replenished. Acquiring foreign technology is thus not a one-shot operation; it is a complicated and dynamic process.

4. An important point picked up by John Page and Howard Pack is that access to and use of technology depends on the policy environment and the domestic regulatory regime. Many countries focus on developing technological infrastructure or on research, pushing hard on the supply side. But this emphasis may not do much for the economy because it is not linked to the productive sector. Technology to increase productivity must take place within firms, as Martin Bell and Keith Pavitt emphasize in their paper.

How much effort firms expend depends in large part on the extent to which the incentive regime demands improved performance, greater efficiency, and better quality. In addition, participation in world trade creates a virtuous cycle. Firms are forced to compete and keep up to date on new and improved processes and products. At the same time, certain beneficial side effects emerge from this process, including design, technology, and technical assistance from foreign buyers (through, for example, specifications for products, information on new products and technologies, training, and management experience), as well as economies of scale and scope.

5. A country's ability to use resources efficiently and develop rapidly depends not only on its policy regime but also on its technological infrastructure. This framework consists of many institutions and agents—some public, some private, and some jointly held—that provide technological information, extension services, productivity advice, and consulting services, and that include universities, public and private research facilities, and specialized institutions which finance technology. Norms and standards, for example, are becoming increasingly important as countries focus on export markets and have to meet product specifications in different markets. Good standards also help technological diffusion because they implicitly embody technology.

6. Too much of the research and development effort in developing countries is expended in the public sector. Much of it does not address the needs of the

productive sector but is concerned instead with reinventing what has already been developed in industrial countries (or it represents an attempt to do basic research without the budget and the support infrastructure required to make a significant contribution). There is, however, a tremendous stock of knowledge and new research that the developing countries are not tapping effectively. Therefore I would argue for less emphasis on research and development as such and more emphasis on making effective use of the technology that already exists.

7. This point is related to the preceding one: research and development varies according to the country and its phase of development. For example, in the mid-1970s Korea set up public laboratories for research and development. The total share of expenditure on research and development in the economy at that time was about 0.5 percent of GDP, and although the government was trying to interest the private sector, 80 percent of the total was being carried by the government. Today Korea spends almost 3 percent of GNP on research, and 80 percent of it is done in the private sector. The basic reason for this change is that there is much more pressure from foreign firms. As Korea gets closer to the world frontier, industry is less able to purchase technology, and firms have to invest their own resources.

8. One of the most important elements in a country's ability to make effective use of technology is an adequate technical human capital base. How to develop technical human capital, through formal secondary education, vocational education, and on-the-job training, is a critical area for further research. In addition, scientists and engineers have to be trained—and trained in a way that gives them maximum flexibility, because it is difficult to predict what skills will be necessary. Because of the rapid change, some areas of inquiry and production will be dead ends. To train high-level research scientists and engineers is an expensive proposition, not only because of the prolonged formal education that is required but also because training typically requires laboratories, supplies, and assistants as well.

A final point is that we have been talking so far about technology narrowly defined. Bell and Pavitt, in their paper, did consider the organization and management of production systems and looked more broadly at the whole system, from research to design, production, distribution, and marketing. Such a view is important because part of what is happening in industry is a shortening of the cycle from design to production. In addition, competition is no longer based just on low unit costs of production but also on the total cost from research and design to the delivery of the product to the final market at a particular time. Many developing countries are falling behind in some of these new systems of competition. Thus we have to help them plug back into the world market. This means going beyond the technology of production or new products toward a better understanding of supplier networks, distribution systems, access to design, access to technology, and so on—another important area for research.

FLOOR DISCUSSION OF THE BELL-PAVITT AND PACK PAPERS

A World Bank participant asked the speakers what they thought about the past ten years of futile attempts to draft an international code of conduct on technology transfers—an effort that had wasted a great deal of policymakers' time in developing countries.

Keith Pavitt said that it was extremely important to study historical examples to understand how technology develops in countries that enter the international market, but his particular focus was on the development of capabilities in firms. He said that if one did not have intellectual property within firms, it did not matter how marvelous the country's policies were. He would be interested, for example, in the growth of the in-house capabilities of Japanese firms in the 1950s and of Korean firms in the 1970s and 1980s. He considered that to be—implicitly or explicitly—technology policy, and he took issue with John Page (discussant) and Howard Pack on the question. He was worried that people assumed firms succeeded because of government policies. The Republic of Korea and Taiwan (China), for example, were outliers in terms of other measures of success, yet the data on patents show that their in-house technological capabilities had reached world-frontier level. Those capabilities, he said, had emerged for reasons beyond human capital—investment in research and development, quality, and so on—far more quickly than in other countries. Why? What had technology policy been in those firms? Had they advanced technologically simply because of their links to international markets? One could argue the point, but he believed the companies had an explicit technology policy that induced Korea's rapid growth in total factor productivity.

Howard Pack said that his paper was quite explicit on the point that Korea and Taiwan (China) had no technology policy in the earliest years of their rapid growth in productivity. About 1980, however, they set up technology institutes to encourage R&D and made a sustained effort to persuade firms to increase their R&D—indeed, they had minimum R&D requirements for firms. Taiwan also realized that technology transfer was going to be more difficult and had begun to spend heavily on original R&D.

In response to Pack's comment that one way to introduce new technology to a country was to import a certain percentage of equipment, a speaker from the Bank asked Pack to comment on China's technology imports in 1984 and 1985.

This session was chaired by Marianne Haug, director, Country Department III, East Asia and Pacific Regional Office, the World Bank.

China had bought massive amounts of equipment from Japan—even transplanting factories rather than acquiring technology—but the equipment went to large state enterprises and had little effect on productivity. By contrast, town and village enterprises, which had been denied access to raw materials or foreign exchange to buy more equipment, had shown a tremendous increase in productivity. Pack agreed that there was no benefit in buying advanced equipment and using it inefficiently.

One participant thought that Pavitt had ignored the critical distinction between social learning and individual learning. Social learning would include such factors as industry norms, folklore, and culture—the sorts of things that are important where you have a concentration of technology users, as in Silicon Valley in California, for example. Pavitt said that the sort of learning the participant was talking about was different from the conventional definition of human capital. It was specific, experience-based, and often collective—which was true of most technological activity. It was learning about specific production techniques that might apply elsewhere, but not everywhere. It was not learning to count; that was what made it different from what one learned in school.

The same participant thought that Pack's argument was somewhat circular. Pack had said that it was important to promote industrial exports without creating distortions. The participant asked how that fit Pack's explanation that there had been productivity growth because of a reallocation of labor from agriculture to manufacturing. For most countries, the policymaker who says, don't create distortions, typically believes in static comparative advantage. That would mean more agriculture, but agricultural expansion would not lead to a reallocation of labor, nor would it lead to productivity growth. Pack responded that the reallocation of sectors presumably occurs naturally as relative factor prices change, so over time one would get normal processing of agricultural commodities. If one wanted to accelerate that, it could be done with a subsidy.

Pack was asked what he meant when he said that competition was key in Korea. The speaker said that it was not competition in the way the World Bank defined it; the playing field had definitely not been level. Intervention had been specifically at the level of the firm, not even the sector. Pack responded that Korea had clearly been protective at the same time that it used export targets, but he disagreed with the participant about the nature of protection: Korea had been relatively neutral in the sense that effective rates of protection had varied little across sectors, at least in the mid-1970s. He had not looked at more recent estimates, which were shakier. Furthermore, profit margins from exporting were about the same as those in domestic markets, so in fact the situation was relatively close to a level playing field. But Korea undoubtedly used export targets to offset some of the selectivity of domestic protection.

Pack attributed Taiwan's success, in contrast to Korea's, to its 35 percent private saving rate and its relatively small government deficit (indeed, surplus). This had produced an excess of exports over imports, which was why Taiwan had foreign exchange reserves of \$85 billion in 1992. Very high saving rates are reflected

in export surpluses. The Reagan and Bush administrations had not understood that this also works in reverse: that a large government deficit will be reflected in an import surplus, which may consist largely of manufactured goods. So not only microeconomic policy mattered; macroeconomic policy was important, too.

A participant from Brazil asked Pavitt to expand on what he had said about the widening gap between firms that can "use" technology and those that can "change" technology. Brazil had imported technology, its economy was opening, and competitive pressure was increasing. Nevertheless, Brazil had firms that were beyond using technology but were not yet capable of changing technology. He wanted to know more about the possibility of firms' getting stuck somewhere in the middle.

Pavitt said that starting from a base of raw materials did not necessarily lead down blind alleys. As recently as World War II, three Scandinavian countries had built up relatively sophisticated technology along a trajectory based on raw materials. Finland—which was technologically more dynamic than Korea and Japan—had undergone a remarkable modernization, moving from raw materials processing into areas such as processing technology and mechanical engineering. Denmark was an interesting example of upstream and downstream links around food. The Swedish trajectory went back a long way, to such industries as wood, pulp and paper, and mining machinery. Why couldn't that happen in Latin America? Possibly, in the areas in which Latin America was operating (such as textiles), there was strong vertical disintegration between the process technologies used internationally that made technology accumulation more difficult. Perhaps macroeconomic conditions in Brazil made it more difficult.

Jorge Katz (discussant) agreed with Pavitt, adding that in the early and middle 1970s Brazil, India, Mexico, and Argentina had begun exporting their dated technology to other developing countries. Having started on a natural trajectory, some of these countries had moved into technology exporting about ten or fifteen years down the line. Something had happened at the end of the 1970s to disrupt the learning structure that had been developing for twenty years. Katz believed that these countries had begun again on a different basis of comparative advantage and that history would repeat itself in due time, once the process industries they were now expanding stabilized.

Paul David (who had presented a paper in an earlier session on technology) asked John Page, Howard Pack, and Keith Pavitt what they meant by terms that seemed to him to be related. Page had distinguished between a high level of technical efficiency and a high level of technological change—a distinction, David assumed, that had to do with the relationship between average-practice technology and the best-practice frontier. Pavitt had talked about technological accumulation as developing the resources or conditions or capabilities—some or all of those, he was not sure—to manage the generation of, and changes in, technology. David wondered if Pavitt saw the development of technological accumulation as a system property held by firms and whether it included a firm's

ability to move toward the best-practice frontier. Pack, said David, had talked about the reallocation of resources as being an intersectoral phenomenon. David wanted to know if a reallocation of resources was also implicit when firms scrapped old equipment and moved toward new equipment in a sector (and therefore toward the best-practice frontier), or if Pack saw reallocation as being the removal of inefficiencies (since using frontier technologies might not be rational in a particular sector, depending on the spectrum of factor prices and local conditions).

Pack said that he defined reallocation as occurring where marginal productivities or factors differed across sectors because of factor-market imperfections. If an innovation's average total cost was less than the average variable cost in the same sector, he would consider it intrasectoral growth of factor productivity, on the basis of calculated marginal products of labor, not on average products.

Page said that the achievement of total factor productivity required the explicit allocation of resources. He did not think that it was clear at what stage in a country's development the benefits of allocating resources to one activity or the other were greatest. If he were in India now, his major concern would be how to get from minus 2 percent total factor productivity growth to zero or 1 percent growth, and then to international best practice. What Pavitt and Carl Dahlman (discussant) had said was that many gains in developing countries had come from moving average practice to best practice, not from moving forward with best practice.

A speaker asked if, in citing figures of minus 2 percent total factor productivity growth in India, Page was suggesting active subsidies. Page said he was not only suggesting it but had the evidence. All one had to do was raise effective rates of protection in the import-substituting sector for twenty years, and (at the expense of consumers) one could actively subsidize the destruction of resources. Marianne Haug (chair) said she doubted that governments would believe it. Page responded that it was remarkable the extent to which governments in Latin America now do believe it. He said that if he had to pick one intellectual origin for changed attitudes toward protection by Latin American governments, it would not be static resource allocation. It would be the realization that the rates of total factor productivity growth had continuously declined and were in severe danger of becoming negative in most of those countries.

A participant said that Paul Krugman (speaker in another session), in discussing regionalization, had concluded that a gravity model best described what was going to happen with trading blocs and had predicted that there would be three major blocs. What were the implications if that came to pass and if there were barriers to the transfer of ideas among these blocs? Would the way technology was transferred vary between a bloc dominated by Japan, North America, or Europe?

Another participant responded that there was a big difference in the way in which Japanese corporations organized for production. Their approach was much more systemic: if a firm wanted to produce a component, it would some-

how have to become part of the system of codesigning and coproducing that component as the model was changing. The Japanese production system required a different type of interaction. If a country such as Brazil wanted to tap into the Japanese system, it would probably have to induce Japanese automobile producers to come to Brazil, since they were not there right now. So there would be some concerns about being in or out of different systems.

For another thing, continued the speaker, countries differ in allowing ease of access to technology, and they were beginning to look at these differences somewhat systematically. Originally, the United States had been more generous in licensing technologies because the marginal returns from licensing were not considered important. As it became evident that some countries, notably Japan, had become very good at assimilating technology rapidly, there was greater concern about protecting the rents that a license implies. There has since been a tremendous tightening. The Japanese pattern has been to ask a higher price for a much shorter time; Japan seems to have internalized the value of the technology much more rapidly. Developing countries complain that it is difficult to get access to Japanese technology unless one is part of the Japanese system.

A participant asked what would happen to Mexico under the North American Free Trade Agreement? Would technology transfers increase its total factor productivity growth or a particular factor productivity growth? U.S. automotive firms had been operating in Mexico for a number of years with very little increase in total factor productivity outside the narrow field of auto assembly. Katz said that the Mexican automotive industry was changing from an industry directed toward local consumers to one that supplied parts and components for American manufacturers. In the process, the industry had been forced to introduce quality control, which it did not have before. Katz did expect technological accumulation to occur through the restructuring now under way.

Pavitt said that there were clearly regional effects—that Taiwan and Korea were influenced by Japan, Finland by Sweden, and European countries by Germany. His advice to policymakers in Latin America would be to be sure to establish links with the Germans and the Japanese because the Americans were not necessarily the best in many areas of technology. He said that one important difference between Latin America and East Asia was that the Asians had adopted Japanese manufacturing methods, whereas Latin Americans were still using North American systems.

Changing the focus, another participant asked to what extent these lessons were applicable to agriculture. There was tremendous agricultural technology policy in place, said the speaker—many national research organizations and the Consultative Group for International Agricultural Research. Nobody seemed to be attacking these organizations. Yet Pack's paper said that from 1970 to 1980 total factor productivity in agriculture had been negative for developing countries. Was agriculture so different from manufacturing that one had to approach it in a completely different way?

Katz responded that patent protection and the role of transnational corporations were becoming central issues in agriculture. Agricultural technology was gradually changing from a public good generated by state institutes to a private good generated by large transnational corporations. In the Uruguay Round much of the discussion about patent protection had been about this gradual transfer. Countries in which transnational firms had a strong hold on the local production of hybrids and fertilizers would probably improve their diffusion rates if those countries accepted stronger patent protection on hybrids and fertilizers.

A speaker from the World Bank added that the new biotechnology was also causing things to move rapidly from the university to the private sector. But one big difference between agriculture and industry was that in agriculture it had been possible to have more publicly funded generic research projects because there were particular crop varieties adapted to local circumstances. In industry there is more variety in the range of products, so the generic approach made less sense.

Pavitt thought that it was interesting that nobody criticized public subsidies of agricultural R&D. A participant responded that appropriability was not an issue when research was not proprietary.

One participant wanted Pack to elaborate on how different energy prices in North America would affect technology change among different trading blocs. Marianne Haug asked if he would also elaborate on his statement that because of changes in pricing and factor inputs, technological change would go in two directions: toward energy and toward the environment. Pack responded that input-output tables showed a high energy component in some industries and a low one in others. Everything else being equal, one would normally expect a shift of output away from energy use. Intrasectorally, one would expect technologies to become more energy-efficient. Rising energy prices do not necessarily lead to energy-saving technological change, but one generally looks for technology changes that save on any factor of production. The change in BTUs per unit of GNP over the past twenty years had been spectacular, even intrasectorally, he said.

Pavitt added that technological development in the U.S., European, and Japanese automotive industries had been heavily influenced by energy prices over the past fifty years. Energy prices had had serious upstream effects on the quality of component inputs and on the development of high-compression engines (as opposed to inefficient ones).

Another participant said that in looking at environmentally friendly innovations, the World Bank had found that increasing regulation, after a bit of a lag, had prompted many environmental innovations. Regulation would target the environment, and the system would respond. The Bank is beginning to see interesting patterns in trade in which competitive regulatory environments had put some countries far ahead of others in capital goods that related to pollution control.

International Capital Flows to Latin America: What Is the Promise?

Pedro-Pablo Kuczynski

Driven by external and internal factors, capital inflows to Latin America rose sharply from about 1 percent of gross national product (GNP) during most of the 1980s to more than 3 percent in 1990. Given the size of external debt owed by the region and the amount of capital still held abroad, the prospects for sustained investment flows depend to a significant extent on global interest rates, the development of domestic fixed-income capital markets, and political and social reforms. The instability that stems from the failure of existing political institutions to redirect new-found saving into improved economic conditions for low-income groups—especially in urban areas—could not only reduce foreign capital inflows but also encourage domestic savings to move abroad once again.

Without a doubt, Latin America is the region of the developing world where the most dramatic change in capital flows has taken place since the start of the 1990s. Only the provinces of southern China near Hong Kong, where massive capital from Taiwan (China) and Hong Kong has gone, can rival Latin America in the magnitude of change. Net capital inflows into Latin America rose sharply from about 1 percent of GNP during most of the 1980s to 2 percent in 1990 and well over 3 percent in 1991.

Net capital inflows to Latin America, which had stagnated at an annual average of about \$9 billion a year following the debt crisis, rose from \$9.6 billion in 1989 to \$18.4 billion in 1990 and doubled again to about \$36 billion in 1991 (ECLAC 1991), or about 3.5 percent of GNP. The sharp upward trend has continued in 1992. Although Mexico has accounted for half of the improvement, the remainder has been spread fairly uniformly throughout the region, even in such countries as Brazil and Peru, which are far from achieving reasonable price stability. Capital has been particularly attracted to financial assets, sparking a boom in regional stock markets, including the most dormant ones, and a decline in domestic interest rates. This trend, combined with direct foreign investment (especially in Chile and Mexico), has helped revive economic growth.

Pedro-Pablo Kuczynski is president and chief executive officer of Nueva Management Inc.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

The sources of this shift are both external and internal. Externally, the sharp drop in short-term dollar interest rates has sparked a reflow of capital from "Latin America's bank": the \$200 billion to \$300 billion of private funds held by Latin American residents abroad. Internally, deregulation, privatization, and the opening up of foreign trade and investment have stimulated the inflow of new capital. While "hot money" does pose problems, rising international reserves have strengthened currencies, lowered inflationary expectations, and assured investors that inflation will not devalue their investments. Given the size of the external debt and the amount of capital still held abroad, the prospects for continued capital inflows depend to a significant extent on what happens to world interest rates. If part of these inflows is to be channeled into investments in essential infrastructure, a determined effort is needed to develop domestic fixed-income capital markets, which have been stunted in the past by high inflation. Political and social stability are further prerequisites. The instability that stems from the failure of existing political institutions to redirect new-found savings into improved economic conditions for lower-income groups—especially in urban areas—could not only reduce foreign capital inflows but also encourage domestic savings to move abroad once again.

The financial results of this boom in capital inflows have been a sharp appreciation in financial and real estate assets and a buildup of international reserves, similar to what occurred in Spain in the second half of the 1980s. Today, official international reserves in Latin America in the aggregate are about \$44 billion—the highest nominal level ever recorded, and equivalent to about five months of imports.

This paper provides a brief overview of these rapid changes, which began in Chile three years ago, and inquires whether we are witnessing a cyclical phenomenon or a secular change. If it is the latter, and the roots of the change are deep enough, we can prudently think of a return to the fairly steady (although maldistributed) 6 percent annual growth that most Latin American countries enjoyed in the 1950s and 1960s and through most of the 1970s.

I. THE INTERNATIONAL SETTING AND EXTERNAL INFLUENCES

The U.S. dollar is the currency in which Latin America's external trade is priced and counted. Because of high Latin American inflation in the 1980s, the dollar has become the *de facto* currency for pricing domestic transactions—and often the actual medium of exchange as well. As a result, the dramatic drop in short-term interest rates in the United States in 1991 and 1992 (table 1) has had a major impact on capital flows to and from Latin America. First, because the bulk of external debt owed to commercial banks is still based on floating rates, debt service declined rapidly during the past year, signaling investors that exchange rates would firm up and that their investments would not depreciate in dollar terms as in the past. Second, declining yields on the funds held abroad by Latin American residents (most of which are probably held in U.S. dollar instruments) began stimulating the return of so-called flight capital.

Table 1. London Interbank Offered Rates on Three-Month U.S. Dollar Deposits, 1980-92
(averages for period shown)

<i>Year</i>	<i>Percent</i>
1980	14.2
1981	16.9
1982	13.3
1983	9.8
1984	10.9
1985	8.4
1986	6.9
1987	7.2
1988	8.0
1989	9.3
1990	8.3
1991	6.0
1990:4	8.2
1991:4	5.1
1992, October 30	3.5

Source: IMF (various years).

An important part of what has happened in Latin America in the past two years thus has its roots in what has happened abroad. The increase in private capital inflows to such countries as Peru, which were still suffering significant financial problems, suggests that external forces affecting the supply of funds, driven by sharply lower interest rates in dollar markets, were very powerful. At the same time, there is no doubt that the expectation of greater domestic stability and a possible economic turnaround, combined with attractive interest rates in most Latin American markets (table 2), provided a demand magnet.

In addition to a decline in U.S. interest rates, the other important external force that has affected Latin American economies in the recent past has been the cyclical decline of export prices as a result of the recession in the United States as well as the incipient recession in other industrial countries. This has affected purchasing orders, to which commodity prices are especially sensitive. The latest

Table 2. Interest Rates for Selected Countries in Latin America, 1990 and 1991

<i>Country</i>	<i>1990</i>		<i>1991^a</i>	
	<i>Interest rate</i>	<i>Devaluation against U.S. dollar</i>	<i>Interest rate</i>	<i>Devaluation against U.S. dollar</i>
Argentina	1,586.0	67.9	60.0	44.1
Brazil	560.7 ^b	93.3	475.3 ^c	84.1
Chile	40.3	11.8	22.3	10.0
Mexico ^d	34.8	10.3	19.3	4.1
U.S. dollar: three-month LIBOR	8.3	—	6.0	—

— Not available.

a. Estimated.

b. Average of last three quarters of 1990, reflecting Collor Plan I.

c. Average of first three quarters of 1991, reflecting Collor Plan II.

d. Twenty-eight-day CETES (Treasury bills).

Source: IMF (1992).

Table 3. Index of Merchandise Terms of Trade for Latin America, 1988–91 (1980 = 100)

<i>Region</i>	1988	1989	1990	1991 ^a
Total Latin America	78	78	77	73
Oil-exporting countries	64	70	73	76
Other Latin America	91	87	81	81
South America, non-oil-exporting	92	86	81	80
Central America and Caribbean	96	96	89	90

Note: Exports and imports f.o.b.

a. Estimated.

Source: ECLAC (1991).

decline in commodity prices comes on top of the similar decline in the terms of trade that took place during most of the 1980s (table 3). The drop was, of course, driven by the sharp drop in oil prices after 1981, which affected exporters as diverse as Bolivia, Colombia, Ecuador, Peru, and, especially, Mexico and Venezuela. Altogether, the terms of trade for Latin America fell 27 percent in the decade 1981–91.

This drop, however, seems to have had a milder effect than previous commodity cycles, when economies were less diversified. In Chile, Colombia, and Mexico, for example, the dependence on exports of one major export product (copper, coffee, and petroleum, respectively) fell to about one-third of merchandise exports, compared with about one-half five years earlier. Obviously, the drop in international prices is an important factor. For example, the current international price of coffee in constant dollars is a third of its 1970 level. But a more important influence has been the tremendous growth in the volume and diversification of exports as a result of the forced devaluations caused by the debt crisis.

From 1980 through 1991 merchandise exports from South America and Mexico rose 85 percent in volume,¹ although their total value increased by only 40 percent (ECLAC 1991, table 9). It is worth considering the argument that the very success of the highly indebted Latin American countries in “exporting their way out of debt”—a prescription often recommended by worried creditors in the early and mid-1980s—contributed to the weakness of the prices of their exports. Although this assertion would probably not be true for petroleum, where Latin American exports are not a central influence on world markets, it probably holds for a number of mining and agricultural products, as well as for certain commodity-intensive manufactures such as steel and basic chemicals.

The upshot of these major changes in the international context as it affects Latin America is that capital movements have become increasingly important—and changes in merchandise trade less so—because of the large size of the external debt in relation to trade flows. At an estimated \$426 billion at the end of 1991, the debt amounts to 3.5 times annual merchandise exports, or 2.8

1. Exports from Central America, except those from Costa Rica, rose only 10 percent because of civil strife.

times exports of goods and services (table 4). Even though the external debt has been constant in nominal terms since 1987 and thus has been declining in relation to exports, the still-high current levels of debt mean that a significant movement in interest rates can have a very large impact on foreign exchange flows. In contrast, the normally expected size of change in export prices has far less impact. The good side of this outcome is that Latin America's export sector, the engine of growth in the past, is perhaps more insulated from economic recession than in the past; the bad side is that the capital account is highly dependent on the international movement of interest rates. This is all the more so because of the large amount of private Latin American assets held abroad.

Latin America's Bank

The cloud of past capital flight does have a silver lining, however; today, most Latin American countries have available a substantial pool of capital held abroad that acts as Latin America's "bank." The current income from this pool—and, increasingly, some of the principal—is likely to return as investors see higher relative yields at home combined with a perception of greater safety.

Estimates of capital flight are by their nature imprecise (see Lessard and Williamson 1987; Kuczynski 1988). The very idea of capital flight conveys a moral judgment implying that the savings should stay at home. Such judgments are rarely based on an analysis of why capital moves: safety and higher returns elsewhere are the basic and justifiable motives. The general assumption is that poor countries should keep their resources at home and rich ones should transfer them to the less fortunate. Thus in the mid-1980s it was argued that Japanese surpluses "should" be invested in the largest deficit country, the United States. These virtuous capital "movements" were contrasted with evil capital "flight" from debt-burdened Latin American countries (see, for example, Morgan Guaranty Trust 1986). In reality, it is not surprising that Latin Americans, even small savers, would buy U.S. dollars to secure high interest rates when they faced inflation at home and could see that most of the capital inflow from international agencies and commercial banks was being used to service the debt.

Table 4. Outstanding External Public Debt and Exports of Goods and Services, Latin America, 1986–91

<i>Item</i>	1986	1987	1988	1989	1990	1991 (estimate)
External public debt (billions of dollars, current prices)	400	428	421	421	435	426
Exports of goods and services (billions of dollars, current prices)	95	108	123	136	151	152
Ratio of debt exports to goods and services (percent)	421	396	342	306	288	280
Ratio of interest due to exports of goods and services (percent)	36.7	30.4	29.0	28.6	24.8	22.3

Source: IMF, *World Economic Outlook*, April 1992.

Measuring capital flight is complicated not only because it involves a moral judgment but also because there are technical difficulties. When a Latin American company deposits funds abroad to buy spare parts, is it considered capital flight? Is a positive errors-and-omissions item in the balance of payments necessarily capital reflow, or could it represent, for example, undervaluation of merchandise exports because of a commodity export tax, common in most Latin American countries?

Estimates of autonomous capital movements abroad are thus inherently imprecise. At best they can give an idea of a range or band. Table 5 gives estimates of outstanding amounts probably held abroad as of 1984, the last year of large-scale private capital transfers abroad. The 1984 capital stock is then adjusted upward to take into account interest earned until 1990 and further private capital outflows (excluding debt service and profit remittances) from 1984 to 1990. Most of these outflows, about \$15 billion to \$20 billion, were from Brazil. It is assumed that the bulk of the assets held abroad were liquid dollar investments, mostly bank deposits and high-quality bonds, and that holdings of real estate and equities were small. This profile is presumed to resemble the average Latin American investor seeking safety rather than high returns abroad. Interest is assumed to be the London interbank offered rate (LIBOR). Overall, these are rather conservative assumptions, especially for the late 1980s (a period when investors had gained experience), and they probably underestimate the total stock of capital in 1990.

The potential earnings on this pool of approximately \$200 billion to \$300 billion are substantial, amounting to between 1.5 and 2 percent of GNP in the region, and can contribute significantly to financing the region's capital needs. From 1984 through 1989, when average net capital inflows were about \$9 billion annually (and consisted largely of official flows), the interest earned on private Latin American capital held abroad was about \$15 billion a year. A shift of this magnitude toward the original source of the capital—the indebted coun-

Table 5. *Estimated Amounts of Capital Held Abroad in "Latin America's Bank"*
(billions of U.S. dollars, current prices)

Country	End 1984		Appreciation 1985-90 (LIBOR cumulative, percent) ^a	End 1990	
	MGT	IMF		MGT	IMF
Argentina	25	24	59.0	40	38
Brazil ^b	17	12		44 ^b	36 ^b
Mexico	53	38		84	60
Venezuela	30	25		48	40
Other ^c	25	20		40	32
Total	150	120		256	206

Note: MGT, Morgan Guaranty Trust.

a. Estimate based on three-month LIBOR: 1985, 8.4 percent; 1986, 6.9 percent; 1987, 7.2 percent; 1988, 8.0 percent; 1989, 9.3 percent; 1990, 8.2 percent.

b. Includes \$15 billion additional outflow in 1988-90.

c. Author's estimate.

Sources: For 1984, Lessard and Williamson (1987, pp. 52-67); for 1990, author's estimates.

tries of Latin America—could obviously make a big difference to the prospects for the region's economic growth. Net foreign capital inflows were conspicuously absent during most of the 1980s, averaging less than 1 percent of GNP a year in spite of a major effort by the International Monetary Fund in 1982–83 and by the multilateral development banks thereafter.

The repatriation of the income earned on assets held abroad can also restore the importance of the market for capital flows and investment decisions. Capital will not return to investments that are perceived as unprofitable. For example, public utilities, which are being privatized currently in Argentina, Mexico, and Venezuela (and were privatized several years earlier in Chile), with the help of domestic and foreign capital, must be able to fund part of their future growth themselves if they are to attract private capital. That was certainly not the case in the 1980s, when state loans financed most of the public utility investments, leaving most state-owned utilities with an excess of debt and inadequate cash flows to cover maintenance and essential operations (OLADE 1992).

Moreover, although there are no data on the ownership of the pool of capital abroad, historical experience suggests dispersed ownership with both large and small investors and savers. There is thus likely to be market diversity in decisionmaking.

Finally, the funds in Latin America's "bank" will, over time, tend to tilt toward investments in equities rather than in debt as the interest differential between domestic debt and international bonds narrows with the easing of fiscal pressures in Latin American countries. (As table 2 shows, this has already begun to happen in Argentina, Brazil, Chile, and Mexico.) The drop in domestic interest rates has coincided in all these cases, as might be expected, with a sharp price rise in domestic equities (table 6). In 1991 the market capitalization of the largest stock exchanges rose 171 percent in dollar terms. Even though the still-narrow domestic equity markets appear to be fully priced, this is only so if one compares today's equity prices with current earnings, which remain depressed. If the comparison is made with potential future earnings, the prospects for a further rise in domestic equity markets—and a continued flow of capital—are good as long as real interest rates are falling.

Table 6. *Short-Term Interest Rates and Equity Markets, 1991*

Country	Decline in short-term domestic interest rates (percent) ^a	Rise in stock market index (percent) ^a	Average price-earnings ratio ^b
Argentina	1,502.2	392	38.9
Brazil	76.2	152	1.5
Chile	16.2	90	17.4
Mexico	12.7	103	14.6
Venezuela	0.4	34	30.5
United States	2.3	31	25.8

a. Converted to U.S. dollars.

b. Year-end prices and estimated earnings.

Sources: Table 2; IFC (1991).

"Hot" Money

It is hard to quantify the amount of "hot" money (capital that seeks quick returns, mainly in fixed-income liquid instruments that carry high interest rates measured in the investor's currency) flowing to Latin America. Hot money was important at the start of the process of recovery in Mexico (1989-90) and Argentina (1991). For Latin America as a whole, however, in 1991 most inflows took the form of direct foreign investment, portfolio investment in equities, and medium-term borrowing in international capital markets.

Nonetheless, the suddenness and size of capital reflows has created problems of monetary management in countries as diverse as Argentina, Colombia, Mexico, and Peru. The sequence is a familiar one: as foreign portfolio investment moves in and foreign exchange reserves rise, central banks, facing a fiscal policy that continues in varying degrees to be loose, begin to tighten monetary policy. Domestic interest rates then stay above the level they would otherwise be at, and so does the exchange rate, despite continued inflation. (No major Latin American country has yet reached single-digit annual inflation.) This policy generates high dollar returns, which in turn generate more capital inflows, and so on.

The virtue of this process is that it is anti-inflationary. Since the exchange rate is the single most important price in the economy, an appreciation generates confidence about future price stability in spite of the large expansion in money supply. This is, for example, what happened in Taiwan (China) after the mid-1980s; an increase of more than 20 percent a year in the money supply coincided with a 3 to 4 percent decline in inflation. Central banks in Latin America may say they dislike hot money, but they like the increase in international reserves and the anti-inflationary effect of higher exchange rates. So do finance ministries, because the fiscal cost of external debt service declines. Treasurers of large companies certainly like it, because their cost of funding declines and their possibilities of cheap equity financing improve.

Then, who dislikes hot money? Presumably, import-competing industrialists, exporters, and those making foreign direct investments. Nevertheless, up to March 1992 there had been no slowdown in direct foreign investment or in exports. In 1991 there was a sharp slowdown in the growth of the value of merchandise exports (ECLAC 1991, table 14), but that slowdown was almost entirely explained by the recession-induced decline in prices on international markets. In fact, the strong 6 percent growth in export volume has continued even in countries where the purchasing power parity exchange rate has appreciated in comparison with the dollar. For example, in Chile, where domestic wholesale prices have risen about 90 percent since 1988 while the exchange rate has depreciated 52 percent against the dollar, export volume grew 11 percent in 1991. In Mexico exports grew 9 percent. The explanation of this contradiction lies in a combination of factors: wage increases have not kept pace with inflation; productivity has increased as idle capacity is put to use; and currencies were substantially undervalued to begin with.

As economic recovery gradually exhausts these factors, export growth could falter, as, for example, in Korea after very large wage settlements in the late 1980s. For the time being, however, that kind of problem is some time off in most Latin American economies.

II. FOREIGN DIRECT INVESTMENT

Foreign direct investment in Latin America stagnated during the 1980s, averaging about \$5 billion a year in nominal terms, and reached its lowest level (\$2.5 billion) in 1986. That was also the turning point for Mexico, where foreign direct investment had declined steadily as a result of the debt crisis and the decision to nationalize banks and foreign currency deposits. The stringent fiscal efforts of the de la Madrid administration began to pay off. Despite the major earthquake the previous year and a sharp drop in oil prices at the beginning of the year, by 1986 investors saw Mexico as a place where assets were cheap and where a turnaround was possible. Erratically at first, divestitures by foreign companies stopped, and new investment started to rise. By early 1991, when Southwestern Bell and France Telecom paid \$850 million for a 49 percent share of the control block of *Teléfonos de México*, foreign direct investment had climbed to about \$5 billion, about 40 percent of the Latin American total and, in real terms, slightly higher than the level recorded ten years earlier (table 7).

Two factors were important for this renewed confidence in Mexico: the perception of future economic recovery and the elimination in 1991 of most restrictions on foreign ownership—a step that had been taken by Chile several years earlier. Restrictions continue to limit the participation of foreigners in banks, airlines, and media companies, but these restrictions are similar to those in the United States.

In comparison with the 1970s and early 1980s, the most interesting change in new foreign investment in Latin America has been its concentration in markets that are rapidly opening up and are increasingly exposed to foreign competition. This is true for Brazil and especially for Mexico, where the average external tariff has fallen to about 13 percent and is due to fall further and where the natural protection of distance is limited, at least as far as competition from the United States is concerned. In that sense the current pattern is quite different from that of the 1960s and 1970s, when the objective of foreign direct investment was to gain a foothold in a country protected by high tariff walls and quotas in order to maintain market share. Now the objective is to be positioned for a period of high domestic growth, especially in such industries as consumer durables where there is long-repressed demand. Such investment also anticipates an export market in industries where labor costs are important and where there are economies of scale because of a potentially large domestic market—for example, in automotive production in Mexico and in the proposed “Mercosur” free trade area dominated by Argentina and Brazil.

Table 7. Net Foreign Direct Investment by the Private Sector, 1981-90
(millions of U.S. dollars, current prices)

<i>Recipient country</i>	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991 ^a
Argentina	944	258	183	268	919	574	-19	1,147	1,028	2,036	2,400
Brazil	2,313	2,534	1,373	1,556	1,267	177	1,087	2,794	744	65	1,200
Chile	362	384	132	67	62	57	97	109	259	587	1,100
Colombia	228	337	514	561	1,016	642	293	165	546	339	400
Mexico	2,835	1,665	461	390	491	1,523	3,346	2,594	2,241	2,633	5,000
Venezuela	184	253	86	-3	57	-444	-16	21	77	451	600
Total	6,866	5,420	2,749	2,839	3,812	2,529	4,688	6,830	4,895	6,111	10,700
Total Latin America	7,971	5,878	3,257	3,098	4,113	2,801	5,358	5,578	5,449	6,685	11,500

a. Author's estimates.

Source: IDB (1991).

Privatization has played a major role in stimulating direct investment. The process started with debt-equity swaps in Argentina, Brazil, Chile, and Mexico,² but subsequent privatizations have involved cash payments. For larger companies in countries with active stock markets (Argentina, Chile, Mexico, and Venezuela), the pattern has been to sell a controlling block through competitive bids and actively promote the public flotation of the rest on domestic and international markets. Privatization thus has implications for both direct and portfolio investment, but, more important, it is a symbol for investors that the state has shed its role as the primary capitalist and dispenser of economic favors.

Whether this form of "concessionaire capitalism" is really on the wane in Latin America will in the end be fundamental to the question of whether Latin American economies are to play a significant role in the international economy. Any major change in ingrained attitudes will take time. Reforming the role of the state does not mean that there will not be big private-sector capitalists; the point is that there are new self-made capitalists. In Mexico, for example, the reprivatization of commercial banks has attracted as many new bidders as established names. The economic stress of the 1980s has bred a new class of Latin American entrepreneurs, many of them relatively young, who took risks at a time when assets were cheap and who are now reaping the rewards. A whole new class of managers is also emerging: U.S. corporations may be having second thoughts about hiring M.B.A.s from top universities, but in Latin America a new generation of Ph.Ds and M.B.A.s is rapidly changing insular attitudes in private and government enterprises.

III. CAPITAL REQUIREMENTS FOR INFRASTRUCTURE

The 1980s left most Latin American and Caribbean countries with lagging infrastructure. Capital inflows from abroad may gradually overcome this gap in some areas, particularly telecommunications, but perhaps not in others, such as water and sewerage. The supply of electricity has reached a crisis in many countries (OLADE 1992). Unless the necessary infrastructure is speedily established, economic growth and the prospects for further rapid growth of foreign capital inflows will be affected.

The telecommunications industry has attracted a great deal of private capital for a variety of reasons. First, deregulation (particularly in the United States and Britain) and mature markets (in the industrial countries) have led telephone companies to look to underserved markets. Whereas in the United States, for example, each telephone line serves two inhabitants, in Latin America the aver-

2. Debt-equity swaps in privatization are a way of making the price appear higher because the price paid is expressed in terms of the nominal value of the loans purchased, which is usually much larger than the cash price paid to purchase the loans. There is, of course, a reduction in external debt. From a financial standpoint the question is whether this reduction can be accomplished at less cost through a central debt repurchase program or through individual transactions. From a legal standpoint, banks have been reluctant to agree to repurchase programs except through agreements under the Brady Plan or through waivers for specific privatizations.

age is one line for every twenty-two residents. (Pakistan, which has one of the most profitable telephone monopolies in the world, has one line for every 110 residents.) The prospects for strong future growth make these companies attractive candidates for privatization. Privatization of Brazil's telephone service (the largest in Latin America), if properly structured, will attract at least \$10 billion in outside capital in the years ahead and comparable domestic capital through the local stock market.

The prospects for electricity-generating facilities are, unfortunately, less clear. The coverage is better than that in telephones, so further growth prospects are less exciting. But more important, except in Mexico and the Caribbean, the big investments in South and Central America probably need to be in hydroelectric power plants, which require terms that the market cannot provide (as the privatized Chilean generating companies are finding out). The choice, given limited capital, will therefore be either to rely on private capital for distribution of electricity and on thermal power plants (with their attendant adverse environmental consequences) instead of hydroelectric ones, or to find a formula that grants official lenders a larger role than in the past decade. Given the demands of Central and Eastern Europe as well as China, the latter is unlikely.

Electric power plants require heavy debt financing, which the international bond markets can help provide for the larger Latin American countries. Of the net capital inflows to Latin America in 1990 and 1991, about 30 percent came from the international bond market, and these funds went primarily to Mexico and Venezuela. In 1991 Argentina and Brazil regained access to the market as investors anticipated a settlement of bank debt sometime in 1992. The international bond market, however, is essentially a medium-term lender in the five-to-seven-year maturity range and requires government guarantees for substantial longer-term financing—guarantees that private electric companies cannot provide. This is also true of domestic bond markets, which are at an incipient stage in all countries except Chile.

The financial requirements for basic infrastructure in electricity, water, and sanitation will therefore continue to require substantial government involvement. Private capital, although welcome in some countries, will need assurances and continuity from governments, a permanent and credible reduction of inflation to single-digit levels, and domestic capital market reforms that eliminate the preferential tax position of governments as issuers in domestic capital markets. The freeing of domestic capital markets, particularly bond markets, is as important as the liberalization of foreign capital that has occurred in recent years—if not more so.

For domestic bond markets—in addition to the obvious tasks that are already high priorities, particularly reducing inflation and budget deficits—governments need to consider eliminating the tax-free status of government bonds, improving the financial health of social security systems, and modifying traditional credit allocation policies that favor government. These steps, which are now under way in several countries, would stimulate reflows of

capital from abroad and help channel necessary funds to commercial infrastructure.

IV. CONCLUSION

Is the recovery in Latin America simply a cyclical fad? Certainly, the recent growth of foreign capital inflows cannot continue at the torrid pace of the past two years; it would be associated with large trade imbalances that would renew calls for protectionism. In any case, the growth of capital reflows is likely to slow down simply because of global events: as the U.S. economy begins to recover, interest rates will rise, and the attraction of quick returns based on large interest rate differentials will diminish.

There are, however, forces at work that argue for sustained economic growth and continued large capital inflows to the main Latin American countries. These include profound institutional and policy changes in economic management, already evident in Argentina, Chile, Colombia, Mexico, and Venezuela and, to a lesser extent, elsewhere. Even in large countries, such as Brazil and Peru, that do not yet have a full reform program in place, the discussion about economic objectives has reached a fairly broad public consensus. Furthermore, assets in Latin America are relatively cheap in international terms and in relation to their potential earning power. In some countries the potential—for example, in real estate, pension funds, money management, consumer credit, and so on—has barely been scratched. Finally, the recovery is taking place at a time when the industrial countries are in varying degrees of recession: as that cloud lifts, the potential for an expansion of exports—an important element in Latin American economic growth in the past—will rise significantly.

If there is an Achilles' heel to the improving outlook, it is the distribution of the benefits of growth. Although there is honest disagreement about what the main objective of restructuring the state should be, there is little doubt that the resources being saved should be redirected to increasing employment and improving basic services for the lower- and middle-income groups that bore the brunt of the recession of the 1980s. This is not so much a matter of state resources as of basic government policy. Otherwise, the already highly uneven income distribution in most of the countries of the region will deteriorate even further and will eventually lead to political instability.

REFERENCES

- ECLAC (United Nations Economic Commission for Latin America and the Caribbean). 1991. *Balance Preliminar de la Economía de América Latina y el Caribe*. Santiago.
- IDB (Inter-American Development Bank). 1991. *Economic and Social Progress in Latin America*. Washington, D.C.
- IFC (International Finance Corporation). 1991. "Quarterly Review of Emerging Stock Markets." Washington, D.C.
- IMF (International Monetary Fund). Various years. *International Financial Statistics*. March. Washington, D.C.

- Kuczynski, Pedro-Pablo. 1988. *Latin American Debt*. Baltimore, Md.: Twentieth Century Fund and Johns Hopkins University Press.
- Lessard, Donald R., and John Williamson. 1987. *Capital Flight and Third World Debt*. Washington, D.C.: Institute for International Economics.
- Morgan Guaranty Trust. 1986. *World Financial Markets*. March. New York.
- OLADE (Organización Latinoamericana de Energía). 1992. *América Latina y el Caribe; El Déficit de Financiamiento de la Inversión Eléctrica en la Presente Década: Posibles Soluciones*. Quito.

COMMENT ON "INTERNATIONAL CAPITAL FLOWS TO LATIN AMERICA:
WHAT IS THE PROMISE?" BY KUCZYNSKI

Albert Fishlow

Pedro-Pablo Kuczynski has written informatively about the recent increase in foreign investment in Latin America and has identified an important change in international capital flows in the past two years. For the first time since the debt crisis, Latin America has returned to market favor. Inflows have multiplied from \$9.6 billion in 1989 to \$41 billion in 1991, with upward growth continuing into 1992. The most favored recipients have been Mexico and Chile, but even Argentina and Brazil have seen large entries in recent months. The issue is whether such flows can be expected to continue, and under what conditions. External capital could contribute to a revival of Latin America's economy after a decade during which lack of such capital sharply reduced economic performance.

On this fundamental question of future foreign investment, Kuczynski answers on the whole affirmatively. First, he sees "profound institutional and policy changes in economic management" in the region; second, he finds that assets in Latin America are "relatively economical in international terms and in relation to their potential earning power"; and finally, he considers that when the recession ends, "the potential for export growth, also a major element in past Latin American economic growth, will rise significantly."

These are important factors. But we may be exaggerating them in relation to shorter-term market conditions that yield a very positive return to a financial commitment to Latin American countries. Simply put, short-term U.S. interest rates are at very low values as a result of recession. Conversely, rates in Latin America are—and have been in recent years—highly positive as a consequence of stabilization efforts.

But this differential cannot be sustained on either side. U.S. rates can be expected to rise with economic recovery, and Latin American rates will have to come down if investment rates in the region are to recover to past levels and permit sustained expansion. Furthermore, some of the capital flow has been motivated by rising stock market valuations in 1991 that offered extremely high returns, and this circumstance, again, cannot be simply extrapolated. In recent

Albert Fishlow is dean of International and Area Studies and Class of 1959 Professor of Economics, University of California, Berkeley.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

weeks there have been signs of concern in several countries. Speculative commitment is not what is needed; long-term commitment to financing higher growth is.

In addition, U.S. policy should be credited with a major role in stimulating foreign capital inflows to Latin America. Certainly, the anticipated North American Free Trade Area requires explicit recognition. Mexico experienced a very large capital flow—almost half of the regional total in 1991—in anticipation of the successful conclusion of negotiations with the United States and Canada. That treaty is expected to provide a stimulus to continuing foreign investment in subsequent years. Mexico sees such flows as essential to the resumption of growth rates of 5 to 6 percent a year. But election of the Democratic party's presidential ticket in November, as the polls are suggesting, together with a widened majority in the Congress, could mean that some renegotiation of the treaty will be necessary. New uncertainties could reduce Washington's foreign commitments and put more pressure on increased domestic savings to finance Mexico's investment. There might also be a comparable reaction to a projected free trade agreement with Chile. In short, we may be seeing a blip rather than a trend.

Third, and with similar consequences, maintaining capital flows at current levels may well require more evidence of a full Latin American commitment to sustained export growth. Kuczynski's identification of export growth as a major element in past Latin American growth is questionable at best. From 1967 to 1982 export volume in the region increased only 2.4 percent a year, against a developing country average of 3.5 percent and a much higher 9.4 percent for Asian countries. Between 1983 and 1991 International Monetary Fund export figures show an improvement to 4.7 percent for Latin America versus a developing country average of 4.1 percent, but Asia's average increased to 10.5 percent. Indeed, if one excludes 1990 and 1991, which incorporate strong declines in Eastern Europe and the former U.S.S.R., the Latin American average of 4.9 percent falls well below the general average of 6 percent. The reality is that Latin America has not caught up to the Asian—or even the average—level of export volume. The region needs to pursue an aggressive exchange rate policy in favor of guaranteeing exports rather than continuing to overvalue as capital flows are attracted. The latter tendency was a key deficiency during the 1970s and early 1980s when the debt crisis emerged. One hardly wants a resumption of the experience of the past decade even before its remaining effects have been erased.

Fourth, a better mix of domestic private and public investment is likely to be necessary for sustained Latin American growth. But foreign flows, unlike the situation in the 1970s and 1980s, are now directed to private sector initiatives and, increasingly, to privatizations. Public investment lags far behind, thereby creating a potentially serious future problem. The commitment to education, health care, and other necessities has shrunk as attention focuses on the budget and the need to eliminate deficits. The short-term consequences can be positive, but one must worry about longer-term implications. Unless the state is capable

of establishing a long-term equilibrium, growth will not be continuous as real needs begin to be felt.

In sum, the new flow of external resources into the region may be less a reward for better Latin American policy and more a response to a combination of short-term forces that favor these flows. That would be truly unfortunate, for few areas so much need to increase investment and growth. But even these alone are not enough. As Kuczynski correctly recognizes, the question of income distribution lurks in the background. Most evidence suggests that the 1980s have seen a further deterioration in what initially had been relatively high inequality. If the new reliance on foreign investment is to prove feasible, serious public policy will have to deal with this problem. That task of providing services to the deprived majority of the population cannot be postponed until after the next round of expansion. States will have to respond now to the immediate need. Latin American countries face a very difficult future as they try to achieve the requisite social, political, and economic equilibria simultaneously.

COMMENT ON "INTERNATIONAL CAPITAL FLOWS TO LATIN AMERICA:
WHAT IS THE PROMISE?" BY KUCZYNSKI

Mario I. Blejer

This paper describes and analyzes one of the most remarkable indications that the debt crisis of the 1980s is reaching its end: the notable increase in capital inflows to Latin America that started about two years ago and that is now reaching important quantitative proportions. In Pedro-Pablo Kuczynski's view this development reflects both "push" and "pull" factors: the push results from a deterioration in the relative return to capital—particularly financial capital—in the United States, and the pull comes from the improvement in economic conditions and financial markets in Latin America. In addition, Kuczynski mentions what he calls "Latin America's bank"—the \$200 billion to \$300 billion in private Latin American capital held outside the region.

Kuczynski gives equal time to both push and pull factors. In my view, however, domestic developments in Latin America are the overwhelming reason for the reversal, and financial conditions in the United States are less important. Indeed, even if financial conditions in the United States had deteriorated further, very little capital would have been directed to Latin America if the region had not undertaken significant shifts in economic policies.

Extraordinary changes have been recorded in the economic performance of the Latin American countries—particularly Chile and Mexico—that have received most of the capital inflows in the past two years. These changes include large reductions in—or the elimination of—fiscal deficits, a strengthening of monetary and credit discipline, and cuts in the economic activity of the public sector, as well as a liberalization of trade and foreign exchange controls that facilitates the international movement of capital. Because these economic reforms have spurred the reflow of capital to the region, it would be worthwhile to treat this issue in a less global and aggregate manner and to try to find a specific relationship between good policies (or the expectation that good policies are to be implemented) and capital inflows.

As for the push factors, I am not convinced that lower interest rates in the United States explain much of the phenomenon. Lower interest rates do reduce the debt burden, with the result that adjustment, economic performance, and

Mario I. Blejer is senior adviser, Europe and Central Asia Region, the World Bank.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

confidence in the government are strengthened. But even though the ratio of debt to exports in the region declined in 1991, the Inter-American Development Bank notes in its annual report that the ratio of debt *service* to exports increased from 26 percent in 1990 to 30 percent in 1991. The direct impact of international interest rates on the debt burden is therefore ambiguous.

The implicit argument in Kuczynski's paper is that the decline in U.S. interest rates reduces or reverses the interest differentials (adjusted for exchange rate fluctuations) between the United States and Latin America and therefore makes investment in the region more attractive. But it seems clear that the main reason that the difference in *expected* returns between U.S. and Latin American investments has shrunk is that there is less uncertainty in Latin America; U.S. interest rates play only a secondary role. Uncertainty about domestic conditions—not interest rate differentials—induced the capital flight in the first place. Interest rates in some countries were more than 2,000 percent a year, and even in dollars, realized returns were huge. (According to table 2 in Kuczynski's paper, annual interest in Argentina in 1990 was more than 1,500 percent, compared with a 67.9 percent devaluation.) That even so, capital continued to flee (or only marginal sums were attracted) reflected high uncertainty about the country's ability to ensure any financial return. Under these conditions, a 5 percentage point change in U.S. interest rates would have made no difference.

Uncertainty was reduced and the expected return on investment in Latin America increased because appropriate domestic financial policies were adopted in most countries.¹ The policies also had positive international consequences, including the Brady Plan and the International Monetary Fund's extended facilities. Because debt reduction accords rewarded good policies, these arrangements enhanced Latin America's access to international capital markets instead of curtailing it, as was initially feared.

Kuczynski mentions that the region benefits from the return of the vast amounts of funds held abroad. I believe that the motivations for this return should be spelled out more clearly. It is not obvious why Latin Americans with funds abroad, if they are rational market participants, would behave differently than other international investors in the presence of profit opportunities. It could be claimed that Latin American residents will be more prone to repatriate funds when profit opportunities arise because their information costs may be lower and there may be complementarities with local sources of finance. It could also be claimed that the contraction in domestic credit required by stabilization policies forced the repatriation of capital to finance existing operations. But there are arguments in the opposite direction. Latin Americans may be more risk averse than other international investors because of their previous bad experience with domestic economic mismanagement and because they are likely to

1. Certain countries, particularly Brazil and Peru, were beneficiaries of capital inflows even though no consistent reforms occurred. This could be interpreted as a systemic phenomenon in which *expectations* of substantive reforms induce a reversal of capital flows, particularly since financial returns are much higher for early investors (provided, of course, that reforms materialize).

have less diversified portfolios. Because the relative weights of the various arguments are unclear, the existence of a large pool of funds abroad does not seem to be a compelling argument in this context.

As for the consequences and prospects for the foreseeable future, there are three questions. How volatile are capital flows? How likely is it that they will continue? And are these flows sowing the seeds of a new debt crisis?

The short answer is that everything depends on the continuation of good policies. The composition of the existing flows is an important indicator of greater stability. Although a large proportion of the capital inflows is invested in short-term bank accounts that can be quickly withdrawn, there is a remarkable increase in more stable financial instruments such as equity and direct investments. The continued development of these more stable instruments is conditional on an increase in confidence which in turn depends on the expectation of continuing good policies.

On the whole, I am optimistic about the prospects for a continued and strengthened flow of capital into the region. In addition to all the benefits that can be derived from the astute management of these inflows, this development provides a clear signal that good policies indeed pay off.

FLOOR DISCUSSION OF THE KUCZYNSKI PAPER

We are at the start of a process here, said Kuczynski in his response to the discussants; it is difficult to say exactly how it will end. He agreed with Mario Blejer (discussant) about the importance of reforms being in place but also felt that changes internationally had helped, allowing the reforms to bloom under bluer skies. Interest rates were important, he stated, because instead of going into fixed-income markets in Latin American countries, investments went into equity markets, where expectations of future performance were higher.

Kuczynski agreed with Albert Fishlow (discussant) that there was a danger of returning to the bad old ways, but he believed that there had been fundamental changes in fiscal policy and debt agreements, so that much of the debt was now at a fixed rate, which was very important. There had also been a change in attitudes, which was why the investments of the 1990s were likely to be more productive than those of the 1970s, when loans were being used to fund the deficits of state enterprises (leading to a high rate of investment but low growth). Finally, while granting the justice of Fishlow's point about the savings rate, Kuczynski said it was important to be careful about where data on savings came from and whether they reflected what was really going on.

In response to Fishlow's concern that capital would leave again if bad policies returned, a participant from Catholic University of Chile said he regarded that possibility as positive. This time, if policies went wrong, there would be some warnings. The same speaker said that one had to be careful in exploring the potential for growth in Latin American countries because the same rate of gross investment in the 1990s was capable of generating more growth than it had in the 1960s and 1970s. In Chile, for example, with an investment rate of 17 or 18 percent, the economy was growing 7 or 8 percent a year—just as, in the first decade of growth in the Republic of Korea, an investment rate of about 20 percent had brought a growth rate of 10 percent. The Chilean participant said it would be wrong to think that to sustain the same rate of growth a country had to return to an investment rate as high as 25 or 30 percent.

One member of the audience expressed the belief that flight capital in Latin America had been a response not so much to the risks of economic uncertainty (which would be healthy) as to the risk of being taxed. Capital movements in response to tax differentials were not always so desirable, said the participant,

This session was chaired by Stanley Fischer, Elizabeth and James Killian Class of 1926 Professor at the Massachusetts Institute of Technology.

© 1993 The International Bank for Reconstruction and Development/THE WORLD BANK

and in this case largely reflected the poor income distribution in Latin America. Middle- and upper-income people had evaded or avoided taxes while the poor carried most of their burden. Kuczynski responded that flight capital was undoubtedly tax avoidance to some extent, as it also was in France or Italy, for example, but that inflation and devaluation—not tax avoidance—had been the main reason for capital flight in Latin America.

A participant from Bolivia said that several Latin American countries, including Bolivia, had been following about the same economic adjustment reforms as Mexico but that Mexico seemed to be attracting more investment. He asked if the panelists thought Mexico's free trade agreement with the United States gave Mexico an advantage over other Latin American countries, at least in the short to medium term. He wondered if the panelists thought policymakers in Latin American countries should negotiate bilateral trade agreements. Kuczynski said he didn't think the North American Free Trade Agreement (NAFTA) was the impetus; the main reason was Mexico's economic reform. The average tariff rate was down to about 11 percent, which was much more than NAFTA would be able to achieve. NAFTA had been a second step, important psychologically, but was not the main reason capital had gone back to Mexico.

A participant from the World Bank's Latin America region asked if people in capital markets felt that there were mechanisms for funding privatized infrastructure or if this required official intervention of some kind. Kuczynski said that pension funds, which had existed for several years in Chile and had been created in Mexico, were possible vehicles for financing infrastructure because they sought long-term investments with stable returns that matched actuarial projections. Kuczynski thought it was essential to develop the bond market in most of these countries. Except for Chile, there really was no bond market, other than for treasury bills, and the market for those depended on both institutional changes and much lower rates of inflation than existed in most countries.

Fishlow said he thought the key question was whether Latin America's new, more market- and outward-oriented approach would be successful enough to generate higher rates of investment. On the basis of the short-term experience in Mexico and Chile, he was concerned that the flow of resources might not be sufficient to finance the enormous investment required in Latin America in the next ten years. In neither Latin America nor the United States was there as serious a commitment to saving as there should be.

The participant from Catholic University of Chile asked that people not assess Latin America as a whole because countries that had begun reforms and sustained them for five to fifteen years were substantially different from countries that had done nothing. Part of what was happening was a contagious effect. Some inflo vs returning to Brazil, for example, probably reflected optimism about what could be done. And Chile the year before recorded a current account surplus—small, but a surplus.

The speaker from the Bank's Latin America region had asked whether there was a policy problem if fiscal accounts were in surplus while there was a current

account deficit. Kuczynski responded that he was not sure there really was a public sector surplus, with one exception. Blejer said that this issue was at the core of what had been discussed. If a balance of trade deficit was being financed by volatile capital inflows, there was a question of how sustainable they were, but since these inflows were being used for direct portfolio investments—a much more sustainable source—he felt there was no reason to worry.

More important, had the fiscal deficit disappeared, and did the region have a fiscal surplus or not? Throughout the region Blejer had observed an extraordinary change in the composition of the fiscal budget, both in sources of revenue (which were more efficient and less distortionary) and in government spending (which now crowded out the private sector less and made foreign and domestic investment much more attractive). This change, which was at the core of policy changes, looks much more sustainable than the overall deficit or result from one year to another.

A participant from the Brookings Institution asked Kuczynski if the hot money had cooled down and if there were reliable data about the direction of change or the kind of investment going to Latin America. Kuczynski responded that hot money moved to where the differentials in perceived prospects were greatest. In 1990 the hot money had been in Mexico; in 1991, in Argentina; and in 1992, in Brazil. But, he noted, the numbers were not reliable.

The same speaker asked if Latin America was substantially different from other capital importers. She asked all three panelists what they saw happening in Mexico. She thought both Fishlow's and Blejer's analyses were correct, adding that in the previous oil boom Mexico had imported mainly capital goods. The point was whether those capital goods were productive or produced results soon enough. The statistics on Mexico were bothersome: Mexico had a \$13 billion current account deficit with a growth rate that was lower in 1991 than in 1990, and according to World Bank research there was no way to say savings hadn't declined. Was it anything other than an act of faith to believe that things were working? Were there any indicators, besides the fact that policies were right—and she agreed that fiscal conditions were dramatically different than at the end of the 1980s—to believe that Mexico would get rid of the increasing deficit and achieve higher growth? Did the current account deficit simply mean Mexico was importing capital—capital that was going to be productive and increase growth?

According to Blejer, capital inflow was the objective indicator that policies were both improved and sustainable. Capital inflows showed that many investors were confident that rates of return were going to be higher. And if capital inflows were really mostly foreign investment and repetition, that was an even stronger indication that there was confidence in Mexico's policies.

Kuczynski gave an example in response to the speaker from Brookings. The Mexican telephone company, he said, had been valued at about \$500 a line four years ago; today it was valued at about \$4,000 a line, which was much higher than the regional Bell telephone companies in the United States. This was possible because of Mexico's growth and reasonable stability and the prospects for

tremendous growth in communications. The announced privatization of Brazil's telephone company had changed its valuation from roughly \$400 or \$500 a line to about \$1,000 a line, which, if open to unrestricted international purchase, would be closer to \$3,000, depending on other variables. This had sparked a tremendous inflow of capital to purchase the stock (which was closed, but people got around that).

What had happened, said Kuczynski, was that Mexico was undergoing serious reform at a time when assets in industrial countries were earning less because of lower interest rates and the recession; when international investors had limited prospects for investments in telecommunications; and when Mexico had a lot of financial assets. The potential for inflows was tremendous. The Mexican banks had been sold for \$13 billion, he said, and there was more to come.

Capital Flows to Developing Countries: Implications from the Economies in Transition?

Susan M. Collins

Are capital flows to developing countries likely to be affected by new demands from Eastern Europe and the former U.S.S.R.? The paper discusses capital flows from private and official sources and concludes that there is little reason for concern about capital flows from the private sector. Few developing countries currently have access to these flows, and the economies in transition are unlikely to attract large-scale private flows during the next few years. In fact, developments in other parts of the world are more likely to influence global capital markets than developments in Eastern Europe.

In terms of official finance, and especially official development assistance, there may be reason for concern, however. Empirical analysis of past experience suggests that large sustained increases in aid to selected countries tended to divert substantial flows from other developing countries. Of course, the current situation is unique: alternative scenarios are possible in which official assistance to Eastern Europe need not divert aid from developing countries.

Recent economic and political developments have raised concerns among developing countries about the availability of international capital flows. As is well known, net resource flows to these countries as a group contracted sharply after 1982. In the late 1980s strong investment demand in many industrial countries, in many cases coupled with low or declining saving rates, prompted widespread concern about a possible global "savings shortage." These concerns were exacerbated as economies in Central and Eastern Europe and in the former U.S.S.R. embarked on market-oriented economic reforms, citing the need to replace or revitalize domestic capital stocks. In this environment it is no surprise that observers worry about the likelihood that the economies in transition could siphon off capital flows that otherwise would have gone to developing economies. Although the economic slowdown in the industrial countries and the associated contraction in investment demand have eased overall concerns about a global "savings shortage," the developing world remains concerned about the possible capital flow implications of the developments in the East.

Susan M. Collins is a senior fellow at The Brookings Institution and associate professor of economics at Georgetown University.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

The implications of the transition are difficult to assess. First, individual economic and political situations are important determinants of private and official capital flows, although the two may respond differently to conditions in these economies and it is important to distinguish between them. These situations, however, remain extremely uncertain. Indeed, each month brings surprise developments.

Second, the official response of the industrial countries is still evolving. The likely magnitudes and forms of financial support from the United States, Japan, and Europe are difficult to predict. For example, there is an ongoing discussion in the United States about how much official assistance should be provided at market terms (through such mechanisms as loans from the International Monetary Fund or nonconcessional export credits) and how much as official development assistance (such as grants or bilateral loans on concessional terms).

At the same time, it is important to put the discussion into perspective. Capital flows to developing countries had run dry well before the revolutions of 1989, and discussions of a perceived global "savings shortage" were quite common. The overall picture for developing countries might have relatively little to do with the economies in transition, even if capital is diverted at the margin.

More generally, such a focus may be misguided for at least three reasons. First, developments in other regions may have much greater effects on international capital markets—particularly on private capital flows. Investment demand in industrial countries may surge as economic activity in the countries of the Organization for Economic Cooperation and Development (OECD) revives in the next few years. Taiwan (China) has announced a major new restructuring effort. And the rebuilding of Kuwait has already created a significant demand for private capital flows.

Second, it may be a mistake to focus on the role of international capital flows to developing countries. Arguments for the view that too much attention has been paid to the availability of external capital point out that internal, not external, saving will account for the lion's share of investment. This view is supported by a wealth of recent and historical country experiences. Krugman (1991) argues that "international capital markets as an engine of growth have been oversold" (p. 3). At the same time, the magnitude of capital inflows in relation to total income or investment in a given economy may understate the role of that capital in development. Foreign direct investment may lead to a transfer of technology and the acquisition of new skills, with important positive spillovers to other sectors. A relatively small loan may provide critical breathing space for a country attempting to avoid a sharp decline in consumption during a period of structural readjustment (Collins 1990). In fact, both arguments contain important lessons. A balanced assessment of the importance of international capital inflows (for the economies in transition, as well as for the developing countries) contains elements of both.

A third reason—which is beyond the scope of this paper—is that capital flows to both groups of countries may remain small and the real issues may turn out to

be on the trade front. For example, will the economies in transition begin producing goods that compete directly with developing country exports in world markets? Will the economies in transition get preferential access to industrial country markets? Poland, Hungary, and Czechoslovakia have already negotiated agreements with the European Community (EC) which may increase their access to some of the EC's protected markets (in, for example, agriculture, steel, and textiles) that are also of great interest to developing countries (see Collins and Rodrik 1991; Messerlin 1992).

This paper examines some relevant features of international capital flows to developing countries, giving special attention to the implications of developments in the economies in transition for the availability of official development assistance (ODA). Section I looks at recent trends in international capital flows to developing countries. Section II discusses the "global savings shortage," focusing on developments in the East in relation to the other parts of the world. Sections III and IV examine concessional flows from official sources and the implications for ODA to developing countries of assistance to the economies in transition. Concluding remarks are in section V.

I. NET RESOURCE FLOWS TO DEVELOPING COUNTRIES: A LOOK AT THE NUMBERS

Table 1 shows the sources of net resource flows to developing countries during the past decade, including net foreign direct investment, net unrequited transfers, net lending from official and private sources, and changes in arrears.¹ The World Bank's Debtor Reporting Service and the OECD also provide these data (although, unfortunately, there can be large discrepancies between the two sets of figures, primarily because the OECD's coverage includes a larger geographic area). Much of the discussion below relies on the OECD data because the OECD is the source of information about official development assistance and thus provides the most comprehensive statistics.²

The table shows that in 1987, after years of decline, net resource flows to developing countries began rising. Although by 1990 they had increased by more than 30 percent (in real terms), the total was still just 73 percent of 1982 levels.

The first three columns of the table disaggregate net resource flows into three broad categories. The first column shows a trend toward an increase in the nominal value of official development finance to developing countries as a group, although the real value of these flows has remained roughly constant. Official financing rose from 38 percent of the total in 1982 to 55 percent in 1990 (including concessional and nonconcessional flows from bilateral and

1. This measure subtracts repayments of principal from gross flows but differs from a frequently used alternative measure—net resource transfers—in that it excludes factor services and transfers.

2. See World Bank (1989) for a discussion of alternative measures of net resource movements to developing countries and the differences between OECD and World Bank data.

Table 1. *Net Resource Flows to Developing Countries, 1982-90*
(billions of U.S. dollars)

Year	Official development finance	Export credits	Private flows	Nominal total	Real total ^a
1982	44.1	13.7	58.2	116.0	175.6
1983	42.4	4.6	47.8	94.8	144.4
1984	47.5	6.2	31.7	85.4	133.7
1985	48.6	4.0	30.5	83.1	128.9
1986	55.8	-0.7	26.7	81.8	100.6
1987	61.5	-2.6	33.7	92.6	98.2
1988	65.5	-2.1	43.8	107.2	105.8
1989	65.5	9.5	48.3	123.3	123.3
1990	78.8	4.6	60.8	144.2	128.9

Note: 1989 prices and exchange rates.

Source: OECD (1991).

multilateral official sources). Concessional aid has accounted for more than three-quarters of official development finance since 1982. (For recent studies of the effectiveness of foreign aid, see Cassen 1987 and Krueger, Michalopoulos, and Ruttan 1989.)

The second column of the table shows that (net) export credits plummeted in the mid-1980s. Even though they recovered during 1988-90, the effect was insignificant because net export credits make up a relatively small share of the total.

Finally, the table shows the familiar "drying up" of private capital flows from half of total net resource flows (\$58 billion) in 1982 to \$27 billion in 1986. Even the \$60 billion recorded in 1990 was only 62 percent of the real value of private flows in 1982. The developing countries' very small presence in international capital markets (OECD 1992) is shown by a look at recent syndicated credits and bond issues. (These instruments account for nearly 80 percent of global borrowing.) Table 2 shows that, of the \$16.3 billion lent in 1991 (excluding credits to the Gulf states), \$12.4 billion went to just four Asian countries. Lending to the economies in transition also fell; only Hungary obtained new credits in 1991.

Table 3 provides information about external bond issues during the same period. Again, the numbers show that few developing countries or economies in transition participated in the market. Following major reform programs, four countries recently regained access to the market: Argentina, Brazil, Mexico, and Venezuela.

Panel A of figure 1 shows the distribution of capital flows across countries in different income groups (based on the United Nations classification of developing countries). The least-developed countries have been relatively insulated from the contraction in flows in the mid-1980s, reflecting their heavy reliance on ODA. Panel B of the figure shows the rise in official flows to these countries from \$12 billion in 1980 (75 percent of their total net resource inflows) to \$14 billion in 1990 (a full 93 percent of the total). Flows to other low-income countries have

Table 2. *Syndicated Lending: OECD, Developing Countries, and Eastern Europe, 1989-91*
(billions of U.S. dollars)

<i>Lending region/country</i>	1989	1990	1991
OECD	99.8	101.0	84.7
Developing countries	16.2	19.8	27.1
China	1.6	1.5	2.2
Gulf states	1.6	1.6	10.8
Indonesia	2.4	3.9	5.1
Korea, Rep. of	0.8	2.0	3.1
Thailand	0.8	1.3	1.6
Eastern Europe	2.4	3.0	0.1
Hungary	0.7	3.0	0.1
Former U.S.S.R.	0.9
Other countries	2.7	0.7	1.3
Total	121.1	124.5	113.2

.. Negligible.

Source: OECD (1992, p. 88).

increased (in real terms) from \$29 billion in 1980 to \$47 billion in 1990,³ \$8 billion of which was additional aid.

As panel B shows, net resource flows to the lower-middle-income countries declined much less from 1980 to 1986 because of the relative constancy of ODA receipts (\$8 billion a year since 1983). The brunt of the collapse in flows was felt by the upper-middle-income group. Net resource flows to these countries plunged more than 70 percent in real terms from 1980 to 1986 but then increased somewhat, to 42 percent of the 1980 level by 1990. Thus the collapse

Table 3. *External Bond Issues, 1989-91*
(billions of U.S. dollars)

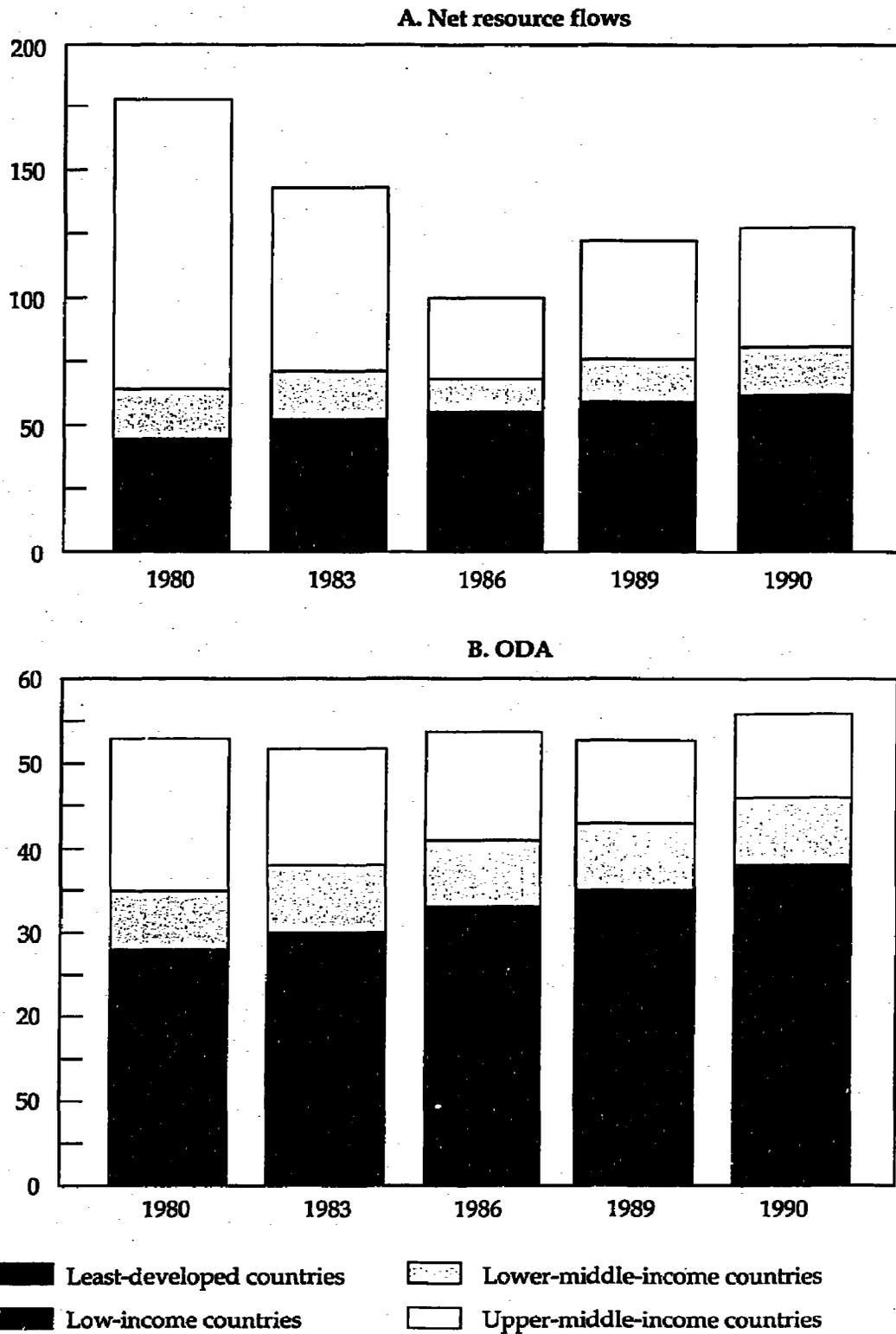
<i>Region/country of origin</i>	1989	1990	1991
OECD	240.4	208.4	271.9
Developing countries	2.6	4.5	8.3
Argentina	0.7
Brazil	1.2
Korea, Rep. of	0.3	1.5	2.4
Mexico	..	0.8	2.1
Venezuela	..	0.2	0.6
Eastern Europe	2.2	1.6	1.5
Czechoslovakia	0.1	0.4	0.3
Hungary	0.9	0.9	1.2
Former U.S.S.R.	0.9	0.3	..
Other countries	10.5	15.4	15.9
Total	255.7	229.9	297.6

.. Negligible.

Source: OECD (1992, p. 47).

3. This group of countries includes Egypt. Part of the increase is thus associated with a rise in ODA to the Middle East from Arab donors.

Figure 1. Allocation of Net Resource Flows and ODA
(billions of U.S. dollars, at constant 1989 prices and exchange rates)



Source: OECD (1991, p. 186).

in international capital flows has been greatest for the (generally wealthier) developing countries that relied on private capital markets.

To summarize, this section makes three main points. First, few developing countries have been able to borrow in international capital markets in recent years, and there is little indication that conditions are about to change. As I point out below, the economies in transition are also unlikely to receive large sums from these sources soon; those economies whose major restructuring efforts succeed can hope to be rewarded with modest flows a few years down the road. Second, however, the sizable amount of private capital that has been invested in Kuwait and the Republic of Korea suggest that capital markets could accommodate quite smoothly large increases in demand from a few countries that are deemed creditworthy. Third, the collapse in international capital flows has been greatest for those (generally wealthier) developing countries that relied on private capital markets. Poorer countries have been cushioned by the relative constancy in the real value of ODA.

II. INTERNATIONAL CAPITAL FLOWS

Suppose developments in the economies in transition do open up attractive investment opportunities. Should developing countries (as a group) be concerned about the capital market implications? The broad answer is yes, but developments in other parts of the world may have much greater ramifications. This section discusses general issues that relate to the question. I begin with a simplistic formulation and then add relevant real-world complexities.

Tradeoffs: A Stylized Framework

The starting point here is the familiar relationship among saving, investment, and net resource transfers (see Collins and Rodrik 1991). We partition the countries of the world into three groups: industrial, developing, and economies in transition. For each of these regions, any excess of domestic investment over domestic saving (an external imbalance) must be financed through a (net) transfer of resources from abroad. External imbalances in one region are offset by imbalances in the rest of the world, so that for the global economy as a whole, total investment is equal to total saving.

Both saving and investment are likely to be functions of real interest rates as well as of other factors. All else being equal, an increase in real interest rates will be associated with increased global saving and a drop in global investment. Suppose for now that there are no restrictions on the international flow of capital, that capital markets work efficiently, and that capital is allocated internationally so as to equate the expected real return on investment. Thus, saving and investment by region and "the" global real interest rate are determined simultaneously. Further, if in autarky the marginal productivity of capital is low in industrial countries in relation to developing countries and economies in transition (where the ratio of capital to labor is considerably lower), we would expect global saving to be allocated disproportionately to the developing coun-

tries and the economies in transition. The resulting capital account deficit in industrial countries would be the counterpart to an industrial country current account surplus and would offset capital account surpluses in developing countries and economies in transition (the counterparts to current account deficits in those regions).

In this simplified world, what would be the effect of an increase in perceived investment opportunities in the economies in transition? All else being equal, the result would be some combination of higher real interest rates, increased global saving, and a reallocation of funds from industrial and developing countries (with reduced investment in these regions) to the economies in transition. The more global saving increases for a given shift of investment demand in one region, the smaller the implied decline in investment in other regions. For example, suppose global saving is very sensitive to changes in real interest rates and investment is not. The end result could be a small rise in interest rates that generates a large increase in total savings and a small reallocation of resources that—before the developments in the transitional economies—would have financed investments in the other two regions. Alternatively, if developments in the economies in transition are associated with other changes that raise global saving (perhaps through reduced defense expenditure that increases government saving), resources could be freed up to finance new investments without a rise in interest rates or a diversion of investment finance from other regions. If, however, saving is relatively insensitive to interest rates and there is no shift in the behavior of global saving, increased investment demand in the economies in transition is likely to lead to a large rise in interest rates and a relatively small net increase in global saving, but a substantial diversion of resources from investments in other regions to the economies in transition.

My own view is that aggregate saving is relatively insensitive to real interest rates. (One way to make the point is to note that levels of saving in industrial countries have been stable as a share of total income during the past two decades, while real interest rates have fluctuated considerably.)⁴ This suggests that the developing countries would have good reason to be concerned about a major increase in investment demand by the economies in transition.⁵ If this were to happen, the developing countries could expect to be hit by a combination of higher interest rates and reduced net resource transfers. Because develop-

4. Skeptics could argue that standard measures of interest rates are poor indicators of the true ex ante return to saving, or that changes over the past two decades in other factors which influence saving have offset the effect of interest rate fluctuations, so that the true interest sensitivity of saving has been underestimated.

5. Overall, industrial countries stand to gain from increased investment opportunities abroad. As net creditors, they gain from increased real interest rates. Furthermore, the diversion of capital from financing domestic projects is, by assumption, the efficient response to investment opportunities abroad that are perceived to be more lucrative. This does not mean, of course, that the gains are distributed equally within the region, and some groups may well end up worse off.

ing countries as a group are net debtors, higher interest rates are welfare-reducing. Further, if developing countries are constrained in international capital markets, even a marginal reduction in their net capital inflows would imply an additional reduction in their welfare. The implications of an increase in investment demand or a decrease in saving in the industrial countries could be traced in a parallel manner. But this is a very stylized scenario. Is the assessment likely to hold up when more realistic considerations are incorporated?

Allocation of Official and Private Capital Flows

In fact, capital is not allocated internationally so that the (expected) real return to investment is the same across regions. As noted earlier, ODA (at concessional terms) accounts for most of the capital flows to developing countries. These flows depend on political, strategic, humanitarian, and other considerations, so it is difficult to predict how official donors will allocate capital and on what terms. Past trends, however, suggest a scenario that might predict the effect on developing countries of changes in the economies in transition.

The stylized market described earlier may not accurately represent the allocation of private capital either—even if real returns are adjusted for risk. Some analysts have concluded that developing countries are often unable to attract capital even after instituting major policy changes that would be expected to enhance the environment for foreign investors. Others argue that differential access to information, “missing markets,” or “bandwagon effects” among private investors effectively constrain the access of developing countries to international capital markets. But if inflows to (some) developing countries are already constrained, an increase in investment opportunities elsewhere may have little or no effect on the volume of capital inflows to these countries—whether or not interest rates rise. Figure 1 suggests that many developing countries are in this boat; the question of whether investment opportunities in the economies in transition will divert investment flows that are currently nonexistent is moot. (Higher interest rates would, however, increase debt service obligations.)

Will developments in the East increase investment demand? For the economies in transition (except for the former German Democratic Republic), capital flows from abroad to finance investments are likely to be quite limited. Because actual investment is apt to fall far short of the amounts that have been suggested by some sources (Collins and Rodrik 1991), domestic saving will remain by far the most important source of financing.

The private sector in the West shows little willingness to invest large amounts in the region in the face of political and economic uncertainties, including the lack of legal protection for property rights. Even successful institutional transitions will take years to develop. And evidence from Latin America suggests that capital inflows may not materialize in the first few years after major reforms—even when those reforms are widely acknowledged to have improved the climate for foreign investors.

Table 4. Estimated Changes in Current Account Balances, 1989-90 to 1994
(billions of U.S. dollars)

<i>Region/country</i>	<i>Scenario 1</i>	<i>Scenario 2</i>
Developing countries	-20	-33
Germany	-50	-60
Japan	+20	+15
Kuwait	-5	-5
Taiwan (China)	-50	-50
United States	+45	+7
Total	-60	-126

Note: A minus sign implies an increase in a deficit or a decrease in a surplus.

Source: Solomon (1991).

Other Elements of the Savings Shortage

There are a number of developments in the global economy that have considerably greater potential for raising global real interest rates than the diversion of funds to markets in the East. Solomon (1991) provides ballpark estimates based on two scenarios for the evolution of the savings-investment balance during the next few years. The main features are reproduced in table 4.

The table shows estimated current account balances in Germany, Japan, and the United States. German unification has already reduced net saving by \$50 billion a year, and this sum could rise to \$60 billion (shown in scenario 2). Japan's current account surplus appears to be on the rise, and the U.S. current account deficit is expected to decline. These developments could generate anything from a modest increase in net saving to a sharp increase in net absorption of external saving.

In the Gulf, war-related destruction has increased demand for investment. Solomon (1991) estimates that investments in Kuwait could lead to a reduction of about \$5 billion in that country's current account balance. Taiwan (China) recently announced a six-year, \$300 billion infrastructure improvement program. Given Taiwan's large stock of foreign exchange reserves, it is unlikely to have trouble financing these outlays. Finally, other developing economies are expected to expand their current account deficits by perhaps \$20 billion to \$33 billion, increasing net investment demand by \$60 billion to \$126 billion a year. These potential imbalances swamp the current account imbalances that the economies in transition could conceivably finance during the next few years.

III. OFFICIAL DEVELOPMENT ASSISTANCE

Table 5 shows that ODA in low-income countries (excluding China and India) averaged \$17 per capita, or 5.6 percent of gross national product (GNP), in 1989. The average reflects a wide range: in 1989 it was more than 20 percent of GNP for some countries, including Lesotho, Malawi, Mozambique, Somalia, and Tanzania. (For a discussion of recent trends, see OECD 1991.) Thus the availability—and the allocation—of development assistance funds are of considerable interest to low-income developing countries.

Table 5. *Indicators of Official Development Assistance by Income Group, 1989*
(U.S. dollars)

Country	Total ODA (millions)	Per capita ODA	ODA as a percentage of GNP
Low-income	21,467	7.3	2.2
China and India	4,101	2.1	0.6
Other	17,366	17.4	5.6
Lower-middle-income	10,973	17.4	1.3
Upper-middle-income	655	2.0	0.1

Note: Table uses World Bank classifications of income groups. ODA amounts are from all sources but exclude "unallocated" amounts. ODA per capita and as a percentage of GNP are weighted averages.

Source: World Bank (1991, table 20).

Until the early 1970s all ODA was provided by eighteen countries, all members of the Development Assistance Committee (DAC) of the OECD (OECD 1991). Other members of the OECD now contribute a small but growing share. Non-Arab developing countries and the former U.S.S.R., together with the economies in transition, provided about 13 percent of the total in the mid- to late 1970s, but these countries accounted for less than 5 percent of the total in 1990. The Arab countries are the largest new group of donors, accounting for nearly a quarter of the total in the mid-1970s. The amount contributed, however, has varied, falling from \$12.7 billion in 1981 (in real 1989 dollars) to just \$1.3 billion in 1989 and soaring to \$6.2 billion in 1990 (after the Iraqi invasion of Kuwait). This sharp increase more than offset the declines from Central and Eastern Europe and from the former U.S.S.R.

ODA from Arab countries is concentrated among a small number of recipients. In 1990 five countries (Egypt, Jordan, Turkey, Morocco, and Syria) received 70 percent of the total. Another 25 percent went to countries in the Middle East or to other Arab countries. African countries south of the Sahara, in contrast, received just 1.7 percent of the total (compared with 32.8 percent of ODA from DAC members); disbursements to developing countries in the Western Hemisphere are also consistently small.

Because DAC members continue to provide the lion's share of concessional assistance to most developing countries—and certainly to the least-developed countries—either directly or through contributions to multilateral organizations, the discussion here focuses on these funds. Note that total ODA as a share of the aggregate gross domestic product (GDP) of the member countries has remained relatively constant during the past two decades.⁶ Underlying the relatively constant share spent by the industrial countries as a whole, however, is a

6. A simple regression shows that of every \$1,000 increase in the total income of DAC members, an average \$3.55 was spent on ODA (*t*-statistics in parentheses).

$$\text{ODA} = -357.990 + 0.00355 * \text{GDP}$$

$$(-0.887) (68.432)$$

$$R^2 = 0.99; \text{Durbin-Watson} = 1.81$$

substantial shift in allocation. The U.S. share has fallen from 34 percent in 1970–71 to less than 20 percent during 1989–90 (OECD 1991, pp. 172–73, 154, 185). The Japanese share rose from 11 to 19 percent over the same period. Some of the smaller OECD members, such as Sweden, Norway, and Finland, have increased their ODA contributions substantially. (In 1989–90 these three countries contributed 6.2 percent of total DAC ODA, or nearly 0.9 percent of their combined income, compared with just 2.7 percent of the DAC total in 1970–71.) These figures point to differences in the way donor countries allocate aid—differences that reflect the domestic arrangements under which assistance is appropriated, historical relationships with groups of developing countries, and attitudes toward assistance. A country-by-country treatment, however, is beyond the scope of this paper.

For a number of reasons, actual aid deviates as much as \$1 billion in some years. First, some assistance is provided in large, discrete amounts, and the timing of these disbursements can lead to significant fluctuations in the levels of aid. Second, domestic developments, such as special budget negotiations, can delay—or alter—assistance flows. And third, flows can be influenced by disasters in a recipient country that lead to a temporary increase in humanitarian assistance.

What Determines Assistance Flows?

Any donor may decide to give more (or less) assistance to a particular country for a number of reasons. First, the donor may increase funding either because the share of income used for assistance rises (by accident or by intent) or because of an increase in domestic income. Second, as noted above, the donor may wish to respond to a disaster or to a change in a country's perceived strategic importance. (Of course, if there is no change in the donor's total outflow, this implies an offsetting change in assistance to other recipients.) And third, a donor may reallocate its support because of developments in other (potential) recipient economies.

In the most general sense, assistance to any one country depends on the economic, political, and other characteristics of each potential donor and of all potential recipients. Although a comprehensive empirical analysis that attempts to take all these factors into consideration is quite complex, it is possible to determine an average relationship between the total per capita assistance received by an individual country and a few of the country's key characteristics.

The analysis relies on OECD data that indicate annual receipts from DAC and other sources for a cross-section of developing countries. The per capita measure provides a convenient indicator for cross-country comparisons. The analysis uses average values of aid and of the explanatory variables from 1987–89 so as to smooth annual variations. The first explanatory variable is per capita income.⁷ The (log of) population is also included as an explanatory variable to

7. Based on World Bank estimates. Territories or countries that are not World Bank members or that have populations of less than 1 million are excluded.

Table 6. *Per Capita ODA for a Cross-Section of Countries: Regression Results*

Variable	All countries			Middle-income	POP greater than 50 million
	Without POP	With POP	Low-income		
Constant	96.35 (4.30)	135.21 (7.68)	57.77 (1.40)	223.8 (5.53)	59.87 (1.74)
Ln(PCI)	-10.00 (-2.95)	-10.86 (-4.28)	2.00 (0.27)	-22.31 (-4.03)	-3.79 (-0.95)
Ln(POP)		-12.67 (-7.09)	-11.42 (-5.78)	-14.63 (-5.00)	-5.59 (-1.98)
Adjusted R ²	0.11	0.50	0.55	0.53	0.33
Number of observations	64	64	27	37	11

Note: The dependent variable is the average of per capita ODA, 1987-89. Figures in parentheses are *t*-statistics. The estimation method is ordinary least squares. PCI, per capita income; POP, population.
Source: World Bank (various issues).

account for populous countries, such as India, that may receive less assistance per person than smaller countries with similar income levels.⁸

The regression results in table 6 confirm that per capita aid is negatively related to per capita income. This variable alone accounts for about 10 percent of the variation in aid levels across developing countries. As shown in column 2, per capita income and population together explain half the average variation in per capita aid across countries. Although that is a respectable proportion for a cross-section analysis, it is clear that a substantial amount of the difference in aid receipts across countries depends on other factors, some of which have been mentioned above.

The first column shows that the estimated coefficients on per capita income and population are each significantly different from zero and that the magnitudes are quite sensible. For example, they suggest that a 1 percent increase in a country's per capita income, given its population, tends to reduce aid by about \$0.11 per capita. In other words, a 50 percent increase in per capita income (for example, from \$500 to \$750) would tend to reduce ODA by about \$5.50 per capita. Similarly, a 1 percent increase in population reduces per capita assistance by about \$0.13.

The next two columns attempt to show whether the determinants of aid differ across income groups (using World Bank income classifications). The third column shows that there is no statistically significant relationship between aid per capita and income level among low-income countries. Nevertheless, it remains true that more populous countries tend to receive less aid per person than less populous ones. The fourth column shows that aid receipts are negatively related to both income level and population for middle-income countries.

8. An additional explanatory variable was also considered: each country's real growth rate. The hypothesis under examination here was that, all else being equal, a country would tend to receive additional assistance when real economic activity was contracting. This variable, however, was consistently insignificant and the estimated coefficient was small.

These estimates suggest that larger populations have about the same negative effect on aid receipts for low- and middle-income countries. All these results predict negative aid receipts for especially large countries. Clearly the average relationship between population and aid is nonlinear. The final column looks at the determinants of aid per capita for the few countries in the sample with more than 50 million people. Unfortunately, population and income levels provide a much less accurate predictor of aid per capita for this group.

IV. SOME LESSONS FROM EXPERIENCE

Is the extension of official assistance to Eastern and Central Europe and the successor states of the former U.S.S.R. likely to affect the availability of ODA to developing countries?⁹ Because of the difficulties of predicting how much assistance might be involved and what—if any—effect such funds would have on the amount of ODA from industrial countries, any response must be interpreted with considerable caution.

In the scenario developed below, I follow a two-step analysis. First, I examine the evidence showing how increases in assistance to individual countries has affected ODA disbursements to other recipients. The point is that on previous occasions developing countries have received large sustained increases in ODA. These instances provide some indication of how donors might react to a new group of potential aid recipients (although there is no relevant precedent to suggest how the West might respond financially to the dissolution of the Soviet empire).

Second, I formulate a benchmark establishing a level for concessional assistance to the economies in transition in the next few years. To determine the benchmark, I ask how much aid these economies would receive if they were treated “like” countries with similar (in some respects) characteristics, based on the empirical analysis of cross-country aid allocations described above. Although the resulting figures are neither predictions of actual flows nor estimates of “appropriate” assistance, they provide a useful point of comparison.

If the economies in transition began to receive annual flows of concessional assistance, the developing countries could stand to lose considerably unless donor countries increased their total contributions. The question, then, is how much of any aid to the East would come from an increase in the total, and how much would come from a diversion of aid?

This section presents one possible scenario based on past experiences in which countries have benefited from large sustained increases in aid from DAC members. I do not look at short-term surges in concessional assistance, under the assumption that it will take many years (perhaps decades) for the economies in transition to raise their per capita income levels above developing country levels (Barro and Sala-i-Martin 1991; Dornbusch and Wolff 1992). That leaves only

9. Note that current OECD definitions exclude from ODA concessional flows to Eastern Europe.

Table 7. ODA for Four Countries with Large Sustained Increases, Selected Periods, 1970-90
(millions of U.S. dollars)

Year	Bangladesh	China	Egypt	Israel	Total ODA	Percentage of DAC total
1970-71	57.0 ^a	0	171.5	258.1	457.1	1.6
1976-80	1,481.7	59.1 ^b	2,357.1	1,516.0	5,378.4	16.0
1981-85	1,698.1	1,010.9	2,412.8	1,902.9	7,024.7	16.4
1986-90	1,731.0	1,771.6	1,978.8	1,470.7	6,948.1	15.0
1989						
PCI (dollars)	180	350	640	9,790	n.a.	n.a.
POP (millions)	111	1,134	51	5	n.a.	n.a.

n.a. Not applicable.

Note: Nominal ODA flows from DAC countries include unallocated amounts and contributions through multilateral organizations. The figures have been deflated using the GNP deflator for DAC members, 1989 = 100.

a. 1971.

b. 1979-80.

Source: OECD data and World Bank (various issues).

four cases of a sustained increase in (real) receipts of at least \$1 billion—Bangladesh, China, Egypt, and Israel (see table 7).

Each of these episodes reflects a special set of circumstances. Bangladesh, a very poor country, was created in 1971 out of the former East Pakistan following a long campaign to gain independence that culminated in a war between Pakistan and India. Bangladesh received small amounts of aid in 1971. These flows surged during the next few years and have since remained high.¹⁰ China did not begin receiving official assistance until it had established diplomatic relations with the United States in 1979. Official aid to Egypt and Israel¹¹ surged during 1975-79, a period that included the negotiation of two Sinai disengagement agreements, the 1977 visit to Israel by Egypt's President Sadat, trilateral negotiations at Camp David in 1978, and the Egypt-Israeli Peace Treaty in 1979.

Table 7 shows that the real value of assistance to these four countries was less than \$0.5 billion during 1970-71 but had risen to \$5.4 billion by 1976-80, more than 15 percent of average development aid during the period. This increase is quite a bit smaller than the amount "predicted" for the economies in transition, but it represented a substantial share of total aid to the developing countries at the time.

The next step is to empirically examine the implications for other recipients of increased aid to these four countries (C4). The variable we are trying to explain is *ODAT4*—the real value of aid from DAC members and multilateral organizations to all developing countries except Bangladesh, China, Egypt, and Israel.

10. Aid to Pakistan declined after Bangladesh became independent, but the combined assistance to the two countries soon far exceeded the real value of the amounts received by Pakistan prior to 1972.

11. Israel receives almost no concessional assistance from multilateral organizations.

Table 8. *ODA Diversion? Regression Results*

<i>Item</i>	<i>OLS level</i>	<i>ARI difference</i>
Constant	-1.8464 (-0.64)	0.5312 (1.25)
Real GDP	0.0035 (9.64)	0.0017 (1.37)
ODA4	-0.6672 (-2.46)	-0.9527 (-3.11)
ρ		-0.5913 (-3.11)
Adjusted R^2	0.91	0.37
Durbin-Watson	1.92	
Number of observations	21	20

Note: Data are for 1970-90. Figures in parentheses are *t*-statistics. The dependent variable is ODA for all countries except Bangladesh, China, Egypt, and Israel.

Source: OECD data and World Bank (various issues).

Note that this measure is based on the OECD concept. The totals include assistance allocated to individual recipients, as well as unallocated amounts. The OECD classifies as "unallocated" funds that donor countries spend on projects which benefit groups of countries, such as general education programs (see OECD 1990). Two explanatory variables are included:¹² the gross national product of DAC members, and ODA4, the real value of concessional assistance to the C4 countries.¹³ If the entire increase came from additional financing with no diversion of ODA from other developing countries, the coefficient on ODA4 should equal zero. If the additional official development assistance to the C4 came, dollar for dollar, from a reduction in aid to other developing countries, the coefficient on ODA4 should equal minus one.

The results are presented in table 8. The estimates show the expected strong positive correlation between income and aid. They also point to a significant negative relationship between aid to the two groups of countries. The results suggest that of every additional dollar to the C4, \$0.67 was diverted from other potential recipients and \$0.33 came from additional financing. The table also looks at the relationship between the first differences of variables.¹⁴ Again, we find evidence of aid diversion. These results suggest that a full \$0.95 of each additional dollar to the C4 was diverted from other countries.

12. The GNP deflator for DAC countries was initially included but was consistently insignificant.

13. Both explanatory variables are treated as exogenous even though ODA4 is likely to be correlated with the error term. (We were unable to find appropriate instruments.) Note that this endogeneity is likely to bias the coefficient estimate on ODA4 upward. Shocks that increase ODA to all countries (for example, developments in the donors) will increase both ODA4 and ODA4, and this will be reflected in the coefficient estimate. Thus, the estimates presented in tables 9 and 10 are likely to understate any diversion of ODA from the developing countries to the four countries that have been singled out.

14. The estimation procedure corrected for first-order autocorrelation of the errors. There appears to be a significant negative relationship between consecutive errors. Unusually high levels of ODA in one year tend to be partially offset by low levels in subsequent years, all else being equal.

Table 9. ODA Diversion from Low-Income Countries? Regression Results

Item	Allocated: total ODA		Allocated: low-income ODA	
	AR1 level	OLS difference	AR1 level	OLS difference
Constant	1.8556 (0.55)	-0.1176 (-0.26)	-0.8252 (-0.40)	-0.0576 (-0.14)
Real GDP	0.0022 (5.39)	-0.0018 (1.59)	0.0015 (5.93)	0.0014 (1.34)
ODA4	-0.4275 (-1.50)	-0.2654 (-0.71)	-0.2798 (-1.47)	-0.2803 (-0.81)
ρ	0.4247 (1.78)		0.1338 (0.51)	
Adjusted R ²	0.87	0.03	0.82	0.00
Durbin-Watson		1.78		2.12
Number of observations	19	19	19	19

Note: Figures in parentheses are *t*-statistics.

Source: OECD data and World Bank (various issues); see text for further discussion.

The diversion result is more worrisome the larger the amount diverted from low-income countries. I turn next to explaining *ODAL4*—ODA to all other low-income countries (excluding Bangladesh and China). The measure available for this group, however, is not strictly comparable to the measure for total ODA used above and is less satisfactory for two reasons. First, the measure comes from aggregating ODA receipts from all countries that the World Bank classifies as low-income and therefore excludes all aid that is listed as unallocated. And second, it includes receipts from all sources, including Arab countries, which is likely to bias the results in the direction of no diversion. There is little evidence that the Arab countries would have provided more aid to other parts of the world if they had provided less to Arab countries and the Middle East. Also, Arab aid to Egypt is highly positively correlated with Arab aid to countries in the Middle East.

For a comparison, we also look at total official development assistance (*ODAWT4*) from all sources, based on World Bank definitions. In both cases the regressions are run in levels and in first differences. To be consistent, we use ODA receipts by the *C4* from all sources as one of the explanatory variables.

The results as reported in table 9 point to three conclusions. First, there appears to have been a smaller diversion (\$0.43 per dollar) from all recipients using the new measure, but this coefficient estimate is not significantly different from zero. Second, the results suggest that \$0.28 of an additional dollar to the *C4* came from low-income developing countries. This ($0.28/0.43 = 0.65$) is approximately proportionate to the low-income countries' share of total ODA from all sources, but the coefficient is also not significant. Third, the explanatory variables explain little of the variation in changes in allocated assistance.

These results, together with those in table 8, suggest that funds that were diverted to the *C4* came from amounts that would otherwise have been unallocated—in other words, funds that would have benefited groups of developing countries. There is only weak support for the view that assistance was

diverted from individual low- or middle-income countries. If we take the two sets of results at face value, they suggest that, all else being equal, an extra dollar to the C4 was generated as follows: \$0.33 in new financing, \$0.24 from unallocated assistance, \$0.15 from assistance to middle-income countries, and \$0.28 from assistance to low-income countries.

A case can be made that it would be in the interests of developing countries for any concessional assistance provided to the economies in transition to remain separate from flows to developing countries. The lesson from experience is that significant funds are diverted when new countries become eligible for aid and that there is relatively little expansion of the overall pie. A separate pie for the economies in transition may guard the developing countries against a repeat performance.

A Benchmark for Aid to the Economies in Transition

Suppose the economies in transition received the same amount of official concessional assistance as developing countries with similar populations and per capita incomes. In this case, a rough estimate of the amounts they would receive can be calculated using the coefficient estimates in table 6. Note that the resulting figures simply provide one interesting benchmark. They do not estimate the level of assistance that would "best" assist these countries or the level that is likely to be forthcoming from industrial country governments.

These calculations are sensitive to a number of factors. The first is which measure of per capita income is used. For consistency I use recent estimates provided by the World Bank. A second issue is whether to treat successor states separately. I have disaggregated the former Soviet republics but not countries in Europe.¹⁵ Third, I use the estimates in the first column of table 6 (based on the entire sample of countries) except for Russia and the Ukraine. Because these states have populations in excess of 50 million, their aid totals can be estimated using the last column of table 6.

The results of this exercise are shown in table 10. The final column shows that if the economies in transition were treated "like" developing countries, they would receive about \$8.2 billion in annual concessional assistance. Of this total, about \$1.5 billion would go to countries in Eastern and Central Europe and about \$6.7 billion would go to the former Soviet republics.

Compared with recent total contributions of aid from the DAC countries, \$8.2 billion is a substantial sum—about 17 percent of the 1989–90 total. It is also large compared with the likely annual increase of about \$1.6 billion in real terms (assuming that GNP grows 3 percent a year and that 0.35 percent of this increase is allocated to aid). Nevertheless, recent developments in the former Communist countries are historic events with potential benefits that are extremely difficult to evaluate financially. And compared with potential reductions in defense spend-

15. The interested reader can calculate predicted ODA under alternative assumptions using the data provided. For example, it may be instructive to split Yugoslavia into separate republics.

Table 10. Annual ODA to Economies in Transition: One Benchmark

Country or republic	Population (millions)	GNP per capita (U.S. dollars)	"Predicted" aid per capita (U.S. dollars)
<i>Eastern and Central Europe</i>			
Bulgaria	8.9	2,210	23.88
Czechoslovakia	15.7	3,140	12.88
Hungary	10.6	2,780	19.18
Poland	37.8	1,700	8.41
Romania	23.3	1,640	14.93
Yugoslavia	23.8	3,060	7.89
Total	120.1	2,279	12.25
<i>Former Soviet republics</i>			
Armenia	3.3	1,412	41.32
Azerbaijan	7.1	1,240	33.02
Belarus	10.3	2,111	22.53
Estonia	1.6	1,942	47.03
Georgia	5.4	1,534	34.18
Kazakhstan	18.7	1,191	21.19
Kyrgyzstan	4.4	941	42.07
Latvia	2.7	2,110	39.50
Lithuania	3.7	1,959	36.31
Moldova	4.4	1,412	37.67
Russia	148.0	2,138	17.82
Tajikistan	5.2	797	41.77
Turkmenistan	3.6	1,151	42.44
Ukraine	52.0	1,613	25.71
Uzbekistan	20.3	842	23.92
Total	290.9	1,780	19.25

Note: Each republic's GNP is estimated as the total GNP of the former U.S.S.R. (estimated as \$1,780 per capita times population) times that republic's share of net national product in 1988. Predicted aid assumes that the economies in transition are treated like developing countries by ODA donors. Predicted aid per capita = $135.21 - 10.86 \ln(\text{PCI}) - 12.67 \ln(\text{POP})$ for all countries and all republics except Russia and the Ukraine. For these two republics (with populations greater than 50 million), predicted aid per capita = $59.87 - 3.79 \ln(\text{PCI}) - 5.59 \ln(\text{POP})$. (See table 6.)

Source: For GNP of the former U.S.S.R., recent World Bank estimates. For other data, IMF and others (1990), World Bank data, and author's calculations.

ing associated with the dissolution of the U.S.S.R., \$8.2 billion is not a large amount.

V. CONCLUSIONS

On balance, the developing countries appear to have more reason to be concerned that developments in the East will reduce flows of ODA than that these developments will divert private capital or push up the real cost of borrowing in international capital markets. Recent data show that few developing countries currently borrow internationally. The constraint on their access appears to be perceptions in the private sector of their creditworthiness. It is unlikely that Eastern Europe will have much impact on that access over the next few years—or that Eastern Europe itself will receive significant private capital flows. The potential for increased current account imbalances in other parts of the world is

considerably more likely to affect international capital markets and the levels of real interest rate.

Official capital flows are difficult to predict, and bilateral and multilateral ODA, in particular, may be the wild card. Evidence to date suggests that total official financing will be measured and that much of the package will be provided as loans or lines of credit on market terms. But many of the calls for assistance argue for concessional financing for the economies in transition—at a time when public support for foreign aid (especially in the United States) appears to be quite low.

Will ODA be diverted from one group of countries to another? A possible scenario comes from looking at the implications of large sustained increases in ODA to individual countries that have occurred in the past. This paper suggests that a substantially greater fraction of these increases appears to have been associated with declines in aid to other recipients than with an increase in the ODA pie. To the extent that official flows are offset by reduced military expenditure in the West, for example, there is little reason to fear a diversion of ODA away from the developing countries. It is difficult, however, to assess how large this offset might be.

REFERENCES

- Barro, Robert, and Xavier Sala-i-Martin. 1991. "Convergence across States and Regions." *Brookings Papers on Economic Activity* 1:107–58.
- Cassen, Robert, and associates. 1986. *Does Aid Work?* Oxford, U.K.: Clarendon Press.
- Collins, Susan. 1990. "Lessons from South Korea." *American Economic Review* (May):104–07.
- Collins, Susan, and Dani Rodrik. 1991. *Eastern Europe and the Soviet Union in the World Economy*. Policy Analyses in International Economics 32. Washington, D.C.: Institute for International Economics.
- Dornbusch, Rudiger, and Helgar Wolff. 1992. "Economic Transition in Eastern Germany." *Brookings Papers on Economic Activity* 1: 235–72.
- Flemming, John. Forthcoming. "The Effects of Competition for Funds between Eastern and Southern Europe." In Albert Giovannini, ed., *Finance and Development: Issues and Experiences*. Cambridge, U.K.: Cambridge University Press.
- IMF (International Monetary Fund). 1991. *Private Market Financing for Developing Countries*. Washington, D.C.
- IMF, World Bank, Organization for Economic Cooperation and Development, and European Bank for Reconstruction and Development. 1990. *The Economy of the USSR: A Study Undertaken in Response to a Request by the Houston Summit*. Washington, D.C.
- Krueger, Anne O., Constantine Michalopoulos, and Vernon Ruttan. 1989. *Aid and Development*. Baltimore, Md.: Johns Hopkins University Press.
- Krugman, Paul. Forthcoming. "International Financial Integration and Economic Development." In Albert Giovannini, ed., *Finance and Development: Issues and Experiences*. Cambridge, U.K.: Cambridge University Press.

- Messerlin, Paul. 1992. "Restrictions on OECD Imports from Eastern Europe: An Overview." Paper presented at the European Bank for Reconstruction and Development conference on "Eastern Europe Trade Policy Issues."
- OECD (Organization for Economic Cooperation and Development). 1990. *Geographical Distribution of Financial Flows to Developing Countries, 1985-88*. Paris.
- . 1991. *Development Co-operation*. Paris.
- . 1992. *Financial Market Trends*. Paris.
- Solomon, Robert. 1991. *Do We Face A Global Shortage of Capital?* Brookings Institution Discussion Paper 91. Washington, D.C.
- World Bank. 1989. *World Debt Tables, 1989-90*. Washington, D.C.
- . Various issues. 1991. *World Development Report*. New York: Oxford University Press.

COMMENT ON "CAPITAL FLOWS TO DEVELOPING COUNTRIES:
IMPLICATIONS FROM THE ECONOMIES IN TRANSITION?" BY COLLINS

András Inotai

My comments on Susan Collins's paper begin with the main issues raised by the author—a savings shortage, foreign direct investment, and official development assistance. I then turn to issues not discussed in the paper—the effect of economies in transition on developing countries and the diversion of financial flows *within* the economies in transition—and make some concluding remarks.

I. THE SAVINGS SHORTAGE

I will deal with three concepts developed by the author: (a) in the global economy total investment is equal to total saving; (b) the industrial countries' current account surplus is financing the capital needs of developing countries; and (c) international economic developments could increase net investment demand by \$60 billion to \$126 billion a year.

- Although it is true that only saved money can be invested, not all savings will be invested. First, there is a time lag in both directions. In some years global savings substantially exceed global investments; in others, previous savings finance higher yearly investments than actual yearly savings. Second, discrepancies are particularly manifest in economies in transition; because of prevailing instabilities and legal and institutional deficiencies in these economies, higher savings will not be converted immediately into higher investment.
- While industrial countries' surpluses are used to finance capital needs in developing countries, in the 1980s many of these countries experienced a net resource outflow, so that this part of the world provided financing for some industrial countries. Currently, some industrial countries have large current account deficits that have to be financed by the current account surpluses of other (mainly) industrial countries. Simultaneously, some developing economies—Taiwan (China) and Hong Kong, for example—have emerged as large potential capital exporters. Therefore, any generalization is problematic.

András Inotai is general director of the Institute for World Economy, Budapest.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

- At least three additional factors may produce a credit crunch. First, according to the Institute of International Economics, financing economic development in the economies in transition to reach U.S. or Western European per capita levels in ten years would require \$1,500 billion a year; this is clearly an illusion. The institute forecasts annual external financing of \$30 billion to \$90 billion, a mere 2 to 6 percent of this sum (Collins and Rodrik 1991). The IFO, a Munich-based economic forecasting institute, puts the annual capital needs of Eastern Europe at between \$75 billion and \$98 billion, and the requirement of the former U.S.S.R. at between \$184 billion and \$235 billion (most of this should be financed by domestic savings). Financial support from the Organization for Economic Cooperation and Development (OECD) has been calculated at \$50 billion a year, about 5 percent of the investments of the European Community (EC) (Ochel 1991).

Second, the EC has announced a one-third budgetary increase between 1993 and 1997 to finance the structural fund and the proposed cohesion funds (*Handelsblatt*, April 9, 1992). This reflects the appeals by the poorer member countries of the Community for increased financial transfers following the decision to provide better market access to East Central Europe (Czechoslovakia, Hungary, and Poland).

Third, economic dynamism may spread from the Asian newly industrialized economies (NIES) to selected countries of Latin America and Eastern Europe. Recent economic data from Latin America indicate high growth rates, and a recovery is predicted in Czechoslovakia, Hungary, and perhaps Poland in the next few years. This will generate increased demand for external financing, particularly if remaining trade barriers do not permit an increase in domestic capital accumulation. The need for capital implies higher levels of saving—not only in the developing or transforming economies but in the industrial countries as well. The substantial changes that increasing savings would require in the budgetary policies of major OECD countries, however, make this a potentially explosive issue.

Higher saving would require higher interest rates, adding to the financial burden of indebted countries. It remains to be seen whether the higher stock of available capital will be used only to finance higher debt burdens or if a part of this capital will be available, even with a higher debt burden, for investment.¹

Better economic policies and successful introduction of market mechanisms are likely to improve resource utilization so that a smaller amount of capital (and external financing) can produce the desired growth. In this case, global

1. If, for instance, the economies in transition, excluding the successor states of the U.S.S.R., get the promised \$39 billion from international financial organizations and Western governments until 1995, their overall indebtedness would rise to \$130 billion. At present interest rates, annual interest payments would be \$1.2 billion for Bulgaria, \$1.1 billion for Czechoslovakia, \$1.3 billion for Romania, \$2.5 billion for Hungary, and \$4 billion for Poland (taking into account a 50 percent debt forgiveness). Higher interest rates would, of course, raise the repayment burden. (For further details, see Schroeder 1991, p. 333.)

demand for capital may become somewhat smaller. It is unlikely, however, that it could compensate for the powerful factors indicating that less capital will be available in this decade.

II. COMPETING FOR FOREIGN DIRECT INVESTMENT

In the 1980s foreign direct investment was increasingly concentrated in the industrial countries. In 1989–90 annual foreign direct investment in the Asian and Pacific region surpassed \$15 billion (compared with the 1985–90 average of \$11 billion), and in Latin America it reached \$10 billion (compared with a 1985–90 average of \$8 billion).²

The economies in transition appeared on the investment market in 1990–91. Inflows to Czechoslovakia, Hungary, and Poland totaled \$3 billion in 1991 (about 1.5 percent of global foreign direct assistance flows in that year). Although this is still a small fraction of foreign direct assistance to the Far East or Latin America, forecasts indicate an increase to the three East Central European countries of \$4 billion to \$8 billion in the next few years.

Although Susan Collins states that the West has little inclination to invest in the economies in transition, recent statistical data seem to contradict this statement—or at least to suggest a less generalized view. A distinction should be made between the East Central European and the other East European economies (Bulgaria, Romania, and Yugoslavia and its successor states) regarding both the activity of foreign direct assistance and the economic, political, legal, financial, and other conditions that attract foreign investors. The recently signed association treaties with the EC provide an additional argument for more investment in East Central Europe.

Certainly international competition for direct foreign investment will become keener; it may or may not divert potential investors from developing countries to economies in transition. In most cases, the two groups are not alternative competing potential host areas. Most foreign investors in Czechoslovakia, Hungary, and Poland have a clear geographic area within which they are looking for investment possibilities. Second, neither the developing countries nor the economies in transition receive very much foreign capital because foreign investors' interest has focused on the industrial countries. Third, the diversion of foreign direct assistance is not a one-way street. It is probable that both groups may divert investment capital from OECD countries in the future. Czechoslovakia and Hungary, for instance, are competing with other potential investment locations in Europe rather than in the developing countries. And fourth, regions or groups of countries do not compete with each other as much as with discrete national economies or regions within the same economy.

2. According to *Economic Reform Today* (Fall 1991, p. 13), Latin America received external private capital flows of about \$40 billion. About half of this capital consisted of returned flight capital (see *Financial Times*, April 6, 1992).

III. WILL THE ECONOMIES IN TRANSITION JOIN THE OFFICIAL DEVELOPMENT ASSISTANCE CLUB?

Calculating the potential diversion of official development assistance to the economies in transition is a hypothetical exercise because none of the economies in transition has been included thus far in the official development assistance club. On the basis of gross domestic product (and other statistical data), it is unlikely that they will be joining. In addition, the OECD countries do not want to see the economies in transition in an official development assistance framework.

But even if the impossible were to happen and *some* of these countries (mainly some successor states of the former U.S.S.R.) became members, official development assistance flows would not be substantially diverted toward the new members. What is more important is that the economies in transition, particularly those of the former U.S.S.R., are no longer important donors to (selected) developing countries. Collins's scenario calculates a diversion of \$8.2 billion, but she does not consider that almost the same amount (\$6.7 billion) was provided to the developing countries by the economies in transition in the late 1980s. An additional point is that the peace dividend likely to result from the political changes in the former U.S.S.R. should compensate for any hypothetical loss caused by a potential resource diversion. Therefore, real fears are not rooted in a resource redistribution within the official development assistance framework. Nevertheless, there are two possible implications for official assistance flows.

First, OECD governments may give relatively less attention to official development assistance because the threat of communism has been eliminated. Second, advanced development, geographic proximity, and better prospects for successful transformation and rapid growth in the economies in transition may turn the attention of the industrial countries increasingly toward Central and Eastern Europe. Most of the commitments proposed thus far are tied to unfavorable conditions, however, and therefore have not been disbursed.³ Another feature that may influence financial commitments to the economies in transition is the financing of environmental cleanup and nuclear safety, an issue whose urgency is fundamentally based on the region's geographic proximity to Western Europe.

IV. DEVELOPING COUNTRIES AND ECONOMIES IN TRANSITION: PARTNERS AND COMPETITORS

Growing political cooperation and the rapid dismantling of military capacities will increase the peace dividend and benefit the economies in transition as well as the developing countries. The opening-up of the former socialist economies and their prospective recovery are expected to create rapid economic growth and an expanding demand for imported goods. Taking into account their economic

3. The Group of Twenty-four promised a total of \$2.8 billion to Bulgaria, Czechoslovakia, Hungary, and Romania for balance of payments support in 1991, but in fact only \$1.8 billion has been spent for this purpose (*Napi Világgazdaság*, April 10, 1992).

weight and development potential, high growth rates in these countries would substantially influence global economic prospects.⁴ It should be stressed that the import reorientation of the Central and Eastern European economies from their traditional Soviet supply sources for raw materials and energy could exert particularly beneficial effects on developing country exports to this area.⁵ Expanding world market orientation and developing intraindustry trade between Eastern and Western Europe may offer additional possibilities.

The EC's agreements with Czechoslovakia, Hungary, and Poland, as well as the extension of the generalized system of preferences designation to Bulgaria and Romania, are considered particularly detrimental to the developing countries. Most of these changes, however, simply put an end to the formerly adverse position of these economies. The absence of trade between the two parts of Europe was one reason why Western Europe's share of world markets declined during most of the 1980s.⁶

In sum, the 1990s and beyond promise both more possibilities for cooperation and growing competition between the two groups of countries. Rapid world economic growth might benefit everyone, even if some countries gain more than others. Nevertheless, competition is increasingly global, so the losers may come from the industrial world.

V. RESOURCE DISTRIBUTION

The question of how resources are distributed within the economies in transition is timely, given the increasing help provided to the successor states of the U.S.S.R. In political terms East Central European countries fear that the West's strong contacts with the former U.S.S.R. will lead it to neglect the in-between area and will thus create an extremely dangerous vacuum in Europe. In economic terms most of the small countries in the region, particularly the more developed ones, are concerned that their development prospects might be curtailed. Indeed, the successor states have received \$24 billion from the Group of Twenty-four to date and may get another \$20 billion this year. According to the *Financial Times* (April 16, 1992), the International Monetary Fund (IMF) will invest \$25 billion to \$35 billion and the World Bank \$12 billion to \$15 billion in these countries during the next four years.

In this respect, the costs of German reunification are a particularly timely lesson (Inotai 1992). Diversion of real resources is only part of the problem,

4. According to the IMF's managing director, Michel Camdessus, economic recovery in the successor states of the U.S.S.R. might add 2 percent to predicted world economic growth rates (*Napi Világ-gazdaság*, April 17, 1992).

5. Hungary's trade balance, traditionally in surplus vis-à-vis developing countries, recorded a deficit of about \$300 million in 1991. As oil and raw material (and later gas) purchases are expected to be diverted to non-Soviet sources, this trend may become characteristic.

6. While the United States and Japan rely on international subcontractors in medium-income developing countries in Asia and Latin America, the EC has concentrated in part on intra-Western European division of labor, with clear (labor) cost disadvantages, and in part on Mediterranean and African countries, which are not able to compete with more developed Asian and Latin American producers.

even if its impact is extremely adverse for East Central Europe. In comparison with the former German Democratic Republic (GDR), these countries are better prepared to implement a functioning market economy within a relatively short period—and at much lower cost—because they can absorb resources more efficiently. Consequently, they could become the success stories for other Eastern European economies despite the pains and burdens of the process. Any setback or failure of reform policies in East Central Europe (or in any of these countries) may discredit market economy concepts throughout the region for the foreseeable future.

More important, any failure of the reform process in the former U.S.S.R. might put East Central European countries in the same box of “countries without a future.” Furthermore, the costs of transformation would be measured on the basis of the experience of the former U.S.S.R. because some analysts incorrectly assume that the developing East Central European economies would need more money than the former GDR.

But the main problem is that ambitious aid programs to the former Soviet republics may seriously undermine exports from the economies in transition—exports that have already largely been destroyed by the collapse of the U.S.S.R. Aid packages often consist of drugs, agricultural goods, and textiles, all of which have been produced by the economies in transition for sale to the former U.S.S.R. Replacement of their goods with free Western products is likely to cause economic problems to deepen.

VI. CONCLUSIONS

With competition for international financial flows, technologies, and markets becoming increasingly global, the simplified model of winners and losers cannot be applied to the developing countries and the economies in transition. Both may be winners as well as losers. Nor will competition leave the OECD countries unscathed. Both groups of countries are far from homogenous in their development level, speed of recovery, and long-term growth prospects. Some will face the challenge successfully; others are likely to be crowded out.

It is extremely instructive that not all countries fear growing competition. For instance, there are few, if any, complaints against support to the economies in transition from the Asian NIES. Anxiety is more pronounced in countries that are less competitive and that have more problematic—and less efficient—economic policies. Many African and Latin American countries, as well as EC countries, seem to fear competition more than the East Asian countries.⁷

The dramatic political and economic changes in Europe are expected to bring about particular attention by Western Europe to the economies in transition. Resource flows and trade preferences granted by industrial countries regularly

7. The poorer EC member countries enjoy Community-level financial support that is 15 times higher per capita than that granted to Central and Eastern European countries. (Shrinking per capita income in the past two years has been widening this gap.) See *Handelsblatt*, April 9, 1992.

show a clear regional bias. The United States gives preference to the Latin American and Far Eastern developing countries, as well as to its main allies in the Middle East. Japan focuses on Asian countries. Oil-rich Arab countries have provided financial support to less-rich Arab economies. Traditionally, Western Europe has established preferential relations with Africa and the Mediterranean region. Central and Eastern Europe have been entering this mainly European (and within it, German) picture recently.

The prospects for investment flows depend only marginally on the possible diversion of resources. Beyond the respective national economic policies pursued, the decisions will originate in the leading industrial countries. First, to create additional sources of financial support, military expenditure should be substantially cut and the saving (at least in part) dedicated to external financing. Second, a decline in international interest rates would benefit both groups of countries. Third, and most important, the global trade liberalization that will follow the ongoing Uruguay Round would not only clearly offset the potentially adverse effect of resource diversion but would also help stimulate additional capital formation and growth.

REFERENCES

- Collins, Susan M., and D. Rodrik. 1991. "Eastern Europe and the Soviet Union in the World Economy." Institute of International Economics, Washington, D.C.
- Fidler, Stephen. 1992. "Capital Tan Turned on." *Financial Times*, April 6, sect. 3, p. 1.
- Graham, George. 1992. "CIS May Need \$20 bn Extra Aid." *Financial Times*, April 16, p. 18.
- Handelsblatt*. 1992. "Deutsches Institut für Wirtschaftsforschung. Vertrag von Maastricht: Die Ausblendung von Osteuropa war ein schwerer Fehler." April 9. Düsseldorf, Germany.
- Inotai, András. 1992. *The Economic Impact of German Reunification on Central and Eastern Europe*. Washington, D.C.: American Institute for Contemporary German Studies.
- Napi Világgazdaság*. 1992. "Magyarország már nem szorul segítségre. A huszonnégyek Kelet-Európáról" [Hungary does not need help. The G-24 on Eastern Europe]. April 10, p. 2.
- . 1992. "145 milliárd dollár a szovjet utódköztársaságoknak? Az IMF döntött a kvótákról" [\$145 bn to the successor states of the U.S.S.R.? IMF decision on entrance quotas]. April 17, pp. 1–2.
- Ochel, Wolfgang. 1991. "Der Kapitalbedarf Osteuropas und der Sowjetunion." *IFO-Schnelldienst* 44, 31:3–8.
- Schroeder, Klaus. 1991. "Westliche Finanzhilfen für Reformen in Osteuropa—Bedingungen und Risiken." *Aussenpolitik* 4:326–35.
- UNCTC (United Nations Centre on Transnational Companies). 1991. "Courting Foreign Direct Investment in the Nineties." *Economic Reform Today* (Fall):13.

COMMENT ON "CAPITAL FLOWS TO DEVELOPING COUNTRIES: IMPLICATIONS
FROM THE ECONOMIES IN TRANSITION?" BY COLLINS

Ishrat Husain

Susan Collins's analysis of the likely effect of the economies in transition on capital flows to developing countries is based on two sets of empirical estimations. In the first single-regression equation, she explains the amount of average official development assistance (ODA) to countries during 1987-89 in terms of per capita income and log of population. These two variables together explain half the average variation. She then asks how much ODA the economies in transition would receive if they were treated "like" countries with similar characteristics, using the earlier cross-country regression results. On the basis of historical evidence about how increases in assistance to China, Egypt, Israel, and Bangladesh affected ODA disbursements to other recipients, she concludes that if the economies in transition were treated "like" developing countries, they would receive about \$8.2 billion in concessional assistance annually.

I have a serious problem with this methodology and therefore with the strength of the paper's conclusions. First, attempts to measure empirically the relative influence of the fundamental factors, even for private capital flows in international markets, have not proved very successful. Collins's effort to measure the determinants of aggregate official capital flows by using twenty-four countries with different motivations and underlying principles—most of them not even obvious to their own bureaucrats—is really heroic. Two donors, the United States and Japan, account for 40 percent of ODA today. In the period Collins refers to, five countries—Israel, Egypt, Pakistan, El Salvador, and the Philippines—accounted for 70 to 75 percent of U.S. ODA disbursements, and nine Asian countries received 80 percent of total Japanese bilateral disbursements. Neither population nor per capita income satisfactorily explains the behavior of the U.S. and Japanese flows. In order to better understand the determinants of ODA, I would guess that we would have to prepare a 24 x 24 matrix in a "sources and uses of funds" framework rather than a simple regression equation for a three-year period at an aggregate level.

Second, capital flows can only be meaningfully analyzed in the framework of a country's macroeconomic aggregates, domestic savings, and investment

Ishrat Husain is chief economist in the World Bank's Africa Regional Office.

© 1993 The International Bank for Reconstruction and Development / THE WORLD BANK

behavior. The pattern of the 1980s is not a valid or reliable guide for the coming decade. More than fifty developing countries are currently undertaking serious policy and institutional reforms to make their economies more efficient, flexible, and competitive. Some of these changes are structural. The decline in inflation, the cut in budget deficits, the reforms of public enterprises, trade liberalization, and other policy improvements should contribute to a strong recovery of domestic savings and investment in these countries. Total factor productivity is expected to respond to improved policies; IMF staff have estimated that the contribution of total factor productivity to the growth of potential output in net debtor countries may rise from an average of 0.5 percentage point in 1983-90 to 1.75 percentage points in 1993-96.

Third, the paper is conspicuously silent on the effects of the enormous debt burden in Latin America and Africa in the 1980s, which resulted in net negative transfers of \$35 billion a year. I think we have to assume that these large transfers were a disincentive to investment and external financing. The consequences of debt (and debt service) reduction will save the countries involved at least \$6 billion to \$8 billion a year. It is also estimated that their investment-output ratio will rise by more than 2 percentage points. What is more important, the reduction in uncertainty and the restoration of confidence among both domestic and foreign investors will have a positive influence on other economic and financial variables. The combination of an easing of the debt burden and policy improvements should boost investment, productivity, and the efficient use of resources. It can be expected to result in higher domestic savings and reduced demand for external resources for the same given level of output, an outcome that is radically different from the pattern in the 1980s.

Fourth, capital flows at the global level cannot be estimated apart from trade flows. A recent study by the World Bank estimates that a 50 percent liberalization of industrial country trade barriers would raise developing country exports by \$50 billion, almost equivalent to the aggregate net resource flows from official sources. Assuming that progress in dismantling trade barriers is not as rapid, even partial liberalization would increase exports by \$10 billion to \$12 billion a year without adversely affecting the budgets of the Organization for Economic Cooperation and Development (OECD) countries.

Finally, on the supply side, the quantum leap in the scale and efficiency of international financial intermediation, the development of liberalized international financial markets, and the integration and globalization of offshore and domestic financial markets will have positive (although modest) beneficial effects in some advanced developing countries.

This brings me to the final issue: how do I perceive the pattern of capital flows in the 1990s? I would start by disaggregating the developing countries into three groups. The first group consists of the successful adjusters. These economies will continue to have access to private capital flows (provided they sustain the adjustment reforms to which they are committed), and they have a much better chance in the 1990s of benefiting from the repatriation of flight capital, from

official export credits, and from nonconcessional multilateral loans. The same economic policies adopted to improve domestic economic performance will also enhance their creditworthiness. The experiences of Chile, Mexico, and Venezuela, which emerged from a debt overhang and gained renewed access to international capital markets in 1990-91, suggests that at least two factors are necessary: strong implementation of an adjustment program and sound prospects for external viability, through either debt reduction or rapid growth. With these factors in place, access can be helped by addressing investor concerns about the risk of default, the risk of nonavailability of foreign exchange, and the ability to trade debt instruments (liquidity risk).

The second group includes the low-income countries that rely on concessional loans from multilateral institutions and bilateral grants. The record of ODA shows that, with some exceptions, this group of countries will be the focus of attention and concessional assistance. Donors are, however, demanding improved economic performance in return. I believe the projected increase in real ODA of 2 to 3 percent a year and some reallocation of ODA from successful middle- and low-income countries that will have access to other sources of financing will be sufficient to meet their requirements without any serious disruption. Again, the key preconditions are improved domestic performance, an early resolution of debt problems, and a benign external environment. If these conditions are not fulfilled, I doubt that the current levels of ODA can be maintained.

The third group comprises the former U.S.S.R., excluding Eastern Europe and the republics of central Asia. Although financial assistance to these countries is important, a more difficult challenge is how quickly we can help establish macroeconomic stability, reasonably functioning institutions, and incentives for production. Unless some progress is made on these fronts, the transfer of financial resources will be of little help. I am not certain what constitutes an adequate level of capital flows to this group, but judging by the recent \$24 billion package announced by the Group of Seven for the Russian Federation, it appears that most of the resources will be provided by such nonconcessional multilateral sources as the IMF, the World Bank, the European Bank for Reconstruction and Development, and the European Community and through officially guaranteed export credits (at market terms), surplus commodity and food aid, and very little concessional aid.

A 1991 OECD survey shows that only \$1.5 billion of concessional aid has been committed so far for East and Central Europe by Development Assistance Committee members, the bulk of it through additional budgetary allocations rather than from the ODA budget for developing countries. To my mind, a more worrisome trend is that several developing countries have recorded an erosion in per capita incomes during the past decade and are now eligible for concessional assistance. The competition from this group of countries in a stagnant pool of concessional resources is more disconcerting than the imminent danger from the economies in transition in Eastern Europe and the former U.S.S.R. Although the international community should monitor official capital flows to ensure that diversion does not take place, the conclusions drawn to date are not yet persuasive.

Conversion of Official Bilateral Debt: The Opportunities and the Issues

Stephany Griffith-Jones

The level and structure of external debt limit the growth prospects of many severely indebted low-income and lower-middle-income countries. Because many of these countries owe a large proportion of their external debt to official creditors, mechanisms such as debt conversion can be useful in reducing their debt overhang; important efficiency gains can be achieved if the use of local currency linked to the debt reduction is productively channeled. This paper examines the lessons learned from swaps of commercial debt and evaluates the potential benefits and costs of swaps of official debt. For debtors, the potential benefits associated with a reduction of debt include an improved climate for domestic and foreign direct investment, a transfer of risk to foreign investors, and access to a source of additional capital for privatization, as well as additional external funding for social programs. The problematic effects include the risk of inflation. For creditors, debt-for-equity and debt-for-development conversions are one way to help countries achieve long-term financial viability and at the same time prompt them to undertake socially desirable programs. The paper also addresses technical issues, such as transparency of operations, ownership of converted claims, and the need for financial intermediaries.

Even as dramatic increases in capital flows and foreign exchange reserves in countries such as Chile and Mexico imply that debt crises are a problem of the past, many other countries are still struggling with a high level of external debt that limits their prospects for growth and development. Because a large proportion of this debt is owed to official bilateral creditors, mechanisms such as debt conversion can be useful, and important efficiency gains can be achieved if the use of local currency linked to the debt reduction is channeled productively.

Section I examines the need and the potential for official debt conversion in heavily indebted low-income and lower-middle-income countries and describes

Stephany Griffith-Jones is a fellow of the Institute of Development Studies, University of Sussex.

The paper draws on the joint work by Griffith-Jones and Percy Mistry commissioned by the Italian government for the Secretariat of the United Nations Conference on Trade and Development (UNCTAD). The author is grateful to the many officials and academics in creditor and debtor countries, as well as in international organizations who commented on the paper and particularly to Percy Mistry for his valuable insights.

current practices in the Paris Club (where official bilateral debt is rescheduled or reduced, or both). Sections II and III review the lessons of commercial and official debt conversions, primarily in middle-income countries, evaluate the potential benefits and costs for debtors and creditors, and address technical issues, such as transparency, ownership of converted claims, and the need for financial intermediaries. Section IV presents conclusions and policy recommendations.

I. THE NEED, THE POTENTIAL, AND THE REALITY OF OFFICIAL BILATERAL DEBT CONVERSION

Between 1982, when the debt crisis exploded, and 1991, total external debt stocks in developing countries grew rapidly, particularly in severely indebted low-income countries. Growth of total debt outstanding and disbursed (referred to in this paper as "debt") more than doubled in these countries, climbing from \$79 billion in 1982 to \$175 billion in 1991 (see table 1). Total interest arrears in this period also increased significantly for these countries, rising from \$1.3 billion in 1982 to \$12.9 billion in 1991, as many low-income countries found themselves unable to service their debt. This debt is a greater burden for the poorest countries than for countries with less severe difficulties, for two reasons. First, the level of debt implies contractions of output and income that are particularly damaging to human welfare. Second, structural weaknesses in many of these countries make adapting to changes in the international environment difficult.

As a result of these problems—and in some cases, of mistaken policies—many of these countries sustained a decline in export performance, which led to a greater deterioration than would have otherwise occurred in the debt service ratio and the debt-to-export ratio (see table 2).¹ By the late 1980s both ratios were higher in the severely indebted low-income countries than they had been earlier in the decade: the debt service ratio in 1989 was double the level of the 1980s, while the ratio of debt to exports was almost five times higher.

This situation contrasts with that of the severely indebted middle-income countries (especially the upper-middle-income countries). Their total debt started to fall in 1991 (see table 1), primarily as a result of actions taken within the framework of the Brady Plan—which reduced private debt—and of a large number of debt conversions. As table 2 shows, these countries also had a better export performance during the 1980s, posting an average growth rate of 3.5 percent. As a result, debt service ratios for this group have declined quite substantially, although they are still fairly high.

The rapid rise of bilateral debt has been an important element in the increase in developing country debt since 1982. In severely indebted low-income countries, bilateral debt rose from \$32 billion in 1982 to \$80 billion in 1991, or more

1. The debt service ratio is defined as loan amortizations plus loan interest payments divided by the level of exports.

Table 1. Growth in Debt Stocks of Severely Indebted Developing Countries
(billions of U.S. dollars)

Category	1982	1985	1988	1991 ^a
<i>Low-income-country debt^b</i>				
Total	79.12	116.74	160.91	175.35
Interest arrears	1.25	3.09	8.00	12.89
Bilateral	32.06	49.45	75.10	80.43
Multilateral	10.61	16.76	26.77	34.73
IMF	3.23	5.56	5.83	6.30
Private guaranteed	18.32	24.11	32.24	29.18
Private unguaranteed	2.90	3.33	3.09	2.88
Short-term	12.00	17.53	17.88	21.83
<i>Middle-income-country debt</i>				
Total	346.15	418.46	485.13	486.54
Interest arrears	4.53	4.81	15.59	40.50
Bilateral	35.23	62.58	90.58	99.59
Multilateral	15.94	24.35	41.71	53.87
IMF	6.97	12.99	14.71	17.54
Private guaranteed	183.49	223.00	255.34	220.69
Private unguaranteed	61.70	50.96	29.17	23.10
Short-term	48.82	44.57	53.62	71.76
<i>Others^c</i>				
Total debt	55.00	50.00	69.00	65.00
Total debt	846.00	1,046.00	1,262.00	1,351.00
Interest arrears	6.10	8.67	25.64	55.50

Note: IMF, International Monetary Fund.

a. Estimated.

b. The estimates for 1991 for this group are higher than those projected in the *World Development Report* because data from other sources suggest that the *World Debt Tables* understates 1991 debt. (It has consistently understated the latest year's estimates.) The 1991 estimates for the middle-income group remain unchanged. No data are included for low- or middle-income countries that do not report to the World Bank but whose debt the *World Debt Tables* estimates in aggregate form (see World Bank 1991, table 1.1).

c. Afghanistan, Albania, Cuba, Iraq, Democratic People's Republic of Korea, Mongolia, Viet Nam, and about thirty island microstates in the Caribbean and the South Pacific.

Source: World Bank (1991).

than 45 percent of the total debt, despite the cancellation of \$8 billion in official development assistance debt from 1983 through 1990 and the successive application of concessional terms granted by the Paris Club under Toronto, Venice, Houston, and, now, "enhanced Toronto" agreements. In the severely indebted middle-income countries, bilateral debt jumped from \$35 billion in 1982 to \$108 billion in 1990 (it declined somewhat in 1991). The increase reflects the effect of exchange rate changes since 1985 and the interest capitalization practices of the Paris Club.

A further source of concern is the increase in multilateral debt. In low-income and middle-income indebted countries, outstanding debt obligations to multilateral creditors rose sharply (see table 1). It is not clear what can and should be done to reduce the multilateral debt burden, given the need for creditworthiness in international capital markets, but the size of multilateral debt service payments

Table 2. Structural Features, Export Growth, and Debt Indicators, Severely Indebted Countries

<i>Item</i>	<i>Severely indebted low-income countries</i>	<i>Severely indebted middle-income countries</i>
GNP per capita, 1988 (U.S. dollars)	288	1,632
Infant mortality, 1987 ^a	102.8	55.0
Annual growth of exports, 1982-89 (percent)	-2.0	3.5
Debt service ratio ^b		
1980	10	36
1982	20	49
1989	23	29
Ratio of debt to exports		
1980	96	196
1982	214	297
1989	493	294

a. Deaths per 1,000 live births.

b. Payment of loan amortization and interest divided by export earnings.

Source: World Bank (1990).

increases the need to reduce and convert bilateral debt, particularly for severely indebted low-income and lower-middle-income countries.

In this sense, export credit agencies need to accept the realities that commercial banks have recognized and offer not just debt cancellation but also debt conversion options on a scale which would reduce bilateral debt to levels that can be serviced. Debt conversion options should be used not just where debt reduction has been insufficient but also to enhance other gains (including efficiency) when undesirable effects—on inflation, for example—are marginal or can be easily counteracted by government policy.

As for official bilateral debt, a number of measures for debt reduction have been and are being implemented for severely indebted low-income countries, but there is evidence that for an important number of those countries progress is still insufficient. (For good discussions of these measures, see World Bank 1989, 1990, 1991; Mistry 1992.) It is disappointing that neither Britain's 1990 proposal (known as the Trinidad terms) nor the Dutch proposal was adopted by the Paris Club. The consensus that was reached in December 1991 (enhanced Toronto terms) and that has already been applied to Benin and Nicaragua dilutes the Trinidad terms considerably. Under the terms of the enhanced Toronto consensus the creditors have several options: canceling 50 percent of eligible maturities; halving interest rates on nonconcessional debt; rescheduling export credit and concessional debt repayments; and capitalizing reduced interest rates in a way that would result in equivalence in net present value terms with the other options.

For severely indebted lower-middle-income countries, the Paris Club agreed in September 1990 to lengthen grace periods and maturities on the basis of three criteria—low per capita income, a high ratio of Paris Club debt to commercial

bank debt, and a heavy debt (and debt service) burden as measured by ratios of debt to gross national product, debt to exports, and debt service to exports. The "10 percent clause," a debt conversion mechanism, was also introduced. The clause allows "creditor countries, on a voluntary and bilateral basis, to exchange up to 10 percent of bilateral official or officially guaranteed nonconcessional loans, and up to 100 percent of official development assistance loans, for debt-equity swaps, debt-for-nature swaps, and debt-for-development swaps." There is also a value limit (\$10 million or \$20 million, depending on the case) that can be used when 10 percent is less than the bilateral nonconcessional debt. In December 1991 the same clause was extended to the severely indebted low-income countries.

Initially, swaps of official bilateral debt were practically nonexistent; indeed, there were limits on debt sales by creditor governments. But emphasis is rapidly shifting toward bilateral official debt conversions for equity as well as for development. Such operations potentially open debt conversions for low-income and lower-middle-income countries, as well.

By early 1992 conversions under this clause were approved by the Paris Club in the cases of Benin, Congo, Côte d'Ivoire, Ecuador, Egypt, El Salvador, Honduras, Jamaica, Morocco, Nicaragua, Nigeria, Peru, the Philippines, Poland, and Senegal. Relatively few conversion transactions had actually taken place under the "10 percent clause," but a number of actions were reportedly being considered, including the following.

- Funding for a \$3 billion Environment Fund in Poland by the United States and France
- A conversion of up to \$10 million of Egypt's bilateral debt by the French government for cofinancing a Social Emergency Fund (with the World Bank). In addition, France and other creditor governments are considering official debt-equity conversions.
- Debt-equity swaps for Morocco by the Netherlands (and some other creditor governments)
- A proposal to use debt-equity swaps to support privatization in Nigeria
- A conversion of official debt by the government of Canada for United Nations Children's Fund (UNICEF) spending programs in Bolivia

Even before September 1990 some European and North American governments were selling (or converting) their Paris Club debt to improve the balance sheets of their export credit agencies. Because these operations were not allowed in the Paris Club framework, they were not publicized. They are, however, interesting prototypes for official debt-equity swaps, showing that it is feasible for an export credit agency to take an equity position in developing country companies, sell official debt to private investors, or both. One creditor agency converted Mexican debt into equity in a private steel company that it later sold, recovering the full face value of its claim.

The U.S. Initiative

Just before the Paris Club initiative was launched, a three-pronged U.S. "Initiative for the Americas" was proposed, including trade and investment measures, and debt concessions. Under the debt sections of the program the United States agreed to reduce stocks of concessional debt (PL 480 and U.S. Agency for International Development obligations) owed by Latin American and Caribbean countries and to accept interest payments in local currency on the remaining debt, to be paid into a fund for the environment. (If the country has not entered into an Environmental Framework Agreement, interest is to be paid in U.S. dollars.) In other words, a commitment to allocate domestic resources to the environment is exchanged for debt reduction. The U.S. Congress has also broadened the use of interest payments for domestic development—specifically, for programs to benefit children.

The Economic Commission for Latin America and the Caribbean (ECLAC 1991) reports that Washington has canceled large portions of the outstanding (primarily concessional) debts owed by Guyana, Honduras, Nicaragua, Haiti, and Bolivia, as well as smaller (in percentage terms) amounts owed by Jamaica and Chile. Implementation of the environmental fund is gradually beginning. Chile—the first country to be granted this concession, in June 1991—, at the time this paper was written, in the process of defining its agreement on the environment.

As for nonconcessional debt owed by eligible Latin American and Caribbean countries to the Eximbank and the Commodity Credit Corporation, the U.S. initiative contemplates sales of a portion of the debt to facilitate debt-for-equity, debt-for-development, or debt-for-nature swaps; these swaps would imply both a conversion and a reduction of debt. Legislation to approve such operations has been seriously delayed, however. The passage of an appropriations bill to fund additional reductions in PL 480 debt is also problematic.

Other Official Debt Conversion Initiatives

Outside the framework of the Paris Club, some developing countries, especially Mexico, have pursued an active strategy as creditors to convert the debt of Central American and Caribbean countries. In three different operations, for example, Mexican investors purchased a privatized company in Honduras, leased farming land in Nicaragua, and agreed to build new hotels in Costa Rica. In addition, debt owed to Central and Eastern European countries and the former U.S.S.R. can be converted, as was done in the sale of loans from the former German Democratic Republic to commercial firms engaged in importing raw materials. Commercial firms are reportedly recovering the full face value of the debt through imports of raw materials from countries such as Zambia. The outcome in this case seems very undesirable, as it implies full prepayment of official debts when Zambia is not even *servicing* the rest of its bilateral debt in full.

II. LESSONS FROM CONVERSIONS OF PRIVATE COMMERCIAL DEBT

Since 1985 conversions of private commercial bank debt have been used extensively in a number of (mainly middle-income) developing countries to reduce debt, promote foreign investment, encourage privatization, and further other development objectives.

Debt-for-Equity Swaps

Table 3 shows that the estimated total volume of commercial debt eliminated through official debt conversion from 1985 through 1990 was \$33.6 billion, or about 15 percent of the total commercial debt of all heavily indebted countries (see table 4). With the exception of Chile, which converted almost 70 percent of its 1985 commercial debt, debt conversions clearly did not overcome the debt overhang of most countries, but they did make a meaningful contribution in several instances. Argentina and the Philippines were able to reduce more than 30 percent of their commercial debt; for all other countries, conversions represented less than 20 percent of commercial debt.

The evolution of actual conversions of such debt is noteworthy. Table 3 shows a rapid expansion in 1987 and 1988 before some countries became concerned about the domestic monetary implications of these operations and began to slow or suspend debt conversions. In other countries—especially Chile—debt conversions grew so rapidly that their very success reduced the stock of debt available for sale. The revival of debt-equity swaps in 1990 in

Table 3. Volume of Debt Conversion by Country, 1985-91
(millions of U.S. dollars)

Country	1985	1986	1987	1988	1989	1990	Total
Argentina	469	—	—	764	1,180	7,038	9,451
Brazil	537	176	336	2,095	942	483	4,569
Chile	323	974	1,997	2,927	2,767	1,096	10,084
Costa Rica	—	7	89	44	124	17	281
Ecuador	—	—	127	261	31	42	461
Honduras	—	—	9	14	47	32	102
Jamaica	—	—	4	5	16	23	48
Mexico	—	413	1,680	1,056 ^a	532	435	4,116
Nigeria	—	—	—	70	304	217	591
Philippines	—	81	451	931	630	378	2,471
Uruguay	—	—	—	104	53	—	157
Venezuela	—	—	45	49	544	716	1,354
Total	1,329	1,651	4,738	8,320	7,170	10,477	33,685

— Not available.

Note: Face value of debt converted under official ongoing schemes. Figures do not include large-scale cash buybacks and debt exchanges.

a. Does not include an estimated \$6 billion-\$8 billion under an August 1987 agreement to restructure debt.

Sources: Central Bank of Argentina; Central Bank of Brazil; Central Bank of Chile; Mexico Ministry of Finance; Central Bank of the Philippines; Bank of Jamaica; Central Bank of Venezuela; International Monetary Fund.

Table 4. Contribution of Debt Conversion Programs to Reduction of Commercial Bank Debt
(billions of U.S. dollars)

Country	Stock of commercial bank debt, 1985	Value of debt conversions 1985-90	Conversions as a percentage of commercial debt
Argentina	25.3	9.5	37.5
Brazil	67.1	4.6	6.9
Chile	14.8	10.1	68.2
Mexico	71.4	4.1	5.7
Nigeria	4.9	0.6	12.2
Philippines	7.6	2.5	32.9
Venezuela	23.6	1.4	5.9
Total	214.7	32.8	15.3

Sources: World Bank (1990); table 1.

countries such as Argentina, Mexico, and the Philippines has been largely linked to privatization because these conversions do not lead to the monetization of foreign debt. Furthermore, if privatization produces efficiency gains, the debtor government saves the revenues previously spent to subsidize inefficient public enterprises. (Of course, if the government swaps debt for equity in profitable enterprises, the reduction in income from those enterprises could have a negative fiscal impact, as discussed in Corden and Dooley 1989.) In addition, several recent agreements to restructure bank debt (especially in the context of the Brady Plan) contain commitments to swap debt for equity.

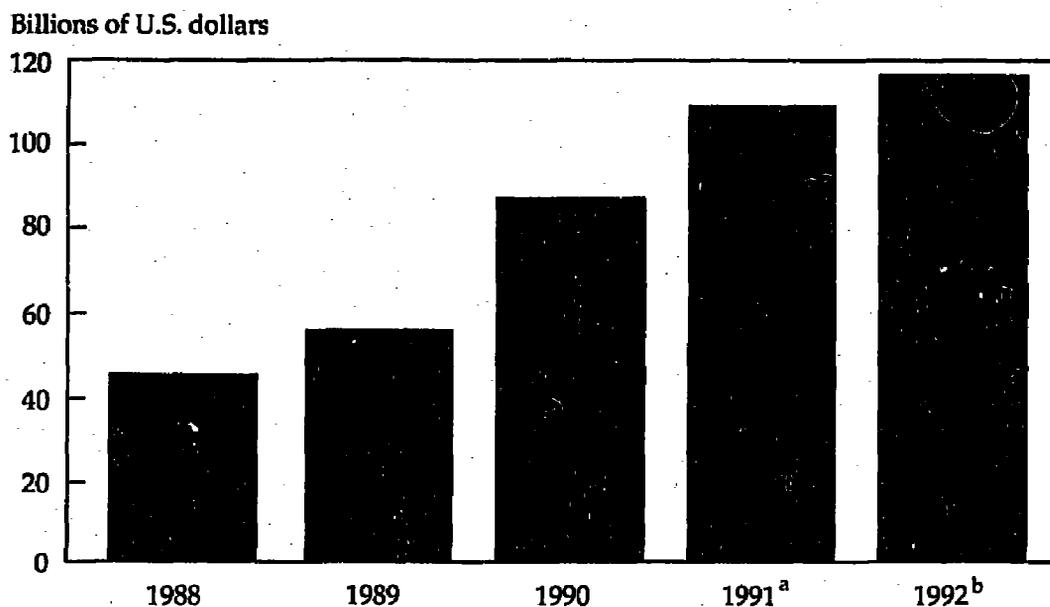
The increased use of market-based debt reduction techniques has been facilitated by—and has contributed to—marked growth in the size of the secondary market. As figure 1 shows, total trading volume in 1990 reached about \$100 billion. (Nederlandse Middenstand Bank, the largest European trader, estimates \$150 billion.) This contrasts with levels in 1983 or 1984, when trading was only \$0.5 billion.

Finally, there has been an effort to streamline and simplify financial procedures, particularly the documentation required for carrying out swaps. That all post-Brady bonds are tradable and assignable has played a major role in facilitating and expanding the volume of transactions. The impressive development of the secondary market for commercial debt points to the potential for similar systems to swap official debt for equity or for development.²

Country results. The economic effects of debt conversions are heterogeneous among countries and sometimes within the same country. Factors that seem to contribute to positive results include a stable macroeconomic environment with low deficits; adequate domestic capital markets that can attenuate or eliminate monetary effects; well-designed debt conversion programs aimed at meeting

2. For a useful analysis of the issues, see Blackwell and Nocera 1988 and Frenkel, Dooley, and Wickham 1989. For a review of debt swaps and policy lessons, see UNCTC 1991; Bouzas and Ffrench-Davis 1990; and Mortimore 1991. For an analysis of Chile's experience, see Larrain and Velasco 1990; Williamson 1990; Lagos 1989; and Aravena 1991.

Figure 1. Total Trading Volume in the Secondary Market for Debt



a. Estimated

b. Projected

Source: *Latin Finance* (October 1991).

objectives such as debt reduction or foreign direct investment; and control of undesirable outcomes, such as excessive monetary expansion or misuse for "round-tripping."³ When the policy framework, the circumstances, and the program design are right, debt-equity swaps have yielded positive results, as outlined below.

- *Major reductions in commercial debt.* In Chile, for instance, swaps helped reduce the debt overhang significantly and enabled the country to regain access to international capital markets. In other cases the effect has been less meaningful (see table 4). Mexico's access to international capital markets was only partly advanced by debt conversions. Both debt reduction and the renewal of capital inflows led to a reduction—or a reversal—of negative net transfers in several cases (for example, Chile and Mexico).
- *Foreign direct investment and a reversal of capital flight.* An important bonus of debt-equity conversions has been their contribution to an increase in foreign direct investment. There is some debate in the literature about how much of this flow is additional; the answer depends on assumptions about how much foreign capital would have entered in the absence of the conversion program. Experience in Chile and Mexico, however, suggests that conversions have influenced potential investors and accelerated the pace of foreign

3. Round-tripping is the practice of converting debt with the sole purpose of obtaining local currency to purchase foreign exchange in the parallel market.

investment. Policymakers stress that favorable publicity linked to debt-equity programs (in the financial press, for example) about the country's economic performance, business climate, and economic openness plays an important indirect role in jump-starting investment flows. Two caveats are important here. First, debt conversion will be effective if it is part of a policy package that makes the country attractive to investors. Second, there may be some tradeoff between applying selectivity criteria to enhance the positive developmental and macroeconomic effects (by demanding new flows to accompany debt conversion, as in Argentina, or restricting the sectors eligible for swaps, as in the Philippines) and the magnitude of the debt conversions carried out.

Debt conversion also can facilitate the return of capital flight. Chile's experience is interesting because a special window was opened to encourage investments by residents. This facility, which provided a smaller subsidy than that for foreign investors but offered an implicit tax and legal amnesty, was successful in attracting returned capital.

- *Export promotion and import substitution.* To the extent that the additional investment goes to tradable sectors (and especially if it brings know-how, additional markets, and more efficient technology), foreign exchange earnings and savings will improve. There is some evidence that an important share of foreign direct investment entering through debt conversions (especially in countries such as Mexico and the Philippines, whose governments had targeted this objective) has gone into such activities.
- *Privatization.* As pointed out above, debt conversion provides additional equity for the companies involved. In some countries (for instance, Chile) it lowers the debt of state-owned enterprises and makes them more attractive to potential private shareholders.
- *Private sector finance.* In countries such as Chile, Brazil, and Ecuador, the debt crisis precipitated problems in the private sector, especially the financial sector. Debt conversion strengthened the private sector, lowered excessive levels of debt, and contributed to a recovery of domestic private investment.

Finally, as regards the positive effects discussed above, transparent government procedures are important to ensure that there are no hidden subsidies to foreign or domestic investors, as was reportedly the case in several debtor countries.

Debt conversion also has problematic effects, although they can be offset through efficient program design and implementation. The first potential problem, inflation, can occur if the swaps are large, debt is exchanged against local currency, the increase in the money supply is not regulated, and compensatory fiscal and monetary measures are neglected. Small swaps (in relation to the money supply) are more easily handled, especially if the expansion of the money supply and the fiscal deficit are small. Brazil's experience, however, is a warning that in a context of high inflation and high budget deficits, conversions can aggravate an already serious problem. If the conversion is made against bonds

placed in the domestic capital market, the monetary impact is diminished, but interest rates may increase.

A second risk is that the net effect on the balance of payments could be negative if none or only a portion of the converted debt was previously serviced, if there was considerable round-tripping and the foreign investment was not additional, and if the flow of profit remittances and capital abroad were higher (on a net present value basis) than the interest and amortization payments that would otherwise be paid.

A third concern is that debt-equity conversions normally imply a subsidy, either to a foreign investor or (less frequently) to a resident. This could lead to an inappropriate allocation of resources unless the operation implies important net efficiency gains. In such cases the magnitude of the subsidy can be regulated by the central bank through market mechanisms (for example, an auction) or through administrative measures, such as fixing a lower value for the local currency swapped per unit of debt.

Debt-for-Development Swaps

Another innovative approach, which can be broadly defined as debt-for-development, involves a wide variety of swaps. Although most of the publicity, and a large share of operations, has focused on debt-for-nature swaps, some broader pioneering operations are under way. In one instance, six banks (from three countries) donated to UNICEF debt obligations in the Sudan valued at more than \$20 million. These funds will be used for clean water projects, with beneficial results for health and the environment.

Commercial debt-for-development swaps can originate in purchases or donations. Frequently, international charitable organizations or creditor governments purchase commercial debt on the secondary market and convert it to local currency. In other cases, banks have donated debt to an international charity or nongovernmental organization, with the condition that the debt be "paid" in local currency for conservation or social programs. The total face value of commercial debt-for-development transactions is only \$500 million to \$600 million, an incredibly small amount in relation to the total commercial debt of developing countries and only 1 to 2 percent of total debt-equity swaps (see table 3).

Experience suggests that the main benefit of such deals is that they emphasize high-priority sectors, such as social spending, and thus may serve as a catalyst for shifts in priorities on the part of debtor governments, donors, and other international agencies. Particularly in countries where such areas have been neglected, this shift will imply important efficiency gains. Furthermore—given favorable publicity, the multiplying effect of donor or creditor effort, and the assurance that the contribution will be channeled to high-priority spending—debt-for-development swaps may encourage additional foreign exchange flows.

Debt-for-development flows tend to have more favorable foreign exchange effects than those resulting from debt-equity swaps; in the former there is no

outflow of profit remittances and capital amortization to offset the reduction in debt service payments (assuming oversight by central bank authorities to prevent abuse of such operations for round-tripping). From the perspective of a debtor country, debt-for-development swaps are more favorable if there is a clear net foreign exchange saving. This occurs, in particular, if the country was (or was about to start) servicing the debt. Otherwise the principal gain is the shift in spending to high-priority sectors.

Given the relatively small magnitude of most commercial debt-for-development swaps so far, the risks of inflation have been marginal. This is particularly true where inflation and budget deficits are low and where local currency proceeds from debt-for-development swaps are regulated in time (for example, through bond issues, as in Costa Rica). Furthermore, if there is a net foreign exchange saving, it will generate a contractionary effect on the money supply when it is used for more imports. Debt-for-development swaps are more attractive from a macroeconomic perspective for countries with relatively low inflation that are servicing (and planning to service) most of the category of debt being swapped. For countries with high inflation, special efforts need to be made to compensate for or sterilize the fiscal and monetary effects if these are meaningfully large.

III. CONVERSIONS OF OFFICIAL DEBT

Although we have had extensive experience with commercial debt swaps, the history of official debt conversions is limited. A significant difference between the two seems to be that debt reduction is more clearly additional in official debt transactions. For commercial debt, countries have other options (Brady Plan deals and debt buybacks). In the case of bilateral official debt, because the "10 percent clause" is negotiated after the multilateral deal is reached, debt conversion seems to imply additional debt reduction. Naturally it is important that this concession not be granted as a substitute for deeper debt reduction in the Paris Club. Similarly, debt reduction linked to debt conversion must not imply a reduction in other aid flows (a consideration if debt reduction is funded from aid budgets) or a reduction in new export credit guarantees—an unlikely possibility, given the limited magnitude of swaps.

Under the U.S. Initiative for the Americas, debt conversion also seems to be additional; if the debt is not reduced through this mechanism, it would not otherwise be reduced at all. Debt for equity or for development will have a similar effect on debt reduction, but the effects on the net balance of payments will be different.

The efficiency gains will differ as well. What is common to both, however, is that debt reduction, because it increases foreign investment, privatization, and social spending, implies additional gains in efficiency.

As discussed above, commercial debt-equity swaps have attracted additional foreign investment (although there is debate in the literature about just how much additionality there was). The reasons for additional investment are two-

fold: first, by decreasing the initial total expenditure, the swaps provide an important bonus up front, thus reducing risk and improving the rate of return, and second, debt-equity programs signal that the government is keen to promote foreign investment.

Whether these signals will work in low-income and lower-middle-income countries is not clear. There may be other problems that deter foreign investors (such as a lack of physical infrastructure, a poorly educated work force, and so on), and investment opportunities may be limited. Nevertheless, if the country does want to attract investors and has taken measures to improve the business climate, official debt conversion could be a valuable catalyst. Where capital flight has been significant, official debt conversions could also encourage domestic capital to return. In countries where nationals have been allowed to participate in debt-equity transactions, commercial debt conversion has been a powerful incentive for repatriation of capital.

To ensure these beneficial effects on the balance of payments, the government must take appropriate measures to avoid or limit round-tripping. In Chile, for example, the government controlled the volume of debt-equity swaps by nationals, monitored the parallel exchange rate, and allowed residents to hold attractive domestic securities. The government can also use more direct controls (as in the Philippines) to verify the use of swap resources by examining invoices, receipts, sales agreements, and other relevant documents and by requiring that funds not used for a project be invested in nontransferable central bank bills in local currency. These precautions will increase the willingness of creditor governments to allow their debt reduction to be used for debt conversion.

For debtor governments, the potential for inflation is the most serious constraint. In two cases, however, conversions would have no effect—or only a marginal impact—on inflationary pressures: when external debt is used to acquire domestic physical assets (companies being privatized, or nature reserves set aside for swaps) and when the amount swapped is very small in proportion to the total money supply (as in debt-for-development conversions).

An important final caveat is that inflation is a far more serious consideration in countries already facing high inflation and high fiscal deficits. Low inflation, low deficits, spare capacity, and an elastic supply response make a limited expansion of the money supply less problematic. Furthermore, to the extent that the economy becomes more efficient, the improved supply response may reduce inflation in the future.

Debt conversion represents only one tool for debt-distressed countries. As such, it is very much a complement of—not a substitute for—more important fiscal and monetary macroeconomic policies.

Macroeconomic Policies

Assessing the monetary impact of debt conversions means making some assumptions about whether the country would have serviced that part of its

official debt if the debt conversion had not taken place. Some indication is provided by the country's past record.

As debt is converted, service payments are reduced in that year and in the future, leading to a decline in net monetary expansion. If the debt is swapped for new investment or development spending and is not compensated, the immediate effect will be a monetary expansion that is partly compensated by the contraction in debt service payments. As the debt is reduced, and ultimately amortized, however, the net cumulative effect on monetary expansion can be zero and in time can become negative.

Finally, there is an indirect contractionary effect on the money supply. If part or all of the foreign exchange saved is used to finance imports, the banking system will absorb money from the private sector, reducing net monetary expansion. These imports will attenuate supply bottlenecks and reduce future inflationary pressures.

Where monetary expansion is thought to be excessive, the financial authorities can take a number of measures to reduce, neutralize, or sterilize the effect. First, the central bank can regulate the redenomination rate to define how much local currency it spends for each unit of debt swapped. Second, it can issue long-term bonds in exchange for the debt. Because the principal is not amortized until the bonds mature, the monetary consequences are delayed (and distributed) as the cost of servicing the debt is transferred to the private financial market in the short run. This does, however, have adverse effects on capital markets as government expenditure pushes up interest rates; furthermore, the interest on the bonds is a drain on fiscal resources. Government paper is an attractive option, but it does require a comparatively well-developed domestic capital market (and a limited fiscal deficit). A third alternative would be a monthly quota of swaps. The total swapped can be varied, if necessary, to limit the monetary effects.

In debt-for-development swaps, inflation will usually manifest itself through increased fiscal spending. Such spending can be offset by reducing other government expenditures or by increasing government revenues. It is also possible to place long-term development bonds on private capital markets.

Balance of Payments

Because debt-distressed countries face severe foreign exchange shortages, a crucial effect of debt conversions is that on foreign exchange cash flow. As noted earlier, an important distinction between debt-for-equity and debt-for-development swaps is the positive effect of the latter on this flow. Particularly if the country was servicing (or was planning to service) that debt, the net foreign exchange effect is likely to be very favorable.

It is difficult to estimate the present value of future foreign exchange flows. In a debt-equity swap the net effect will depend on (a) whether that part of the debt was going to be serviced and, if so, whether the transfer would be less (in net present value terms) than the likely future repatriation of profits, dividends, and

capital generated by the foreign investment; (b) how much foreign direct investment is additional (and whether it will help generate other investments); and (c) whether round-tripping is small or can be controlled. In development swaps, the net foreign exchange impact depends on the same factors, except that there is no repatriation of funds.

The net foreign exchange effect can be improved, in part, by government regulations. The real (as opposed to the projected) outcome also depends, however, on events in the future. If the overall economic situation improves, fears of negative effects on the balance of payments can prove largely unjustified. In Chile central bank officials report that foreign investors began in the early 1990s to cut back on profit remittances because the economy was so successful.

If a balance of payments crisis does occur, it is likely to lead to increased profit remittances and capital repatriation, exacerbating the foreign exchange crisis.

The positive effects of debt conversion can be enhanced if they are part of an overall plan that assures sufficient debt reduction to free the economy of the effects of excessive debt and if they are accompanied by a policy package designed to ensure sustainable growth.

In development swaps, too, the net foreign exchange impact of official debt conversion will depend on whether that debt was being serviced; whether the debt reduction is funded by additional contributions or under existing aid budgets; and whether round-tripping is small or can be controlled.

What about the effects of official debt conversions on creditor governments? These can also be positive. First, there is the general point that debt reduction improves the overall value of the residual portfolio of debt (see Claessens and others 1991). Second, in the case of official debt-equity conversions, export credit agencies will sell their debt. If the price at which they sell is higher than the expected net present value of future debt service payments, they will realize a profit. Although this calculation is clear in economic terms, it may be obscured in practice by the accounting and provisioning regulations of some export credit agencies, particularly those that are essentially government departments and that maintain a fictional value for the debt well above its real economic value (for a detailed discussion, see Mistry and Griffith-Jones 1991). Export credit agencies with a greater degree of autonomy and financial responsibility are forced to value the debt at realistic prices that reflect past and likely future servicing ability. It is important that officials not be required by accounting and provisioning regulations to make decisions that are economically incorrect. Third, if the debtor country's economy becomes more efficient (for example, as a result of privatization or increased social spending), future debt service payments are likely to increase. Fourth, creditor governments can be assured that official debt conversions will be used for high-priority government spending (for instance, health, education, and nutrition); this guarantee justifies additional debt reduction (which is also in the interests of the debtor).

Finally, the indirect benefits include enhanced trade, export, and investment as indebted economies recover. Creditor governments, for their part, can use

official debt conversions as a way of funding international public goods, such as environmental protection.

Technical Issues in Official Debt Conversions

Several technical issues need to be dealt with to enhance the implementation of swaps. Some of the most relevant are outlined here. Most of these activities refer to debt-equity swaps because debt-for-development swaps are technically simpler.

Transparency and competitiveness. Five "prices" determine the net gains and losses to debtors and creditors: the discounted price of the debt on secondary markets; the redemption price (that is, the proportion of face value that the debtor agrees to convert to local currency); the transaction fees and taxes; the price (in local currency) of the asset to be acquired; and any sweeteners that are offered to encourage the investment. Experience shows that it is essential that these prices be transparent and equitably applied to all official creditors to avoid perceptions of inequities among creditors and to avoid excessive subsidies to foreign investors.

Legal and technical problems. Official debt claims—in contrast to commercial bank claims—vary widely. This is not a problem if the claims are to be canceled (in development swaps), but it is if the claims are to be converted to equity. (Some creditor governments see this as a major obstacle, while others seem to overcome these problems with ease.) Claims are structured differently for each source of credit within and across a wide range of creditor countries; they are not particularly easy to assign or transfer to third parties. Many clauses have implications for conversion that are not yet fully understood. To avoid complexity, it is possible in many instances to novate existing claims or to exchange them for promissory notes with standard features that reflect the maturities and coupon obligations of the existing bilateral claim. If the volume of official debt sales in secondary markets and bilateral debt conversions grows, the Paris Club could be instrumental in standardizing such exchange instruments.

A particular complication arises in the case of insured export credit agency claims that usually provide less than 100 percent indemnity and leave the policyholder with 5 to 25 percent of the claim but with full rights and obligations to protect in the rescheduling (or conversion). In some instances the policyholder remains the titular owner of the full claim and is responsible for obtaining full recovery, even when the export credit agency has paid out the indemnity. These "tail" claims present major legal and technical problems for export credit agencies that are considering debt reduction or conversion. In some instances the "tails" have had to be bought out at face value and in other instances at negotiated discounts (which are generally above the secondary market price). Although there is no easy answer to this problem, creditor governments could

provide tax or other incentives to "tail-holders" to sell their claims at discounts to export credit agencies or to intermediaries that trade in debt.

Ownership of converted claims. Some governments of Organization for Economic Cooperation and Development (OECD) countries that have undertaken large-scale divestitures of public assets are reluctant to become shareholders of corporations in debtor economies. Debtor governments are also not particularly anxious to have them as shareholders. Unlike banks or other private creditors, governments and most export credit agencies that are government departments cannot directly undertake official debt-equity swap operations on their own. They are therefore left to (a) sell their claims directly—for example, in the secondary market—and withdraw from further involvement; (b) sell their claims on a negotiated basis to state-owned entities (which will pursue swap opportunities); or (c) transfer official claims from the primary source to the equity investment promotion arms of governments that specialize in such investments and already have large portfolios in indebted developing countries.

The last option is perhaps the easiest and most practical. Either the investment agency acts on its own account (after the debt is transferred to it from the export credit agency at an agreed-on transfer price), or the investment agency can act as a trustee in managing the government's foreign asset portfolio through equity conversions.

Financial intermediaries. An important operational issue is whether export credit agencies should rely on specialized intermediaries or should develop their own in-house capabilities. There are concerns that the use of intermediaries may involve conflict of interest issues because the intermediaries are such large players on these markets. There are also concerns about the advisory fees and transaction costs levied by financial institutions. The question is whether there is a case for the interim development of a specialized submarket in trading official claims (which need to become increasingly standard and liquid in their structural characteristics). Specialized market-making, bid-offer pricing, settlement services, and transactions knowledge would certainly be required to make the market work. It could later be integrated into the private secondary market when private and official debt claims become almost indistinguishable in their trading features.

IV. CONCLUSIONS

Although debt-for-equity transactions have absorbed the vast bulk of swaps involving commercial debt (a pattern that is likely to continue), it is important that conversions of official debt direct a greater proportion of funding to poverty alleviation, social sector development, and conservation. To help alleviate debt buildup, the additional debt reduction granted for this purpose should not be funded out of existing aid budgets, so as to ensure an additional foreign exchange saving for the debtor economy.

Debt-development conversions seem well suited for funding social services. Additional local currency resources could be accommodated relatively easily within existing programs or through replication and expansion of ongoing activities. It may be desirable to develop specific proposals to attract and accommodate additional resources on the basis of programs that have been developed but still need resources, such as the National Programs of Action adopted at the 1990 World Summit for Children or the Polish Environmental Fund.

In all these cases debtor governments need to take the lead, but they may require technical assistance from international institutions in program preparation, selection of an appropriate debt conversion mechanism, and bilateral or multilateral negotiations with creditor governments (through a special support group or the Paris Club). The cooperation of creditor governments (some of which may wish to take the lead) is, of course, crucial. For official debt-equity conversions, international institutions can play an important role by supporting the development of well-structured market mechanisms in creditor and debtor countries. Collaboration and consultation with private financial intermediaries will be important.

The potential for debt-equity swaps may be very great. If such swaps are linked to the privatization of state enterprises, they have the advantage of being broadly noninflationary. These conversions are particularly well suited for countries with a high level of official bilateral indebtedness and a large number of public enterprises in the process of privatization. They are suitable for funding private sector infrastructure and can benefit stabilization and structural adjustment programs by helping low- and middle-income countries attract significant foreign direct investment. Debt-equity swaps imply larger commercial returns for export credit agencies. Indeed, they may end up being far larger in volume than debt-for-development conversions. It would, however, be unfortunate if such swaps were to preempt the valuable complementary role that debt-development conversions can play.

REFERENCES

- Aravena, José. 1991. *Debt Reduction Schemes, Theoretical and Empirical Issues for Chile*. Louvain, Belgium: Faculté des Sciences Economiques, Université Catholique de Louvain.
- Blackwell, Michael, and Simon Nocera. 1988. "Debt-Equity Swaps." IMF Working Paper. International Monetary Fund. Washington, D.C.
- Bouzas, Roberto, and Ricardo Ffrench-Davis, eds. 1990. *Conversión de Deuda Externa y Financiación del Desarrollo en América Latina*. Buenos Aires: Grupo Editor Latinoamericano.
- Claessens, Stijn, Ishac Diwan, Kenneth A. Froot, and Paul R. Krugman. 1991. *Market-Based Debt Reduction for Developing Countries: Principles and Prospects*." World Bank Policy and Research Report 16. Washington, D.C.
- Corden, Max, and Michael P. Dooley. 1989. "Issues in Debt Strategy." In Jacob A. Frenkel, Michael P. Dooley, and Peter Wickham, eds., *Analytical Issues in Debt*. Washington, D.C.: International Monetary Fund.

- ECLAC (Economic Commission for Latin America and the Caribbean). 1991. "Preliminary Overview of the Latin American and Caribbean Economy." Dic. 1991. Santiago.
- Frenkel, Jacob A., Michael P. Dooley, and Peter Wickham, eds. 1989. *Analytical Issues in Debt*. Washington, D.C.: International Monetary Fund.
- Larrain B., Fernando, and Andres Velasco. 1990. "Can Swaps Solve the Debt Crisis? Lessons from the Chilean Experience." *Princeton Studies in International Finance* 69. Princeton University, Department of Economics, Princeton, N.J.
- Latin Finance*. 1991. October. Miami, Fla.: Euromoney.
- Lagos, Ricardo. 1989. "Debt Relief through Debt Conversion: A Critical Analysis of the Chilean Programme." Master's thesis. University of Sussex, Institute of Development Studies, Brighton, U.K.
- Mistry, Percy. 1992. *African Debt Revisited: Procrastination or Progress?* The Hague: Foundation on Debt and Development.
- Mistry, Percy, and Stephany Griffith-Jones. 1991. "Interim Report on Conversion of Official Bilateral Debt." United Nations Conference on Trade and Development, Geneva.
- Mortimore, Michael. 1991. "Debt-Equity Swaps." *ECLAC Review* (September). Santiago.
- UNCTC (United Nations Centre on Transnational Corporations). 1991. *Transnational Banks and Debt-Equity Conversions. Report to the Secretary General*. EIC/10/1991/1. New York.
- Williamson, Mary. 1990. *Chile's Debt Conversion Programme*. ECLAC/CET. Santiago.
- World Bank. 1989. *World Debt Tables 1989-90*. Vol. 1. Washington, D.C.
- . 1990. *World Debt Tables 1990-91*. Vol. 1. Washington, D.C.
- . 1991. *World Debt Tables 1991-92*. Vol. 1. Washington, D.C.

COMMENT ON "CONVERSION OF OFFICIAL BILATERAL DEBT:
THE OPPORTUNITIES AND THE ISSUES," BY STEPHANY GRIFFITH-JONES

Jonathan Eaton

The paper by Griffith-Jones proposes swaps of official debt for equity and for "good policy" (where "good policy" could involve, for example, the environment, UNICEF, or development). In my comments I briefly review the possible benefits of the swaps that have occurred so far, which have chiefly involved swapping the private debt of severely indebted *middle-income* countries. I then look at whether the factors that contributed to the success of these swaps are present in the proposed conversions of official debt, which is largely owed by severely indebted *low-income* countries. Using a simple model, I examine some problems with reducing debt in exchange for good policy on the part of the debtor and conclude that for several reasons, debt relief is an inappropriate means of influencing policy in debtor countries.

I. HAVE SWAPS BEEN SUCCESSFUL?

Although they have been criticized for a number of reasons, swaps of debt for equity involving private creditors have had some desirable effects. First, swaps gave countries an opportunity to replace foreign public debt with domestic public debt. This exchange removed issues of international diplomacy and international sanctions from the question of repayment. Second, the swaps facilitated a switch from foreign portfolio to foreign equity investment. Consequently, the risks associated with domestic shocks and international market conditions are now shared with foreign investors. Third, to the extent that swaps have reduced government debt, they have lowered anticipated future taxes, making domestic investment more attractive.

Two questions are relevant here: What specific features of these swaps accounted for their success? Are these features present in debt-distressed low-income countries?

Four features helped the conversions of debt owed to private creditors succeed: (a) the debtor countries involved (largely severely indebted middle-

Jonathan Eaton is professor of economics at the Institute for Economic Development, Boston University, Boston, Mass.

income countries) had attractive private investment opportunities—in fact, in some cases direct foreign investment rose dramatically after swaps began; (b) the governments involved owned valuable national enterprises whose privatization provided a means of financing conversions; (c) a considerable amount of flight capital was held abroad; and (d) these countries were making large net transfers to private creditors, while little new lending was forthcoming. The reduction in outstanding debt thus probably reduced subsequent net transfers to creditors.

There is no strong evidence that any of these four characteristics are present in the official debts of the severely indebted low-income countries (with the possible exception of Nigeria). Table 1 provides evidence on characteristic (d). It shows the cumulative (undiscounted) net transfer from official creditors to individual countries in this group between 1970 and 1990, along with the number of years in which net transfers were made to creditors during those twenty-one years. Note that the average country in this group made a net transfer to official creditors in only one year, while the modal country *never* made a net resource transfer. In an overwhelming majority of years, official creditors lent these countries more than their debt service obligations for that year.

Table 1. *Official Flows to Severely Indebted Low-Income Countries, 1970–90*

<i>Country</i>	<i>Cumulative flow (billions of U.S. dollars)</i>	<i>Years of net repayment</i>
Burundi	1.55	1.00
Egypt	31.80	5.00
Equatorial Guinea	0.32	2.00
Ethiopia	4.22	0.00
Ghana	3.32	2.00
Guinea	1.96	3.00
Guinea-Bissau	0.70	0.00
Guyana	1.17	0.00
Honduras	2.60	1.00
Kenya	4.29	0.00
Liberia	0.96	3.00
Madagascar	3.40	0.00
Malawi	1.98	0.00
Mauritania	2.53	0.00
Mozambique	2.35	0.00
Myanmar	5.59	1.00
Niger	2.00	0.00
Nigeria	(1.02)	7.00
São Tomé and Príncipe	0.22	0.00
Sierra Leone	0.67	0.00
Somalia	3.25	0.00
Sudan	9.95	0.00
Tanzania	6.04	0.00
Uganda	2.95	0.00
Zaire	4.24	3.00
Zambia	4.13	0.00
Total or average	110.98	1.08

Note: Negative numbers are in parentheses. Numbers may not sum to total because of rounding.

Source: World Bank, *World Debt Tables, 1990–91 Edition*.

What matters for both the creditor's and the debtor's current and future welfare is not the stock of debt but the net resources transferred between the two. Debt reduction affects these transfers only if official lenders do not reduce new lending by an offsetting amount. If official creditors make their annual lending decisions with a targeted net transfer in mind, the outstanding stock of debt will not affect the amount of the net transfer. Debtors would be better served by a promise of increased net transfers than by debt reduction per se.

II. WHAT ARE CREDITORS BUYING WHEN THEY SWAP DEBT FOR GOOD POLICY?

Reduction of debt in exchange for increased domestic expenditure on the environment, development, or some other meritorious policy may sound like a worthy undertaking, but there are at least two causes for concern.

The first applies to the issue of conditionality in general: why should official creditors have to "buy" good policy from debtors? One answer is that creditors have an interest in the policy that the debtor country lacks. This may be the case, for example, with policies that affect the global environment, such as the protection of rain forests. Another answer is that debtor governments may not act in the interests of their own nationals, and creditors may want to intercede on their behalf, as with human rights issues. A third answer is that conditionality may provide a well-intentioned government the means to make a commitment to an optimal, but time-inconsistent, policy. For example, the government may decide on low inflation or low taxes on investment in order to encourage investment. But once investments have been made, the government may be tempted to raise inflation or taxes. Anticipating these policies, capital flees abroad. By making future loans conditional on low inflation or taxes, the lender gives the government the incentive to follow through on its promises, which may then be more credible. As a consequence, investment may become more attractive.

The second cause for concern is that even if official creditors do in fact have reason to influence the policies of debtor governments, offering debt reduction as a "carrot" to reward good policy may be a very bad way to do it. For one thing, the debtor may be unsure about how much of a concession is being offered. As discussed above, what matter for the debtor are future net transfers, and what debt reductions imply for these is uncertain. If the government thinks that lenders will offset debt relief by lending less in the future, it may find doing what the creditor asks not worth the trouble. Furthermore, to the extent there is any carrot at all, debt relief lets the debtor eat it before arriving at the desired destination. If debt swaps *do* lower anticipated future transfers, they do so all at once—when the debt is reduced—whereas the sought-after policy is probably long term. What does the creditor do if the debtor does not follow through on the quid pro quo of the swap?

Indeed "sticks" may be better incentives than "carrots." The problem with offering to reward good policy rather than threatening to punish bad policy is that the first approach creates an incentive for a debtor to act as if it wants to

pursue bad policies in order to get a reward for not doing so. Knowing that a debt-for-nature swap may be in the cards, for example, a debtor may degrade the environment to try to show that it must be "bought off" not to pollute. The result may be even *more* environmental degradation than would occur in the absence of any program!

The following example is from Eaton and Engers (1992).¹ Say that the debtor can decide in each of two periods whether or not to pollute. Its pollution inflicts an amount of harm c on the creditor. The debtor may be a "true" polluter, meaning that it benefits from polluting by some amount b , or it may not benefit at all. Say that c exceeds b , so that in either case pollution is globally inefficient because world costs exceed world benefits. The creditor, however, does not know whether the debtor benefits from polluting or not. Initially, it assesses the probability that the debtor is a "true" polluter at p .

The problem is that the creditor's willingness to offer relief in exchange for not polluting depends on its assessment that the debtor is a true polluter. The "false" polluter has an incentive to try to make the victim think that it is a true one so that it can get relief in exchange for staying clean.

Say that each party discounts the future by just a small amount. There are then three equilibrium outcomes. Which one emerges depends on the relationship among c (the cost of pollution to the creditor), b (the benefit of pollution to the true polluter), and p (the creditor's initial belief about the probability that the debtor is a true polluter).

If pc exceeds b , in each period the creditor will offer relief in an amount slightly more than b as a reward for not polluting. The creditor never learns whether it needed to offer this much to avoid pollution. The efficient outcome of no pollution is achieved, but the false polluter free-rides on the creditor's ignorance about its preferences, getting rewarded for not doing something it didn't really want to do anyway.

If pc is less than b but more than b/c , in period 1 the creditor offers the debtor relief in an amount slightly less than b in return for not polluting. This is rejected by the true polluter but is accepted by the false one. Relief thus coaxes the false polluter to reveal itself, but the polluter cannot extract compensation again, since the creditor learns it does not have to offer the reward. The creditor also realizes that if the debtor refuses the offer in period 1, it must be offered at least b in period 2. So if the first offer is turned down, in period 2 the creditor offers the higher amount. The outcome is now one in which the true polluter actually pollutes in period 1, while the false polluter is bought off not to. The debtor must be paid almost b because in refusing the offer and polluting—thus pretending to be a true polluter—it could get b in the next period. There is no pollution in period 2, when true polluters are paid b not to pollute and the false polluters (now revealed) are paid nothing. Here the false polluter can free-ride in only one

1. Eaton, Jonathan, and Maxim Engers. 1992. "Threats and Promises." Boston University, Department of Economics, Boston, Mass.

period. The creditor puts up with first-period pollution by true polluters because the cost of compensating free riders over two periods exceeds the benefit of stopping all pollution. The outcome is inefficient, since the cost of pollution to the creditor exceeds the benefit even to a true polluter. But less pollution occurs than if no relief were offered, in which case true polluters would pollute in both periods.

Finally, if pc is less than b/c , the best the creditor can do is to offer no relief in period 1. True polluters and some proportion of false polluters then pollute in period 1. (If all false polluters did not pollute, the creditor, realizing this, would offer those that did pollute an amount just above b not to pollute in period 2, creating an incentive for false pollution as well. Enough false polluters must pollute in period 1 to remove the creditor's incentive to offer relief in period 2.) In this outcome, not only is there pollution, but there is *more* of it than if relief in exchange for nonpollution were ruled out altogether. The creditor's ability to provide relief to debtors that it can identify as true polluters later on creates an incentive for false polluters to pollute. They do so to send a false signal that they are true polluters.

This example illustrates the importance of being clear about the motive for debt relief. Here, the creditor provides relief that it would not provide otherwise in order to change the debtor's behavior. Inefficiency arises from the debtor's attempt to increase the amount of relief.

Alternatively, the creditor might wish to provide relief simply to reduce future net transfers from the debtor. In this case it should begin by offering the amount of relief (or aid) that it would want to offer to the best-behaved debtor and should threaten to reduce future aid by an amount b if good policy is not forthcoming. Using such a "stick" to punish undesired behavior avoids the potential for opportunistic behavior by debtors who might otherwise embark on atrocious policies in order to be rewarded for not continuing them.

III. CONCLUSION

Humanitarian as well as developmental concerns may justify increased aid for severely indebted low-income countries, but it is doubtful that debt relief, by itself, will contribute to their growth or welfare unless it is accompanied by a commitment on the part of official lenders not to offset debt relief with less new lending. Moreover, it is important to distinguish between reducing future net transfers to creditors and encouraging better policy as objectives. The two goals are best met by different approaches. Making relief contingent on good policy can create an incentive for debtor governments to initiate bad policies in order to be rewarded for discontinuing them.

FLOOR DISCUSSION OF THE COLLINS AND GRIFFITH-JONES PAPERS

A participant from the Overseas Development Administration in London said that Susan Collins's model was too narrow: to compare flows between developing countries and the economies in transition simply on the basis of gross national product (GNP) and population was slightly misleading because the sets of countries were different. Their GNPs were similar, but in many other respects the social structure in the economies in transition was so much more advanced that it was a mistake to categorize them as competing with developing countries for the same aid funds. A number of European countries had stated clearly that they did not intend to divert funds from traditional programs, and the Organization for Economic Cooperation and Development (OECD) was now hotly debating how to compare lists of countries eligible for aid. For a few years the economies in transition would need aid—preferably not diverted from developing countries, but additional resources. What the participant found interesting was how quickly these economies would be able to move toward concessional funds. There the prospects were somewhat less depressing, since he thought the economies in transition could become creditworthy fairly rapidly. A participant from the Congo cautioned that establishing a separate "pie" might minimize the issue of diverting aid from developing countries but would not eliminate it. If the increase in the aid pie were not greater than the separate pie, the amounts available to developing countries would obviously decrease.

Griffith-Jones thought there was a good chance that Poland, Hungary, and Czechoslovakia would be members of the European Community by 2000: at least, they want to be. Wouldn't access to the European market attract much more foreign direct investment to the economies in transition? Griffith-Jones thought it might have been simpler if Collins had started with empirical data on concessional flows rather than trying to estimate flows econometrically, which was not so relevant because the economies in transition had European links. Moreover, in determining international interest rates, Collins had focused on the savings-investment equation. Shouldn't she also have considered the mixture of tight monetary policy and expansionary fiscal policy that the Germans were now pursuing and that the United States had pursued ten years earlier?

Collins responded that her aggregate grouping was merely a first-pass look at the situation and that she could not agree more about the importance of looking at individual countries. In developing this "aid diversion" equation, which was

This session was chaired by S. Ibi Ajayi, professor of economics at the University of Ibadan, Nigeria.

© 1993 The International Bank for Reconstruction and Development/THE WORLD BANK

not meant to be used for forecasting, she had looked at historical aggregate numbers that, she agreed, concealed much of what went on when individual countries were considered. It was important to look at countries in more detail, and she was not forecasting what would happen in the long run in terms of the diversion of funds to the economies in transition. She fully agreed that internal developments would determine whether those countries attracted international flows or generated domestic savings. Developing countries as a whole accounted for only a small part of international flows, and it was true that donor countries had announced that they would avoid those diversions.

But—on average—developing countries had cause for concern, Collins reiterated. For developing countries as a group there appeared to have been a diversion of funds. If there were some way to put a fence around the assistance that went to developing countries, so much the better, said Collins, although in practice that would of course be difficult. Official development assistance was a small part of the pie in some respects, she said, but for some countries it was not at all small, and it was determined differently from private capital flows. Equally important, domestic savings were likely to be a much more important source of finance than either official development assistance or private capital flows.

Another cause for concern, said Collins, was the relationship between the economies in transition and the developing countries. Historically, the economies in transition had given developing countries substantial amounts of development assistance, and that had been eliminated. Whether the two groups would now become partners or competitors was a hot issue. Probably they would be both, and she hoped that industrial countries would react in such a way as to help to minimize that competition and would work toward more of a partnership. S. Ibi Ajayi (chair) interposed a comment on this subject, somewhat as devil's advocate. He could not cite numbers, but surely there would be competition between the two groups, since they were all drawing from the same pool. So it would be realistic—in the short term at least—to expect some competition. In the long run, the outcome was less clear.

A participant from the United Kingdom asked Griffith-Jones how the export agencies, which hold most Paris Club debt, had reacted to these debt-conversion schemes. The debt was usually held on their books at totally artificial prices, he said, and they had to be bought out. Had they been able to arrive at sensible bases for pricing when they sold to the private sector? The United Kingdom, he said, having privatized half the public sector, did not want to be left holding large stocks of, say, overpriced public sector enterprises in Nigeria, either. If one was in the debt game for good causes, he continued, an export credit agency was, on the whole, going to look to some other part of its parent government to buy out its claims. That was going to come primarily from aid funds, and he suspected that it was mainly going to be nonadditional. Was there not a danger, then, of this being a zero-sum game for the debtor country?

As for whether the United Kingdom would want to hold assets in Africa, Griffith-Jones said that it did not have to hold the assets for any length of time; it

could just sell them in turn, as the Belgians had. Mexico, for example, had swapped debt for a domestic steel company that was losing money. The purchaser turned the company around and sold it at a profit. So maybe there is an opportunity for the U.K. Treasury to make money, if that is what it is trying to do. In response to the question of whether the conversion of official debt, which tends to be owed by low-income countries, was relevant for poor countries, she said that in talking about investment opportunities, one must be careful not to compare Chile now with Ecuador now. One should compare them in 1984 and 1985, when they were coming out of a financial crisis with a massive debt overhang and prospects did not look good. In Chile and Mexico a reduction of the debt overhang and improved policies have radically changed the climate. Such a result would be more difficult in poorer countries, but not impossible. The price of the debt there was, of course, lower than it had been in Chile and Mexico. But even the poor countries could kick-start foreign direct investment if they simultaneously improved their economic policies and offered investors larger subsidies. Why not try it in a country such as Bolivia, which has improved its economic policy but so far has not been able to attract investment flows?

Griffith-Jones agreed that poor countries do not necessarily have valuable state enterprises to sell off. Ecuador, for example, did not have a particularly large state sector. Other countries might not want to privatize or might find it difficult because the price of their commercial debt had gone up. One had to examine options country by country. To Jonathan Eaton's (discussant) question about capital flight, she responded that Nicaragua had hyperinflation and a socialist government, but she was sure that some bank accounts (owned by Nicaraguans) in Miami could be encouraged to return. A large amount of capital had fled the Philippines, and even in the poorer African countries she suspected that there had been considerable capital flight. It might have happened on a smaller scale in those countries, but the return of even a small amount of capital would probably help them. Later in the discussion Ajayi commented that there were many different estimates of the amount of such flight capital and that the problem for developing countries in creating and maintaining a macro-economic environment good enough to attract capital back was that they had to keep running harder all the time just to stay in place.

As for how these debt-conversion deals would affect net transfers—new money and new aid flows—Griffith-Jones saw two key questions. First, would any debt reduction for development come from an existing budget, or would it be additional? Second, if debt conversions were very large, would it inhibit new flows from the export credit agencies?

Some donor governments and some nongovernmental organizations liked debt-development swaps, she said. Whether it was rational or not, they were willing to commit additional money where they could see a specific link with spending on children, with trees being planted, and so on—and they liked the connection with debt reduction. One had to be very careful, however, that aid was not siphoned off for debt reduction. Taking it out of existing aid budgets

would only be playing games and would mean the transfer of aid from not-so-heavily indebted countries to very heavily indebted countries, which might or might not be desirable.

The British participant had asked Griffith-Jones how export credit agencies had reacted to these schemes. Their reactions varied, she said. On its seven hundredth anniversary, Switzerland created a Sw F 700 million facility for debt reduction, which it preferred to link to good works. That was a clear window of opportunity for additional funds. Belgium, by contrast, had an autonomous export credit agency that operated on purely commercial grounds, marking down the value of the debt to what was realistic.

The export credit agencies' reaction depended, too, on their level of loan-loss provisioning. If existing provisions were adequate, they might be willing to go further because they did not have to take a loss—the money didn't have to go through the treasury. In countries where the treatment of debt was unrealistic, the operation was more difficult. One had to look at each operation to see whether there was additionality and what the effect would be on both debtor and creditor countries. It was important to look at these deals in detail and not to let enthusiasm overcome reason.

ROUNDTABLE DISCUSSION

The Health of the Public: A Public Responsibility?

The panelists for this roundtable discussion were Julio Frenk, the founding director-general of the National Institute of Public Health of Mexico; Eyitayo Lambo, a health economist with the World Health Organization Regional Office for Africa, Brazzaville, Congo; Uwe Reinhardt, James Madison Professor of Political Economy at Princeton University; and W. Henry Mosley, chairman of the Department of Population Dynamics and director of the Institute for International Programs at the School of Hygiene and Public Health of the Johns Hopkins University. The moderator was Lawrence Summers, vice president, Development Economics, and chief economist, the World Bank.

Eyitayo Lambo had been asked to give his thoughts on the role of the public and private sectors in health care. Lambo said that "the public" in the subtitle could be defined in two ways: as the government or as people in general. Either way, the answer was a qualified yes, that it was a public responsibility. There was absolutely no doubt, however, that both the private and the public sector had roles to play. He identified several roles for the public sector, including:

- Developing explicit, community-oriented health policies, which would be used as the basis for developing prospective and medium-term health plans and annual budgets. Many developing countries—especially in Sub-Saharan Africa—do not have such policies.
- Leading and managing the implementation of health plans and ensuring that finance is available for health services for the most vulnerable citizens, especially women and children, and for health services that would not otherwise be provided
- Ensuring acceptable and accessible health care to the population that is not covered by private health services, especially in the poor rural areas where about 70 percent of all Africans live
- Ensuring a standard of care that is compatible with the resources available and provides a balanced system of preventive, promotive, curative, and rehabilitative care
- Gathering and distributing information (including health systems research) for public planning and regulation
- Overseeing the activities of agents in the health sector and fostering partnerships among all concerned with health and health care

- Promoting social and physical environments that are conducive to health and ensuring that citizens are empowered to make informed decisions in matters that affect their health
- Coordinating donor contributions in the health sector, supporting community health initiatives, and rationalizing the roles of different players in the health care system.

As for the private sector, Lambo mentioned several roles, including filling the resource gap for health development by providing coverage and access for essentially private groups, improving efficiency and the quality of care in the health sector by fostering competition, and supplying some services the public sector cannot afford—for example, radiotherapy, computerized tomography (CT) scans, and other high-technology services. Another function, he said, was to offer health consumers a choice to supplement and to enlarge the field through innovation, which the flexibility of the private sector allows.

Julio Frenk felt that a dichotomy should not be created between public and private roles and that the debate had been oversimplified, for three main reasons. First, rapid demographic and epidemiological changes in many middle- and low-income countries were making the distribution of world health problems more heterogeneous, while at the same time the distribution of health conditions and disease within countries was becoming more inequitable. Second, people spoke of private enterprise as a new player in health care when it already dominated certain sectors (including equipment, pharmaceuticals, and hospital construction). And many private sector participants were totally unregulated, operating in extreme *laissez-faire* conditions that approached chaos. Third, private sector institutions themselves were highly heterogeneous. Some nonprofit (often religiously affiliated) organizations functioned more like their public counterparts than like private organizations. Private facilities ranged from small shops run by individual entrepreneurs all the way to multinational chains. Finally, health services themselves included everything from sanitation to cosmetic surgery. People talked about public and private health care as two polar extremes, when the real world contained many combinations—and it was a creative mix that held the most promise.

Lawrence Summers asked W. Henry Mosley how he would set priorities. Mosley responded that the government should be in the business of producing health, welfare, income growth, and all the other aspects of development. The private sector was engaged in the production of goods and services for profit, and there was nothing wrong with that, but the production of profits might or might not produce health. Governments themselves could also make investments that did not produce health, but they should always try to make the choices that would.

Mosley used the example of tobacco and smoking to illustrate the kinds of choices governments have to make. First, it has to understand which health problems exist in the community and to make the right choices about how to

prevent or ameliorate them. Tobacco-related conditions represent the most serious health problem in the world today, accounting, for example, for almost 30 percent of all deaths in the United States. In many countries governments are engaged in promoting the production and consumption of tobacco. That is the wrong choice. The private sector is also engaged in these activities—also a wrong choice because it is a disease-producing endeavor.

Second, governments must be clearly aware of the range of interventions available and must decide how cost-effective each option is. Options available for treating tobacco-related diseases and conditions include surgery for lung cancer, open-heart surgery, coronary bypasses, and lung transplants. Clearly, supporting lung transplants is not a sensible option for the government. Reducing and ultimately ending smoking, especially through taxes (which would also increase government revenues), might be more cost-effective. Other appropriate choices include education, regulation, and such agricultural policies as encouraging the production of alternative crops.

Third, governments must have some way to measure health if they are going to produce it. Mosley used the measure of healthy life years: how many healthy life years could be saved or produced by preventing a disease (for example, with measles vaccine or by discouraging smoking) or by rehabilitation (for example, restoring vision through cataract surgery)? One could discount an investment today by how many healthy life years were gained in the future and thus weigh investments.

Finally, in addition to knowing the burden of a disease, the interventions available, and how much health a given intervention would produce, it is important to know the cost of the intervention (which, in the case of smoking, might include the cost of producing behavioral change or of protecting the environment from tobacco).

Once we understand the problems and choices, it is much clearer where the private and public sectors should become involved. Some interventions—such as clinics to help people stop smoking and nicotine patches—are more appropriate to the private sector. What can governments do? Mosley listed five kinds of interventions: information, regulation, taxes and prices (to influence behavior), the direct provision of services (such as immunization), and research (to expand the array of choices). Most crucial, he said, was information, including information for policymakers and health care providers. Mosley said that he had been engaged for almost fifteen years in developing and promoting oral rehydration therapy for diarrheal diseases but still had to stop misinformed doctors from prescribing expensive antibiotics or telling mothers to stop breastfeeding. Health care providers in poorer countries are more interested in selling products than in producing health. The pharmaceutical industry is an essential part of a health system, but there must be appropriate incentives (and disincentives) to ensure that the public health and welfare are protected.

Summers asked the next speaker, Uwe Reinhardt, about his involvement in the debate about health care in many countries, including the United States, and

what lessons he drew from his experiences in designing health care policies. Before answering, Reinhardt took issue with Mosley's remark that the government's function was to produce health and the private sector's function was to produce profits. Any economic sector had two roles: to provide services to some people and to provide income to others; some of that income was profits and some was wages, and any economist would tell you that wages include a huge profit component.

Reinhardt suggested a 3 x 3 matrix to clarify the issues. In the rows he listed who owns and operates the real resources—the clinics, hospitals, and so on. That could be the government (as in the United Kingdom), it could be a private but nonprofit organization, or it could be a private, for profit (investor-owned) enterprise. In the columns he suggested listing who controls the money transfer from households to doctors, nurses, and health economists. That transfer could be accomplished by the government (as in the United Kingdom and in Canada, where the government taxes households and then reimburses providers); through business-mandated benefits (where businesses collect taxes from workers' paychecks and funnel them to the doctors); or through private means (unregulated, as in the United States, or regulated, as in Germany). He noted parenthetically that the German government owned few health care delivery resources and offered no insurance at all but fluttered over private insurers and health care providers, dangling a huge Damocles sword and letting it drop whenever they failed to dance to the government's tune.

The point of this matrix, Reinhardt said, was that no country is in only one cell in this grid; they are all over the map. The United Kingdom, Sweden, and Germany have private systems. Canada does not, but then again it does—it is called the United States.

Reinhardt's main conclusion was that every country, industrial or not, will always run a two-tier health care system. That is the ideal, the best you can achieve. The large bottom tier of the system would serve 70 to 80 percent of the people. That tier, where health care was viewed as a public good, would be publicly financed by people paying into an insurance pool according to their ability to pay. A private system of health care would serve the top 20 percent or so of the population, with private hospital rooms and gourmet food if they wished. In that tier, people would pay through private insurance or out of pocket, there would be no public health provision, and if people wanted to buy unnecessary operations that might kill them, that would be their problem and their choice. Reinhardt thought that the one-tier universal health system advocated by left-wing or centrist thinkers would never come about. He felt that failure to recognize this made the best the enemy of the good, which is the tragedy of health policy.

What would Reinhardt do, practically? In the Philippines, for example, he would sell to the private sector the hospitals in Manila that do transplants, and he would use the money from the high overhead of often useless tertiary care given only to the elite for clinics that would provide secondary care to attack the

diseases of the poor. The rich would be encouraged to get their problems treated abroad.

Summers asked the other panelists if they thought a two-tier system was inevitable and invited their reaction to Reinhardt's implicit question: how do you keep public money out of, or reduce its role in, the upper tier?

Frenk said that it was not an ideal to aspire to, in reforming the health care system; it was the situation we have now. Many countries have a two- or three-tier system. The ministry of health or a similar organization, together with some charity, takes care of the poor. In Latin America a second tier of social insurance or a social security system takes care of the part of the population that works in the formal sector of the economy. Finally, the private sector takes care of the elite.

One problem with this kind of system, said Frenk, is that subsidies still go to the top tier. The top tier is rarely completely private or without some transfer of funds from the government—even if such transfers are in the form of tax deductions or public subsidies for research and development. Another problem is that even elite consumers have imperfect information and require protection (if only in the form of quality assurance) from dangerous services. Many countries do not have the required set of institutions to ensure adequate standards of care. Their medical professions are not organized and provide little professional self-regulation. The court systems are not well enough developed to control provider behavior through malpractice suits. There are no effective means for accrediting hospitals and other facilities.

Finally, said Frenk, there is an ideological question. Poor health and disease are unequally distributed along social dimensions. Is it the role of the health system to correct that unequal distribution? If not, then we can live with a two- or three-tier system; otherwise, we must seek alternative interventions.

Summers asked Mosley, who had advocated policies designed to get the maximum bang for the health buck, what he thought about the goal of operational efficiency, as opposed to responding to what consumers want. Summers said that he understood Reinhardt's view to be that, in the end, consumer sovereignty would reign and a two-tier system would emerge.

Mosley said he was not sure that he could adequately answer such a question, but if you truly followed the premise of making health care decisions on the basis of the cost-effectiveness of the interventions in producing health, the poor have more disease and thus would presumably have more equity. As for consumer sovereignty, that relates to having fully informed consumers and providers because a full range of information about choices is needed, and there are serious information gaps at every level. Few consumers realize that they have a greater risk of dying from smoking than from homicide or a car accident, and they don't know the cost-effectiveness of all the options available for managing heart disease or breast cancer, or most other chronic conditions. But few providers (physicians) do either, so we have market failures.

Summers asked if Lambo supported the idea that we must plan for two tiers, and if he did, what he saw as the role of public money in such a system. Lambo

said that funding the bottom tier would not be a problem if the government diverted what it was now spending to subsidize care for the elite. If it were a question of the government's saving money by getting the elite to pay for their own health care and using that money to provide care for the other group, that would be fine. The problem was getting the government to follow through; it saves the money, but somehow the money is not used for the bottom tier. That was where the problem lay.

Reinhardt agreed. In Manila, he was told, the two hospitals that do transplants absorb two-thirds of the health department's budget and it is unlikely that those transplants are equally shared. The money for the large bottom tier must come from somewhere. There could be less defense spending, but ultimately the "haves" must pay for their own health care and must also be asked at the cashier's window to pay for the bottom tier. The reason the American poor are suffering is that the price of being nice to them is too high. In Canada a coronary bypass can be bought from a surgeon for \$1,200. In New York City, it costs \$10,000. It is not surprising that everyone in Canada is insured—there it is very cheap to be nice—whereas in the United States there are 37 million uninsured individuals. Reinhardt said he always warns countries not to let prices be driven so high by the medical establishment that kindness is priced out of the soul of the "haves," as is happening in the United States. The United States cannot afford to cover the uninsured—women and children—until it has costs under control. Doctors will drive that system up, said Reinhardt; only hard-nosed economists can prevent this folly.

This explains why economists are so popular, said Summers, before asking each panelist for a final comment.

Julio Frenk returned to his criticism of the multitier system. The rich are a tiny minority in developing countries, he said, agreeing that they should be left alone and that their health care should no longer be subsidized. But there is a growing urban middle class that is completely unprotected and often at the mercy of primitive, totally unregulated private health services with no quality assurance or consumer protection. He questioned the notion that chronic diseases are solely the legacy of the rich and that easily preventable infectious diseases are a problem only of the poor. The poor suffer high rates of both chronic and infectious diseases, he argued. Chronic diseases are more prevalent among the urban poor than among the rich, and the population is aging rapidly. Frenk agreed with Reinhardt that it was critical to separate the delivery of services from the financing of services. The government does not have to have a monopoly on delivering services, but it must offer an alternative—the saying in the private system is that the first operation you get is a wallet biopsy, and if it turns out negative you don't go on to the next procedure. Innovations for expanding coverage for the urban middle class—such as prepayment schemes and private insurance that exerts pressure to control costs and ensure quality—can be tested gradually. Meanwhile, we can begin to improve quality in the private sector and find ways to introduce competition into the public sector, make private sector

care more responsive to consumer choice, and pay providers according to their performance. The critical distinction concerns not so much private and public but who controls the system. Consumers currently have little power or protection in either sector; both sectors must be reformed to produce better health outcomes.

Henry Mosley agreed with Frenk, adding that in his experience there was a transfer of health resources from the poor to the rich. The rich pay more, but they also expect more. Because they are paying for open-heart surgery and other advanced techniques, they expect well-trained doctors and the best facilities, which are expensive. The government generally provides these resources, which means, effectively, taking them away from the poor. This is an economic and political reality in the health system that must be dealt with. He agreed that much can be done to make health care more economical; the government should not only encourage providers to offer innovative, low-cost health care but should encourage consumers to demand it.

Eyitayo Lambo urged greater efforts to increase equity in the system so that the well-to-do pay for their own health care and only the people who need help are subsidized. He deplored a proposed national health insurance plan in Nigeria to have everyone insured under the scheme, irrespective of income, pay the same amount, which would be equally matched by the government or the employer. Such a scheme, if implemented, would mean subsidizing people who did not need assistance, that is, people in the high-income group.

Reinhardt concluded by saying he has just begun working on health policy in developing countries and found it gratifying and refreshing after working in the United States where, as Summers had mentioned, the ratio of talk to action is so frustratingly high. Developing countries, he said, are more sincere about the problem and have more hope that things can get done. The mistake we have made in the United States, he said, is to allow the health system to be configured by the providers, who act in their own professional and financial interests. He hoped there would be more coherent policies in the developing countries and a more rational allocation of resources and subsidies.

Foreign Investment Promotion in the Developing Countries

Foreign investment can play a critical role in boosting economic growth in the developing world. But developing countries should not wait for foreign investment to come to them. They must learn how to promote their countries as good places to make an investment.

This innovative series of videos demonstrates how to organize a campaign to attract foreign direct investment. Specialists in foreign investment explain what investors are looking for when planning an investment, how to find sources of foreign investment, and what new programs countries are using to encourage private investment.

In the first series of videos the basic principles that have been used to attract foreign investment are presented. The second series highlights success stories of export-oriented development that has been done with foreign investment or other foreign-private collaboration. Examples are presented from Mauritius, the Caribbean and Central America, and Asia.

The third series focuses on the special needs of Sub-Saharan Africa. Specialists explain how to get the climate right for investment, and they review newly emerging privatization and debt-equity swap programs that offer a promising role for foreign investment.

Each of the three series can be purchased separately. Videos are available in English or French and in three VHS formats: NTSC, PAL, and SECAM.

Produced by the Foreign Investment Advisory Service of the International Finance Corporation, the Multilateral Investment Guarantee Agency, and the World Bank

See reverse for ordering information.

Ordering Information

Foreign Investment Promotion in the Developing Countries

Series I: Approaches and Techniques in Marketing a Country

Video 1: The Public Relations Industry Approach (35 minutes)

Video 2: The Institutional Approach (58 minutes)

US\$85.00 / Order English NTSC / Stock # 12020

English PAL / Stock #12021

French SECAM / Stock #12022

French PAL / Stock #12023

Series II: Success Stories of Foreign Collaboration in Export-Oriented Investment

Video 1: Mauritius (44 minutes)

Video 2: Caribbean / Central America (42 minutes)

Video 3: Asia (70 minutes)

US\$120.00 / Order English NTSC / Stock #12024

English PAL / Stock #12025

French SECAM / Stock #12026

French PAL / Stock #12027

Series III: Improving the Foreign Investment Climate in Sub-Saharan Africa

Video 1: Getting the Climate Right in Sub-Saharan Africa (21 minutes)

Video 2: Privatization in Sub-Saharan Africa and the Role of Foreign Investment (58 minutes)

Video 3: Debt-Equity Swap Programs in Sub-Saharan Africa and the Role of Foreign Investment (56 minutes)

US\$120.00 / Order English NTSC / Stock #12028

English PAL / Stock #12029

French SECAM / Stock #12030

French PAL / Stock #12031

Complete Set: Series I, II, & III

US\$250.00 / Order English NTSC / Stock #12032

English PAL / Stock #12033

French SECAM / Stock #12034

French PAL / Stock #12035

For a detailed brochure describing these videos write to: The World Bank, Office of the Publisher, Room T-8062, 1818 H Street, NW, Washington, DC 20433, or fax 202-676-0581.

*To order any of these videos
please use the coupon that follows.*

*Extensive Research Library
on 17 Compact Discs*

**Compact International Agricultural Research Library,
Basic Retrospective Set 1962-1986 (CIARL BRS)**

Now agricultural researchers worldwide can have instant access to 190,000 pages of essential research material without leaving their computers. The Consultative Group on International Agricultural Research (CGIAR), an informal association of 40 public and private sector donors, is creating one of the world's largest full-text libraries on CD-ROM. And CIARL BRS comprises the first publication in this major research collection.

CIARL BRS is a set of 17 CD-ROM discs that provides a comprehensive library of agricultural research. At the touch of a few keys, users access

- complete text, images, and catalog records for some 1,350 titles originally published by 20 agricultural research centers worldwide
- key books, serials, and proceedings published between 1962 and 1986, totaling 2,000 documents
- more than 190,000 pages of materials converted to text
- over 50,000 monochrome images, including all graphics and photographs
- 1,900 selected color images — such as maps and field guides

The Master Disc of the set provides indexes to the entire collection — a set of fully searchable catalog records and a handy finding aid for easy subject access to documents. The retrieval system is designed to give users many search options. These options range from simple searches for beginners to advanced search techniques for more experienced users. In mere seconds, a search can find every occurrence of a word or phrase.

Documents in CIARL BRS come in 11 different languages; however, most materials are in English, Spanish, and French. All interface prompts, text messages, user aids, and the reference guide are provided in these three languages.

CIARL BRS requires an IBM-compatible microcomputer with 550K free memory, hard disk, and standard CD-ROM drive. Any monitor — from monochrome to VGA — can display text and monochrome images, while color images can be displayed on EGA or VGA color monitors.

US\$1,950.00 / Order Stock #12179

For more information about CIARL - BRS write to: The World Bank,
Office of the Publisher, Room T-8048, 1818 H Street, NW, Washington, DC 20433,
or fax 202-676-0579.

*To order the CIARL - BRS,
please use the coupon that follows.*

Two Valuable Resources

Index of Publications & Guide to Information Products and Services *January 1993*

The latest edition of this handy reference includes two new features that make it easier than ever to use. A short description is included for each title published after January 1, 1992, and a new section lists, by language, all non-English publications available.

This edition of the **Index of Publications** lists more than 750 World Bank publications in print as of January 1, 1993. Each title, listed alphabetically, includes detailed bibliographic information. Titles published after January 1, 1991, also include an ISBN number, and beginning with this edition, titles published after January 1, 1992, include a short description.

Four separate sections list titles alphabetically by author, by 130 countries or regions, by 88 subject categories, and by eight languages (other than English). The **Index of Publications** also includes ordering information and a list of more than 60 authorized distributors and booksellers of World Bank publications.

*104 pages / ISBN 0-8213-2238-9 /
Complimentary / Order Stock #12238*

Directory of Libraries The World Bank Depository Library Program *January 1993*

The World Bank Depository Library Program, with more than 170 participants, provides free worldwide access to World Bank publications.

This fifth edition of the *Directory of Libraries* provides the names and locations of the World Bank Depository Libraries. It also lists the name of the librarian responsible for the collection and the telephone, telex, or facsimile numbers for the library. A statement of the criteria used for selecting depository libraries and a list of distributors of World Bank publications are also included.

Depository libraries receive all literature made available through the Bank's Office of the Publisher. Participating libraries are required to catalog the publications and make them available to the public for research and reference. The publications are also available to other libraries through interlibrary lending.

*44 pages / ISBN 0-8213-2333-4 /
US\$6.95 / Order Stock #12333*

*To order either of these volumes
please use the coupon that follows.*

CUSTOMERS IN THE UNITED STATES:

Complete this coupon and return to

World Bank Publications
Box 7247-8619
Philadelphia, PA 19170-8619
U.S.A.

To have your order shipped faster, charge by credit card by calling (202) 473-1155 or send this completed order coupon by facsimile by dialing (202) 676-0581.

CUSTOMERS OUTSIDE**THE UNITED STATES:**

Contact your local World Bank Publications distributor for information on prices in local currency and payment terms. (A complete list of distributors follows this coupon.) If no distributor is listed for your country, use this order form and return it to the U.S. address. Orders that are sent to the U.S. address from countries with distributors will be returned to the customer.

Title	Stock Number	Price

Subtotal US\$ _____

Postage and Handling* US\$ _____

Total US\$ _____

* If purchase order is used, actual postage will be charged. If payment is by check or credit card, postage and handling charges are \$3.50 per order. For air mail delivery outside the U.S., include an additional US\$6.00 per item.

CHECK METHOD OF PAYMENT

- Enclosed is my check payable to World Bank Publications.
 Charge my VISA MasterCard American Express

 Credit card account number

 Expiration Date

 Signature

- Bill me. (Institutional customers only. Purchase order must be included.)

PLEASE PRINT CLEARLY

Name _____

Firm _____

Address _____

City _____ State _____ Postal Code _____

Country _____ Telephone _____

Distributors of World Bank Publications

Prices and credit terms vary from country to country. Consult your local distributor before placing an order.

ARGENTINA
Carlos Hirsch, SRL
Galeria Guemes
Florida 165, 4th Floor-Ofc. 453/465
1333 Buenos Aires

**AUSTRALIA, PAPUA NEW
GUINEA,
FIJI, SOLOMON ISLANDS,
VANUATU, AND WESTERN
SAMOA**
D.A. Books & Journals
648 Whitehorse Road
Mitcham 3132
Victoria

AUSTRIA
Gerold and Co.
Graben 31
A-1011 Wien

BANGLADESH
Micro Industries Development
Assistance Society (MIDAS)
House 5, Road 16
Dhanmondi R/Area
Dhaka 1209

Branch offices:
156, Nur Ahmed Sarak
Chittagong 4000

76, K.D.A. Avenue
Kulna 9100

BELGIUM
Jean De Lannoy
Av. du Roi 202
1060 Brussels

CANADA
Le Diffuseur
C.P. 85, 1501B rue Ampère
Boucherville, Québec
J4B 5E6

CHILE
Invertec IGT S.A.
Americo Vespucio Norte 1165
Santiago

CHINA
China Financial & Economic
Publishing House
8, Da Fo Si Dong Jie
Beijing

COLOMBIA
Infoenlace Ltda.
Apartado Aereo 34270
Bogota D.E.

COTE D'IVOIRE
Centre d'Édition et de Diffusion
Africaines (CEDA)
04 B.P. 541
Abidjan 04 Plateau

CYPRUS
Cyprus College Bookstore
6, Diogenes Street, Engomi
P.O. Box 2006
Nicosia

DENMARK
SamfundsLitteratur
Rosenoerns Allé 11
DK-1970 Frederiksberg C

DOMINICAN REPUBLIC
Editora Taller, C. por A.
Restauración e Isabel la Católica 309
Apartado de Correos 2190 Z-1
Santo Domingo

EGYPT, ARAB REPUBLIC OF
Al Ahram
Al Galaa Street
Cairo

The Middle East Observer
41, Sherif Street
Cairo

FINLAND
Akateeminen Kirjakauppa
P.O. Box 128
SF-00101 Helsinki 10

FRANCE
World Bank Publications
66, avenue d'Iéna
75116 Paris

GERMANY
UNO-Verlag
Poppelsdorfer Allee 55
D-5300 Bonn 1

HONG KONG, MACAO
Asia 2000 Ltd.
46-48 Wyndham Street
Winning Centre
2nd Floor
Central Hong Kong

INDIA
Allied Publishers Private Ltd.
751 Mount Road
Madras - 600 002

Branch offices:
15 J.N. Heredia Marg
Ballard Estate
Bombay - 400 038

13/14 Asaf Ali Road
New Delhi - 110 002

17 Chittaranjan Avenue
Calcutta - 700 072

Jayadeva Hostel Building
5th Main Road Gandhinagar
Bangalore - 560 009

3-5-1129 Kachiguda Cross Road
Hyderabad - 500 027

Prarthana Flats, 2nd Floor
Near Thakore Baug, Navrangpura
Ahmedabad - 380 009

Patala House
16-A Ashok Marg
Lucknow - 226 001

Central Bazaar Road
60 Bajaj Nagar
Nagpur 440010

INDONESIA
Pt. Indira Limited
Jl. Sam Ratulangi 37
P.O. Box 181
Jakarta Pusat

IRELAND
Government Supplies Agency
4-5 Harcourt Road
Dublin 2

ISRAEL
Yozmot Literature Ltd.
P.O. Box 56055
Tel Aviv 61560
Israel

ITALY
Licosa Commissionaria Sansoni SPA
Via Duca Di Calabria, 1/1
Casella Postale 552
50125 Firenze

JAPAN
Eastern Book Service
Hongo 3-Chome, Bunkyo-ku 113
Tokyo

KENYA
Africa Book Service (E.A.) Ltd.
Quaran House, Mfangano Street
P.O. Box 45245
Nairobi

KOREA, REPUBLIC OF
Pan Korea Book Corporation
P.O. Box 101, Kwangwhamun
Seoul

(continued on reverse)

Distributors of World Bank Publications
(continued)

MALAYSIA

University of Malaya Cooperative
Bookshop, Limited
P.O. Box 1127, Jalan Pantai Baru
59700 Kuala Lumpur

MEXICO

INFOTEC
Apartado Postal 22-860
14060 Tlalpan, Mexico D.F.

NETHERLANDS

De Lindeboom/InOr-Publikaties
P.O. Box 202
7480 AE Haaksbergen

NEW ZEALAND

EBSCO NZ Ltd.
Private Mail Bag 99914
New Market
Auckland

NIGERIA

University Press Limited
Three Crowns Building Jericho
Private Mail Bag 5095
Ibadan

NORWAY

Narvesen Information Center
Book Department
P.O. Box 6125 Etterstad
N-0602 Oslo 6

PAKISTAN

Mirza Book Agency
65, Shahrah-e-Quaid-e-Azam
P.O. Box No. 729
Lahore 54000

PERU

Editorial Desarrollo SA
Apartado 3824
Lima 1

PHILIPPINES

International Book Center
Suite 1703, Cityland 10
Condominium Tower 1
Ayala Avenue, Corner H.V.
dela Costa Extension
Makati, Metro Manila

POLAND

International Publishing Service
Ul. Piekna 31/37
00-677 Warszawa

For subscription orders:

IPS Journals
Ul. Okrezna 3
02-916 Warszawa

PORTUGAL

Livraria Portugal
Rua Do Carmo 70-74
1200 Lisbon

SAUDI ARABIA, QATAR

Jarir Book Store
P.O. Box 3196
Riyadh 11471

SINGAPORE, TAIWAN,

MYANMAR, BRUNEI
Information Publications
Private, Ltd.
Golden Wheel Building
41, Kallang Pudding, #04-03
Singapore 1334

SOUTH AFRICA, BOTSWANA

For single titles:
Oxford University Press
Southern Africa
P.O. Box 1141
Cape Town 8000

For subscription orders:

International Subscription Service
P.O. Box 41095
Craighall
Johannesburg 2024

SPAIN

Mundi-Prensa Libros, S.A.
Castello 37
28001 Madrid

Llibreria Internacional AEDOS

Consell de Cent, 391
08009 Barcelona

SRI LANKA AND THE MALDIVES

Lake House Bookshop
P.O. Box 244
100, Sir Chittampalam A.
Gardiner Mawatha
Colombo 2

SWEDEN

For single titles:
Fritzes Fackboksforetaget
Regeringsgatan 12, Box 16356
S-103 27 Stockholm

For subscription orders:

Wennergren-Williams AB
P. O. Box 1305
S-171 25 Solna

SWITZERLAND

For single titles:
Librairie Payot
1, rue de Bourg
CH 1002 Lausanne

For subscription orders:

Librairie Payot
Service des Abonnements
Case postale 3312
CH 1002 Lausanne

TANZANIA

Oxford University Press
P.O. Box 5299
Maktaba Road
Dar es Salaam

THAILAND

Central Department Store
306 Silom Road
Bangkok

TRINIDAD & TOBAGO,

ANTIGUA
BARBUDA, BARBADOS,
DOMINICA, GRENADA,
GUYANA,
JAMAICA, MONTSERRAT, ST.
KITTS & NEVIS, ST. LUCIA,
ST. VINCENT & GRENADINES
Systematics Studies Unit
#9 Watts Street
Curepe
Trinidad, West Indies

TURKEY

Infotel
Narlabahçe Sok. No. 15
Cagalolu
Istanbul

UNITED KINGDOM

Microinfo Ltd.
P.O. Box 3
Alton, Hampshire GU34 2PG
England

VENEZUELA

Libreria del Este
Aptdo. 60.337
Caracas 1060-A

