2005

A B C D E

Annual World Bank Conference on Development Economics

Lessons of Experience

Edited by

François Bourguignon and Boris Pleskovic

Annual World Bank Conference on
Development Economics
2005

Lessons of Experience

Edited by
**François Bourguignon and
Boris Pleskovic**

A copublication of the World Bank and Oxford University Press

Contents

Infrastructure and Development

Trade and Development

About This Book

The Annual World Bank Conference on Development Economics is a forum for discussion and debate of important policy issues facing developing countries. The conferences emphasize the contribution that 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. Conference papers are reviewed by the editors and are also subject to internal and external peer review. Some papers were revised after the conference, sometimes to reflect the comments by discussants or from the floor. Most discussants' comments were not revised. As a result, discussants' comments may refer to elements of the paper that no longer exist in their original form. Participants' affiliations identified in this volume are as of the time of the conference, May 3–4, 2004.

François Bourguignon and Boris Pleskovic edited this volume. Editing services were provided by Kim Kelley. The book was designed by Naylor Design, Inc. Book production and dissemination were coordinated by the World Bank Office of the Publisher.

Introduction and Summary

FRANÇOIS BOURGUIGNON AND BORIS PLESKOVIC

The Annual World Bank Conference on Development Economics is one of the world's best-known series of conferences on development. It seeks to expand the flow of ideas among scholars and practitioners of development policy from academia, government, and the private sector around the world. By fostering a better understanding of development and the problems developing countries face, the conference aims to enhance policymaking at the World Bank and at its partner institutions. It also provides a forum for exposition by academics and practitioners as they seek to identify and elaborate on new ideas and issues pertinent to development.

The 16th conference was held in Washington on May 3–4, 2004. The theme was lessons of development experience, which was divided into four topics: lessons of experience, behavioral economics, infrastructure and development, and trade and development.

The conference opened with remarks by James D. Wolfensohn, president of the World Bank, and a keynote address by François Bourguignon, chief economist and senior vice president, Development Economics. These were followed by two papers on lessons of development experience: a half-century of development (1950–2000), and the evolution of thinking on economic development. The second day began with a keynote address by Nobel Laureate Vernon L. Smith on markets, capital markets, and globalization, followed by papers on behavioral economics, infrastructure and development, and trade and development. The rest of this introduction summarizes the opening address and two keynote addresses, followed by papers on behavioral economics, lessons of experience, infrastructure and development, and trade and development.

François Bourguignon is Chief Economist and Senior Vice President, Development Economics, at the World Bank. Boris Pleskovic is Research Manager, Development Economics, at the World Bank.

Opening and Keynote Addresses: The Future of Development—The Next 10 Years

In his opening address, **James D. Wolfensohn** expresses hope that the conference will provide participants with the opportunity to discuss the theory of development and to remark on practices and events of the past, as well as to relate how their thoughts and experience might guide the World Bank in future initiatives. Wolfensohn discusses the challenges we face—such as inequity between countries and an imbalance around the planet with regard to gross domestic product, poverty, and hope—and identifies key issues for the coming century: poverty, social justice, and the related areas of migration and openness of trade.

Wolfensohn emphasizes the interdependence between developed and developing countries and points to the integral need for rich and poor countries alike to seek poverty reduction and establish hope. The interrelationships of trade and migration, and the other interrelationships that exist in the world, are such that we must find ways to ensure that those who are less fortunate have more of their needs satisfied and more opportunities for the future.

Both worlds must come together and work together: resources need to be made available by developed countries, and developing countries need to build up capacity and find ways to keep that capacity in the home country. Both sides have obligations to fulfill, but together they can achieve a more equitable world.

Wolfensohn argues that the issues of development and poverty are as important and as urgent as the questions of Iraq, Afghanistan, and the West Bank and Gaza. He emphasizes that, as we look forward to the next 10 years, we must determine how to move beyond statements, how to achieve a more equitable future, and how to get people to understand that we live in an interdependent world.

In his keynote address, **François Bourguignon** considers two issues. First, is the evolution of the global distribution of income improving or worsening over time? It is well known that the degree of inequality in the world is very high, certainly higher than within most individual countries. Some people argue that this distribution is worsening over time, and many blame the globalization process for this trend. But others claim the distribution is improving, and that globalization is responsible for that evolution. Second, redistribution is taking place in the world, but what kind of redistribution? There is a great deal of inequality, and if the world were a single country, a community mechanism would probably push for redistribution. In his analysis, Bourguignon focuses on two types of international policies as potential instruments for the redistribution of actual and potential income: official development assistance and trade liberalization.

Bourguignon provides the results of an analysis of the evolution of per capita income across countries over the past decade and presents measures of the distribution of that income, weighted and unweighted for population. He also presents the results of simulations on the impact of aid, remittance flows, and trade barriers across the distribution of global income, by decile. The data and simulations presented support three main conclusions. First, there has not been an unambiguous

evolution in the world distribution of income. If one weighs countries by their populations, there has been a positive evolution over the past two decades. Absolute poverty has declined, and therefore, in a certain sense, inequality has fallen. But if one views the world distribution in another way, by looking at the evolution of specific countries—especially those in Sub-Saharan Africa, the situation is growing worse. So when people say the distribution has worsened or has improved, it is important to keep in mind this double perspective. Second, redistribution through official development assistance is extremely limited, and is cancelled out, to a large extent, by trade restrictions that limit the market access of poor countries. Bourguignon recognizes the importance of worker and profit remittances as factors of redistribution as well; however, the analysis of these flows is more complex. Bourguignon therefore concludes that income redistribution does not capture a country's potential for the future—and he argues that the objective of aid is not to redistribute income today in order to increase immediate consumption but rather to transfer the potential for growth from rich countries to poor ones. Official development assistance should be aimed at improving those conditions, in particular by helping poor countries meet the United Nations' Millennium Development Goals. In the end, trade flows, capital flows, and migration flows should be organized so that they maximize the growth potential of the poorest countries. Only with this kind of objective will we be able to one day say that there has been unambiguous improvement in the world distribution of income.

In a keynote address delivered on the second day of the conference, **Vernon L. Smith** opens with a quote from philosopher David Hume, on trade: "Manufacturers, gradually shift their places, leaving those countries and provinces which they have already enriched, and flying to others, whither they are allured by the cheapness of provisions and labor; till they have enriched these also, and are again banished by the same causes" (1777). Smith argues that without exchange and markets, people cannot engage in task and knowledge specialization. And it is this specialization that is the key to all wealth creation. There is no other source of sustainable human betterment. We all function simultaneously in more than one world of exchange, and those worlds overlap. We live first in a world of personal exchange, trading favors and friendship and building reputations based on trust and trustworthiness in small groups and families; and second, in a world of impersonal exchange through markets, where communication and cooperation gradually emerge in trade with strangers.

Smith first discusses markets for commodities and services, which have been studied extensively in the laboratory, and then applies the learning from various experiments to interpret the nature and function of such markets in the world. These markets are the foundation of wealth creation. Second, he considers markets for capital, or stock markets. Capital and stock markets are far more volatile and unpredictable than are existing commodity and service markets, but this is to be expected, as their function is to anticipate the commodities and services of the future.

Smith also discusses globalization—the worldwide extension of such markets—and characterizes this as nothing more than a new word for the ancient process of

migration and human quest for betterment that began centuries ago when our common ancestors first walked out of Africa. Commodity and service markets are the foundation of wealth creation. The fact that stock markets supply capital for new consumer products explains why they are inherently uncertain, unpredictable, and volatile, and—given investor behavior—why they tend to bubble and crash. Stock markets are far more uncertain than markets for commodities and services because stock markets must anticipate innovations, or the new commodities and services of the *future*. Smith believes that globalization is a good and peaceful word. In the wise pronouncement commonly attributed to the great French economist Frédérich Bastiate, "If goods don't cross borders, soldiers will."

Behavioral Economics

Sendhil Mullainathan argues that economists conceptualize a world populated by calculating, unemotional maximizers—and that this view shapes our understanding of many crucial elements of development economics, from how rural villagers save, to how parents decide on whether to send their children to school. Psychological research has, however, documented the incompleteness of this perspective. Individuals have self-control and time inconsistency problems. They can give into short-run temptations and later regret it. They can have strong feelings about others that drive them to commit both generous acts and spiteful ones, and they often passively accept defaults rather than make active choices. They let the institutions around them make choices for them, and they may misread new data in a way that conforms with their beliefs.

In short, the rational maximization model may not be a very good approximation of human behaviors. In this paper, Mullainathan presents some of the psychological evidence that may help policymakers better understand a few core behavioral issues in development economics, such as savings, education, and property rights. Mullainathan also emphasizes that this evidence provides us with new interpretations for a variety of behaviors in these contexts and enriches the set of policy tools we should consider. This suggests the need for some dramatically new tools, but also suggests that small cost changes may dramatically improve the efficacy of existing policies.

Lessons of Experience

Richard N. Cooper presents development as a global policy objective that dates from the 1940s. Relative to expectations at that time, the world economy performed outstandingly well during the second half of the 20th century. Worldwide growth in average per capita income exceeded 2 percent per year (historically unprecedented), many poor countries became rich, infant mortality declined, longevity increased, and

diseases were contained and sometimes vanquished. Poverty by the World Bank definition of $1 per day (in 1985 U.S. dollars) declined dramatically, and the number of people living in poverty was halved—despite a more than doubling of the world population. Variations occurred over time and space, with rapid growth concentrated in Europe and Japan early in the period, then moving to East Asia, Southeast Asia, and South Asia. Growth in the 1950s and especially the 1960s exceeded that in later decades. Examples of high growth could be found on every continent, but on average Sub-Saharan Africa fared much less well than other regions. Declines in national per capita income were rare, and were concentrated in Africa. Civil disorder was common but not the universal cause of low growth. Median world income gained relative to that of the well-off, but both of these groups surged ahead of the poorest.

Cooper notes that world exports grew more rapidly than output, often leading the way. Many countries gradually shifted their exports from primary products to labor-intensive manufactured goods, and, as development proceeded, to more sophisticated manufactures and services. The fraction of the labor force devoted to agriculture declined significantly. One country after another achieved social stability, created the right incentives for effort and risk-taking, and engaged constructively with the world economy, which facilitated economic growth. Those that lagged behind failed to meet one or more of these conditions. Civil and political liberties also spread during this period, although less certainly and less securely. Cooper concludes that on the whole, the past 50 years were a good half-century for mankind, and that the substantial poverty and misery that still exist should not detract from these achievements.

Gustav Ranis discusses the course of development thinking and associated development policy over the past six decades. He first addresses the early postwar consensus, with theory focused largely on the revival and extension of classical dualism, and policy concentrated on creating the preconditions needed for development and the severing of colonial ties viewed as tied to the market. He then traces the increasing awareness of the role of prices, the rejection of various types of elasticity pessimism, and a diminishing reliance on the developmentalist state as the main actor. The international financial institutions gradually moved toward increased reliance on structural adjustment lending associated with conditionality and reform at both macro and micro levels of policy, embodied in the Washington Consensus and its extensions. The third section illuminates the search for "silver bullets" over time, in both the theory and policy arenas. Here, Ranis demonstrates the never-ending search for dimensions of development that can be identified as critical, or key, to achieving success.

Ranis concludes with a personal assessment of where we are and where we will be (or should be) headed in the effort to achieve the Third World's basic objective of human development fueled by equitable growth. He points to the future directions in development thinking, including theoretical framework, two-way relationship between growth and improvements in human development, decentralization and democratization, and liberalization. Ranis also outlines necessary policy reforms in

resource-rich countries and an easy access to foreign capital as extensions of the Dutch Disease problem.

Infrastructure and Development

Rémy Prud'homme presents infrastructure as a subset of the notion of capital. He dismisses several definitional characteristics of infrastructure and argues that infrastructure as an analytic concept was essentially absent from the economist's tool box for nearly two centuries. By contrast, during the 1990s, a vast body of literature introduced infrastructure as a determinant of production functions, with a view to estimate its contribution to economic growth. Unlike productive capital which is homogenized by market forces, politically driven infrastructure may—and often does—consist of white elephants as well as highly useful roads. Why and how does infrastructure contribute to development? It is a space-shrinker, it enlarges markets, and it operates like the lowering of trade barriers. In urban areas it can be shown that infrastructure helps enlarge the effective size of the labor market and of the goods or ideas markets, thus increasing productivity and output.

Prud'homme remarks that institutional and financial regimes have a direct impact upon the socioeconomic efficiency of infrastructure. Because infrastructure always has a government dimension but can also have a private dimension, the menu of institutional options available is quite large and ranges from direct government provision to unsubsidized concessions with various forms of public-private partnerships, such as subsidized concessions or shadow tolls. Three mechanisms have to be taken into account: the welfare loss often associated with tolls and prices, which implies that in such cases, all other things being equal, nontolled infrastructure projects are better than tolled ones; the cost advantage usually associated with private production, which implies that, all other things being equal, privately managed infrastructure is better; and the distortionary impact of taxes, which implies that, all other things being equal, toll-financed infrastructure projects are better than tax-financed ones. Prud'homme notes that a simulation using a small model that combines these three mechanisms suggests that the more private options, in particular the shadow-toll option, are economically superior to the more government-oriented options. The problem is complicated, however, when the public-finance dimension of the various options is considered.

Forecasting errors and associated risks are characteristic of infrastructure projects, with costs generally underestimated and patronage overestimated by large amounts. Errors of 50 percent or more seem to be the rule rather than the exception. But an understanding of the various reasons for such errors is useful in allocating the related risks between government bodies and private partners. Substantive risks (those linked to changes in project scope or design) and pure economic risks (those associated with the macroeconomic environment) are not insurable and should therefore be borne by the public entity. But technical risks (those linked to errors in forecasting costs and usage) should be borne by private enterprises. Prud'homme concludes

that institutional errors can only be reduced by strategic changes in institutional design and contracts.

Trade and Development

Riccardo Faini argues that international economic integration has been on the rise since at least the mid-1980s, but that the present episode of globalization has a number of distinctive features that make it more vulnerable. First and foremost, the nature of foreign direct investment (FDI) has changed. In the past FDI was mainly directed to establish production facilities in foreign markets, with a view to catering to foreign consumers and circumventing trade restrictions. More recently, however, investment decisions by international firms seem to be increasingly driven by the desire to cut production costs. Hence, trade costs play a very different role in this context: they no longer encourage multinational firms to invest abroad with a view to gaining better access to foreign consumers. Trade and FDI have increasingly become complements. And as a result, any step back in either dimension may negatively affect the other.

Similar considerations apply to the links between trade and migration. The two have typically been seen as substitutes, and the creation of NAFTA was, in fact, hailed by then President Salinas of Mexico as reflecting the desire of his country "to export goods, not people." This relationship may still hold true, but increasingly less so. In particular, the growth in service trade, one of the most dynamic components in the expansion of international trade, very much depends on the ability to supply such services *in loco* through a firm's own personnel, and is therefore, positively linked to migration.

The main findings of Faini's paper can be summarized as follows. First, while trade and FDI policies have become steadily more liberal throughout the world, immigration policies in receiving countries have grown quite restrictive. Only for skilled workers has immigration policy in the main receiving countries become more generous. Second, at least during the 1990s, import trade liberalization fostered not only trade but also inward investment, confirming that trade and FDI toward developing countries have been, most recently, largely complements. Third, the presence of a skilled labor force is a relevant factor in attracting FDI. Moreover, open trade policies and the stock of FDI have a positive impact on the incentives to acquire education. This set of findings highlights the possibility of a low-equilibrium trap where the lack of human capital discourages FDI and inadequate investment from abroad limits the domestic incentives to acquire education. Rich countries, by encouraging skilled immigration from relatively poor countries, definitely aggravate such a risk. Fourth, there is little relevant evidence to support the contrary argument of a brain gain, where the possibility for skilled workers to migrate abroad raises the return to education and the investment in human capital in poor countries.

* * *

As in previous years, the planning and organization of the 2004 conference was a joint effort. We wish to thank several anonymous reviewers for their advice, as well as Aehyung Kim, Gerard Rice, and David Rosenblatt for their useful ideas and suggestions. We would also like to thank conference coordinators Theresa Bampoe and Leita Jones, whose excellent organizational skills helped ensure a successful conference. Finally, we thank the editorial staff for pulling this volume together, especially Cindy Fisher from the Office of the Publisher, and Kim Kelley, a consultant for the Office of the Publisher.

Opening Address

The Future of Development:
The Next 10 Years

JAMES D. WOLFENSOHN

Let me tell you what I believe to be the challenges that we're facing now, and, frankly, why I need your help—and why I think the Annual Bank Conference on Development Economics is not only an opportunity for you to discuss with each other theory and to remark on practice and events of the past. Perhaps we can also relate your thoughts and your experience to the ways in which we might move in the future. Because it is the pragmatic sense going forward that I'm most concerned about at a time when I am deeply concerned about the level of development and, indeed, the level of peace on our planet.

I think you all know my assessment of where we stand. An assessment that in the developing world, five out of six billion people on the planet, a split of 80/20 in terms of income, experience inequity within countries, and between countries. My concern is with what I call an imbalance on the planet in relation to GDP, in relation to poverty, and in relation to hope.

Again, rather simplistically, it seems to me that if someone has an investment in hope—in the future and in their family, and for an opportunity to make a living— they're a lot less likely to go out and shoot you. And if instead of hope there is despair, then there is a much greater likelihood that they can be influenced.

I thought this issue of hope was brought to the attention of the world on September 11, 2001, when the notion of the developed and the developing worlds having a wall between them was shown to be fallacious. And for the first time we came upon visible evidence of the problems of Afghanistan and, indeed, problems not of the Afghan people but of people who were using Afghanistan as a haven. Evidence of those problems was now on Wall Street and at the Pentagon. And for me, the image of those buildings falling down was an image of the two worlds coming together in a highly visible way.

This notion of interdependence is my preoccupation at the moment and is something that I am really concerned about and would look to you for guidance. But to

---

James D. Wolfensohn is the President of the World Bank.

me, the notion of the developed and developing worlds, with an imbalance between them in terms of pace of development, technology, wealth, or in terms of the future, is the issue that keeps me awake at night as I look forward to the next 25 years and recognize that our planet will move from six billion to eight billion and that, very broadly, seven billion of these people will live in developing countries—while the rich countries will have grown by maybe 50 million people. Europe as we know it today will be smaller and older. The issue of youth is one in which 2.8 billion people on our planet today are under 24 years of age, and obviously if two billion more come onto the planet in terms of net population increase, many more than that are born. So we have a situation today where 1.5 billion young people under the age of 15 are looking for work, so we need to create a billion jobs in the next 25 years.

This presents me with a problem, which, of course, leads to the attendant issues of migration (another big issue that I hope you might help me with), as it leaks into the question of trade and the openness of trade. A former Mexican leader indicated at the time of the NAFTA negotiations that it would be better to import work rather than export people.

So this interrelationship between the developed and the developing worlds, the tensions arising from it, and the fact that where once you could contain terror, where once you could contain migration and its impact, in today's world of communications, these are significantly bigger issues. I do not believe that we currently have the methods needed to deal with these questions of development, either in terms of organization or in terms of intellect.

Let me share my thinking on this. At the time of the U. N. Millennial Assembly, the world leaders all came together and in different ways recognized this challenge. The real element to come out of the Millennial Assembly was the Millennium Development Goals—which said that the key issues for the coming century are poverty and social justice, or social goals that need to be achieved.

If any of you have ever gone back to read the assembly speeches, you'll recall that they are, one after the other, reiterations on the same theme. The world needs greater equity. If the world is to be stable, it needs social justice. There is an integral need for rich and poor countries alike to acknowledge the war on poverty and move forward; to understand that poverty somewhere is poverty everywhere; and that the interrelationships on trade and migration—and the other interrelationships that exist in the world—are such that what we must get on to is making sure that those who are less fortunate become at least satisfied, and more fortunate.

Here I must say that the paper Dick Cooper prepared for this conference is, I think, a remarkable *tour d'horizon* of the past 50 years and points to significant progress. But nonetheless, at the Millennial Assembly in September 2000, people took note of the world as they saw it, pushed firmly by developing countries which are not satisfied with the development achievements that have been made to date—whether measured at $1 or $2 a day, or measured in some purchasing power parity. We can play with estimates of whether there are 1.2 billion people who are really poor, or 800 million instead. But having a $1 or $2 equivalent is still a long way

from where it is in the wealthy countries. And, of course, people living in poverty know a lot more now because of communications.

But, in any event, at the Millennial Assembly, these objectives were set and then reiterated at meetings in Monterrey and Johannesburg, and subsequently at the New Partnership for Africa's Development exposition (NEPAD), which Africans thought was the real deal. Here it was recognized that if there was to be development, it would be a bargain between two sides; that they needed to come together and work together; that resources needed to be made available; and that there were responsibilities. Responsibilities for developing countries to build up capacity but not export capacity, which has, of course, been a singular problem; to build the mechanisms and the framework needed to support legal, judicial, and financial systems that work; and in terms of behavior, to focus on the issue of corruption.

This was not a hidden condition imposed by the World Bank or the International Monetary Fund. These were the terms of a freely agreed upon partnership asserted at the assemblies in Monterrey, Johannesburg, and NEPAD.

So I think you'd have to say that some progress has been made by developing countries, significant progress in some countries, but that a huge challenge still exists in terms of the pace and the expectations of people, which occasionally have led to a revolution. In other cases expectations have led to economic and political turmoil.

On the other side, the wealthy countries said, "We'll help you build your capacity, we'll open our markets for trade, and we'll increase aid."

Now we are in a situation where everybody is talking to this proposition, simplistic as I have described it, of the two sides having obligations that they need to fulfill, and that together the result will be a more equitable world. But, in fact, the issue that I constantly face is the issue that, first of all, the development in too many developing countries is more an of an indicator of the direction to go rather than movement in that direction.

At these last meetings (NEPAD), when we discussed corruption, I said that it takes me about 24 or 48 hours in any country to know who the crooks are. You know if the president is a crook, if his wife or cousin is a crook, how many of the cabinet members are crooks. It is not difficult in almost any developing country in a very short time to understand where the real elements of corruption are.

But there's still power in the elements that are corrupt, and it's difficult to do something about it. And at the core of the issue of corruption is a cancer that needs to be excised.

There is now a lot more movement toward building institutions, but I think all of us here know that building institutions is not just a matter of changing the law. It's also about having a change in culture, and building the capacity to do so. If you have lawyers who have been corrupt for 30 years and you get a few honest judges, you still have corrupt lawyers. If you change the law, you still have to train people in how the law has changed. So the pacing of this has proven to be difficult, but the direction has certainly been established.

The other element that I find difficult is the rich countries. I need hardly tell you that Doha has not been a huge success thus far in terms of agriculture and in terms of other opening of trade. And I need hardly tell you that aid has not been a huge success.

So when you come down to the issue of aid, you have a serious problem, as we look forward, in terms of the level of aid and the actual cash that is passing over to countries to deal with the fundamental social issues that they're seeking to address. And I find that extraordinarily worrisome. Grants, for example, imply that you're going to have to increase your lending in the future because there are not going to be any repayments.

To give you an idea of what's at stake, we currently have $115 billion outstanding with the International Development Association, or IDA, on which we have no countervailing claim as we do in the case of the World Bank, which is a bank that borrows money. We could forgive that amount tomorrow. But if we do, we won't get $3.5 billion cash flow coming back, which is relied on annually to make part of the $8 billion which is given.

So we can turn it all into debt relief and turn it all into grants, but that has the impact of doubling the contribution for IDA, which, of course, raises the same question again. These questions of debt relief, of new debt, of grants, of trade even, are all linked in some way.

And then the leaders come back to the proposition that we've got elections to run, we've got crises to face, we have our own growth to deal with, we have our own poverty to deal with, we have our own unemployment to deal with, we have expatriation of jobs to deal with, and we have an election to win.

So my big worry is that we have this huge build-up in terms of the development assembly. We have intellectual recognition of this imbalance. You have people on each side espousing the right ideas—but you don't have the action.

I see my friend Pedro Malan in the front row. I'll just comment that when we were in the G-8 meetings—and let me say I'm not often in G-8 meetings, so I'm not name-dropping. But I did happen to be there on that day when President Lula of Brazil came into the room, and we had in the room President Hu of China, Vajpayee of India, Obasanjo of Nigeria, Bouteflika of Algeria, and Thabo Mbeki from South Africa. Lula came in—a man who, by the way, I think is continuing what Pedro and Fernando Enrique Cardoso started, this amazing and critically important social revolution in Brazil.

He came in and said, you know, "I'm a trade union leader, and my parents would have been very honored to see me here with you, President Bush, and you, President Chirac, and you [the other leaders], and I feel deeply honored at being at a G-7 meeting. But it does occur to me as I look at my other colleagues from India, China, Africa—that maybe next year we should be the G-7 because, after all, we represent five-sixths of the world. And as we look forward, it will be seven-eighths of the world." And the look of shock on the face of others that were there was really palpable.

And what I think Lula was saying was, listen, you've got to take us seriously. We are not a Hollywood event. We are real. Of course, this was further exemplified

in the Doha Round meetings in Cancun. But what I find difficult is that really a joining of forces on the issue of development has not yet taken place. You have a statement about the issues on all sides. You have recognition of the issues, in ways that I think many of you would be proud of. But in terms of acting on the issues, on either side, the commitment is far from perfect.

In fact, if you look at the overall statistics on the use of our resources, I come back to the simple proposition that today we're spending $900 billion on defense. We have tariffs and agricultural protection at the level of around $300 to $350 billion—and we're spending $68 billion on what I described as development. Of course, that does not include internal expenditures on development. This is only for overseas development assistance.

But if one could really look forward in these next 10 years and deal with these issues in a pragmatic way, we wouldn't be dealing with it by spending $900 billion on defense. We would be putting a lot more resources, intellectual and financial, into aid, and trade, and the questions of rebalancing the world.

Another thing is that, even within what we're now doing, as we look forward to the next 10 years, at what we've started—and I've seen the commentary—some believe, as mentioned in Professor Ranis's paper, that the poverty reduction strategy papers (PRSPs) do not sufficiently take into account the views of developing countries, and that they are not sufficiently doing what they're supposed to do. The strategies aren't sufficient, but they're a hell of a lot better than we had in the past. And they're continuing in terms of participation.

But we also face the challenge in the next 10 years of leadership within the countries, in bringing people together within the countries, but also among ourselves—in bringing together the development efforts that are already taking place. The cooperation of IFIs, the bilateral institutions, civil society, which is now huge compared with even 10 years ago, both in terms of numbers and in terms of transfers of financial resources. And then, of course, there's private sector interest, which goes up and down in countries with regard to overseas investment, but where there is a need to create within countries a framework in which investment can comfortably take place.

What we have before us is a lot of work that must be done within the framework of existing resources coming together. And those challenges are very real, because there are "boxes" in which people put institutions.

In reading the papers I've noticed that many of you put the World Bank in a box in terms of what we are and are not. And from what I read there are a lot of commonly held beliefs about our slavishness to the Washington Consensus and to our being behind the eight ball in terms of moving forward. I'm sure many economists have a view of our institution that I would have to say is frequently, though not in the case of anyone in this room, outdated. And I can say this because I've been traveling around to universities, talking to students, and hearing what they've been told.

But today's World Bank is a different institution. This is an institution that does not have the answers to all these questions but does have a sense of direction, which may or may not be right but is today more humanly based. We are not starting so much at

the macro level, although we do understand that growth is essential. But as mentioned in some of the papers at this conference—and as I think we believe—growth alone is not sufficient.

And so we need to look at the issue of how we get to a more equitable future, how to spread these resources, how we get the will, how we get the global balance so that people can understand that we must now move to planetary considerations. When we look at Sub-Saharan Africa, with dozens of countries and 670 million people, how we can get through this and look first to regional development and then to global interdependence is a challenge that I think we're not yet ready for.

And I think what we need from many of you is not just commentary on the past, but ways to help us move forward in the future, to help us to benefit from the past, and to give us remedies. Push us, push the world, become advocates in relation to bringing about change. It's not enough to give the commentary.

The problem that I'm facing and that practitioners are facing is, How do we get movement? How do we move beyond the statements? How do we get people to understand that this is an interdependent world, where, in 2050, 40 percent of the global GDP is likely to be in developing countries? How do we educate our kids so that they know something about Islam and other cultures? How do we educate our kids in developed countries to face the future of development? And they're not being educated this way.

How do we get kids in developing countries to love, not hate? How do we deal with the madrassah questions? How do we confront the issues that face us around the world when we are comfortable with the old systems and do not fully recognize that the world has changed in an absolutely fundamental way?

How do we deal with the question of technology, not just industrial technology but information technology, which is changing our world? In Andrha Pradesh, India, every village is now linked by fiberoptic cable. What difference does that make to the potential of that state and to the potential of their knowledge, and their expectations, and their hopes?

My worry, as I approach nearly 10 years in this job, is that we do a lot of talking and a lot of analyzing, but that what we need is a kick, a sense of anger, a sense of urgency. To me the issue of development and poverty is as important and as urgent as is the question of Iraq, the question of Afghanistan, and the question of the West Bank and Gaza.

To me, this is the urgent issue. But we're not yet there. We quietly examine it. We comment on it. We get people to say the right things. We score points off each other—when, in fact, the issue is survival and peace. And that is why I hope that this conference will come up with some prescriptions, some ideas, some risks and approaches that can help us and the poor, the uneducated, and practitioners to try and make a difference.

Thank you very much.

Global Distribution and Redistribution: A Preliminary Review[1]

FRANÇOIS BOURGUIGNON

Allow me to begin this keynote address with the sad news that I have to share with you. Our community of economists suffered a huge loss during the weekend with the passing away of Jean-Jacques Laffont. He had been sick for some time now. His family and his friends thought that he would recover from the illness, but finally the disease caught up with him and he died last Saturday.

This is very sad news for the whole community. He was 57. He leaves behind a huge contribution to economics. As you know, in the second part of his career, he became very much interested in development issues. He applied his knowledge, his expertise in industrial organization and regulation economics, to the reform of industrial sectors and infrastructure in many developing countries: China, several Latin American countries, in particular in Argentina, several African countries and the transition economies.

He came several times to ABCDE, and was involved with various projects in the World Bank. I think that with him our discipline is losing an intellectual leader and a wonderful human being.

What I will be talking about today is in line with Jean-Jacques' commitment to development, and it is very much in line with what Jim Wolfensohn just said a few moments ago. I will discuss global distribution and redistribution—a combination of two strands of the development literature. First, global distribution has raised very much interest and much debate lately with some analysts asserting that things are improving, because inequality in the world is going down. Supporters of this view emphasize that inequality among world inhabitants, and poverty in absolute terms and in relative terms has decreased. Meanwhile, others assert that exactly the opposite is going on, and they tend to associate the process of globalization with the process of increasing inequality and increasing relative poverty in the world.

François Bourguignon is Chief Economist and Senior Vice President, Development Economics, at the World Bank.

The present level of inequality of living standards in the world and the rates of absolute and relative poverty are extremely high. Between-country inequality indices far surpass almost all individual country indices. There are two ways to reduce poverty and inequality: generate faster growth in the poorer countries of the distribution and/or redistribute from rich to poor countries. This address can be thought of as an initial attempt to evaluate these two forces. I start by reviewing how the impact of growth has affected the global distribution of income across countries in 1980–2002, and I then proceed to study the redistribution that takes place via aid flows as well as the potential income that can be generated from liberalizing trade.

There has been an active literature in recent years on the world distribution of income. The literature has evolved from the early discussions of the inter-country (or international) distribution of income where countries are weighted equally, with China having the same weight as Barbados.[2] Later studies still treated countries as the unit of analysis, but weighted their influence on world income distribution by population size.[3] More recent papers have analyzed the "true" world distribution of income, combining estimates of within-country inequality with those of inter-country inequality.[4]

I will focus on the inter-country distribution of income, abstracting away from the admittedly critical element of within-country inequality, in order to evaluate international policies that can potentially affect the inter-country distribution of income. However, unlike the literature on macroeconomic convergence,[5] I will emphasize population-weighted results as a more accurate measure of the human impact of international policies and macroeconomic growth. Besides fitting with this address's international policy perspective, between-country inequality accounted for approximately 60 percent of overall inequality across the world's citizens during the last decade.[6] As a result, it is an interesting topic of study in its own right, but my primary reason for ignoring within-country inequality is to facilitate the simulation of the impact of policies. In future work, one might incorporate within-country distributional elements; however, then one would have to estimate how aid and other cross-country flows are distributed among the citizens of a given recipient country. Incorporating within-country heterogeneity would also raise the issue of consistency between household survey data and national accounts data in estimating the world distribution of income. In brief, the analysis here should be considered as a rough first-order approximation on global distribution and redistribution.

One of the inter-country redistribution channels is the allocation of foreign aid by high-income donors to developing country recipients. To the extent that official development assistance (ODA) is in grant form, these transfers can be thought of, in the short run, as pure redistribution of national income.

A successful implementation of multilateral trade liberalization would increase the world's and many countries' welfare, with interesting distributional consequences. In effect, existing protection may be seen as a way through which countries affect the world distribution of gains from trade, and therefore, the world distribution of income. Trade protection or liberalization may thus be seen as a redistribution device in the same way as ODA. Using the World Bank's Linkage Model of computerized general equilibrium simulations prepared for the *Global Economic Prospects* report,

I will analyze the inter-country redistribution of income resulting from full or partial merchandise trade reform.

From a global national accounts point of view, it might be thought that worker remittances technically also serve as a channel for between-country income redistribution—the same being true of profit repatriation on foreign direct investment. However, this redistribution is not comparable to ODA or liberalization of trade in the sense that remittances and repatriated profits simply are the counterparts of a service delivered by migrants or foreign companies. Similar to trade in goods, world redistribution potential lies in the openness of countries to migrants and foreign capital. But the analysis of the actual redistribution taking place through these channels is less straightforward than in the case of aid, or trade.

In my address today, I will first describe the historical trends in inequality. Then, I will discuss the impact of current and increased aid flows and the potential impact of trade reform. I will conclude with a discussion of some tentative policy implications of the quantitative results.

The Current Inter-Country Distribution of Income

The current degree of between-country inequality in the world would likely be a cause for severe social strife if it existed within a single country. The existing international distribution of income is deeply rooted in the history as well as the geography of present-day nation-states. It is as much a consequence of the distribution of the factors of production—labor, capital, and natural resources—as that of access to technology, development of human capital, opportunities for trade and a variety of domestic institutional, social and cultural factors that affect countries' productivity in employing the factors of production.

For the analysis below, data are from the 2004 *World Development Indicators*. A number of countries were eliminated from our sample due to an incomplete time series of PPP (purchasing power parity) corrected GNI for the period under study. We arrived at a constant sample of 138 countries representing approximately 91–92 percent of world population over 1980–2002, for the historical analysis below. (For the case of the former Soviet republics during the pre-1992 period, we used data on net material product growth rates to extrapolate back to this earlier period.)

Figure 1 reveals more details on the inter-country distribution of income in 2002. For example, in population-weighted terms, the poorest 40 percent of countries received just over 10 percent of world GNI, while the richest 20 percent got more than 60 percent of world gross national income. If we compare the two extremes, the ratio of the top vintile's income per capita to the bottom decile's income per capita was 32 to 1.

We can also look at the data from the point of view of applying equal weights to countries—giving China and Barbados the same weight. One can see from the figure that the poorest 40 percent of countries have even less than before—just under 8 percent of global national income.

FIGURE 1.
2002 Distribution of Global GNI (1995 PPP Dollars): Decile/Vintile Shares
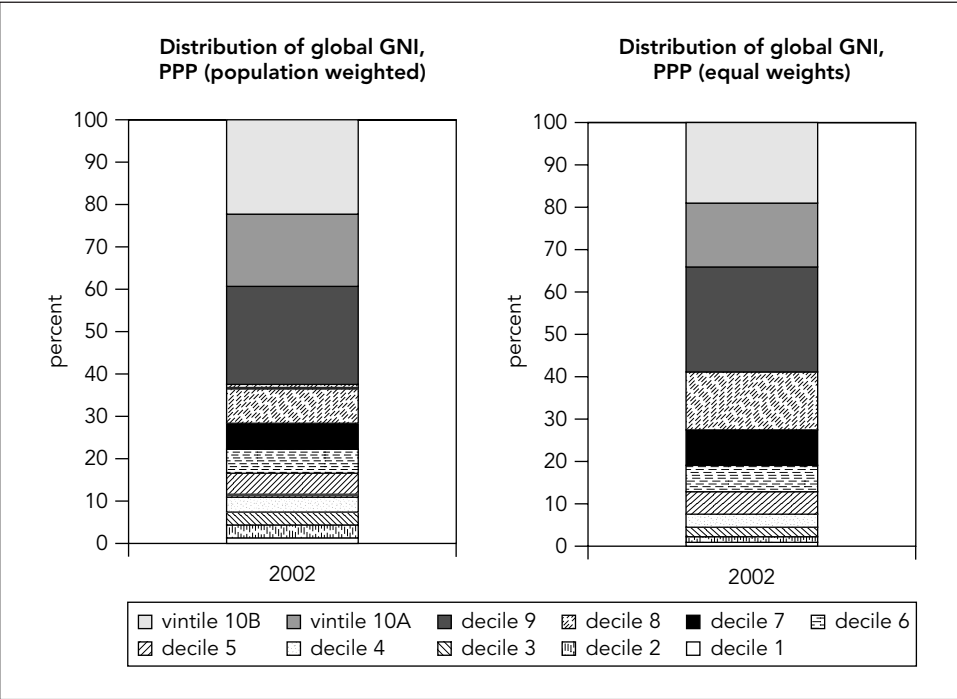
Source: Author's estimates based on data from the *World Development Indicators* 2004.

Historical Trends

To represent inter-country distribution of world income over the 1980–2002 period, we derived Lorenz curves and calculated standard inequality indices. The figure below reveals the evolution of nine deciles and the top two vintiles of countries' GNI, PPP over the period.

Figure 2 reveals a clear increase in the income share of the bottom 70 percent of the countries, weighted by countries' population. There is a noticeable improvement in deciles 2 through 6, matched by shrinking shares of the seventh and eighth deciles. The poorest decile improved somewhat during the middle of the period; however, this progress was reversed in the latter part of the period. The figure also reveals an increasing income share of the top vintile. Comparing 1980 and 2002, there is no Lorenz dominance: the improvement in the distribution due to a higher share for the poorest deciles is offset by an increasing share of the top vintile.

The trends of summary statistics in figure 3 also reveal the decreasing overall inequality; however, there is no uniformity among inequality measures, reflecting the absence of Lorenz dominance. Thus, the Atkinson indices show there is a slight increase in the inequality measure in the second half of the period if one assumes a high level of inequality aversion (e = 5), and a continuous decline otherwise.

FIGURE 2.
Historical Trend in the Distribution of Global GNI
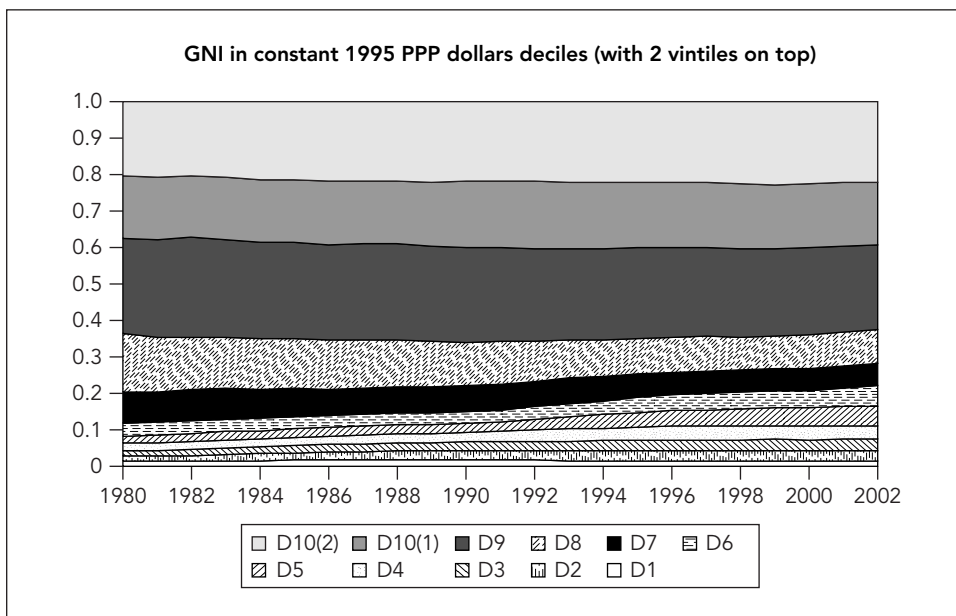
Source: Authors' estimates using data from *World Development Indicators* 2004.

FIGURE 3.
Trends in Population-Weighted Inter-Country Inequality
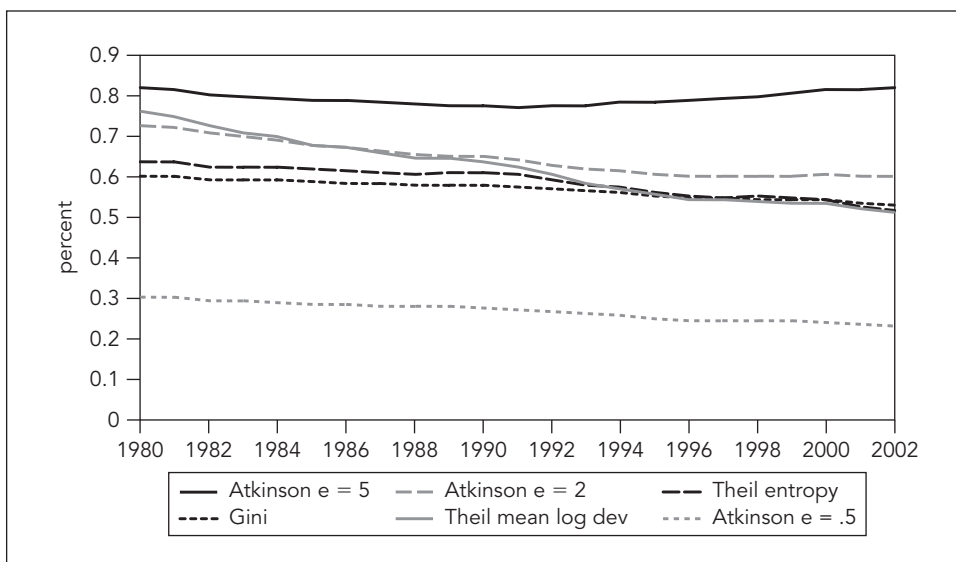
FIGURE 4.
Annual Growth of GNI per Capita per Decile, 1980–2002
(population-weighted)
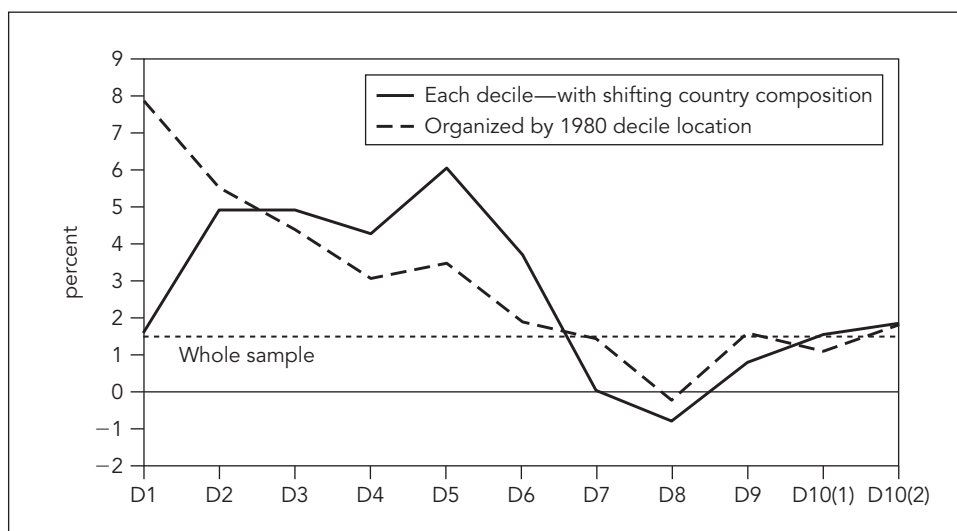
Source: Author estimates based on data from *World Development Indicators* 2004.

Even though most inequality indices show a strong and persistent decline, it is the case that some countries have been left behind global trends. Figure 4 illustrates this phenomenon. If, on the one hand, one traces the growth of income per capita of each decile allowing countries to move in and out of those deciles, and on the other hand, one traces the growth of income per capita of countries that started the period in a given decile (thus fixing the composition of the deciles in 1980), then one sees a rather different pattern. The lack of progress of the bottom decile relative to the world GNI per capita growth is due to the fact that this decile is abandoned early on by fast-growing China, which rapidly moves across the deciles from 1980 to 2002. Meanwhile, a group of countries from the second decile or even higher experienced negative per capita income growth over the period and sank into the bottom decile.

Another view of this phenomenon is to track the performance of income per capita for the countries that were in the bottom decile in 1980. Table 1 summarizes the results. We see that not only did these countries (other than China) not perform as well as the average, there were actually negative growth rates for several countries over the period. In addition, if we look at the performance of the 26 countries that replaced China[7] in the bottom decile in 2002, those countries experienced an average annual *negative* growth of about −0.85 percent.

If we give equal weights to every country, then the remarkable progress of large, previously poor countries is muted. We see in figure 5 that there is then a clear trend of increasing inequality measures if we treat each country as an individual. This type

TABLE 1.
Breakdown of the Performance of Income per Capita: Bottom Decile
(population-weighted)

Bottom decile of 1980	Annual growth rate (1980–2002)
Malawi	−0.10%
Guinea-Bissau	−0.19%
Mozambique	1.52%
Congo, Rep. of	−0.07%
Chad	1.26%
Burundi	−0.94%
China	8.20%
Pop-weighted average of decile	*7.86%*

Bottom decile of 2002	Weighted averages of group's annual growth rate (1980–2002)
Same countries as above, except for China	1.00%
26 more countries, of which:	−0.85%
a) 19 countries with negative growth (Tajikistan, Congo (DRC), . . .)	−2.26%
b) 7 countries with positive growth (Nepal, Bangladesh, . . .)	1.77%
Pop-weighted average of decile	*−0.76%*

Source: Authors' calculations based on data from *World Development Indicators* 2004.

FIGURE 5.
Trends in Inter-Country Inequality, Countries with Equal Weights
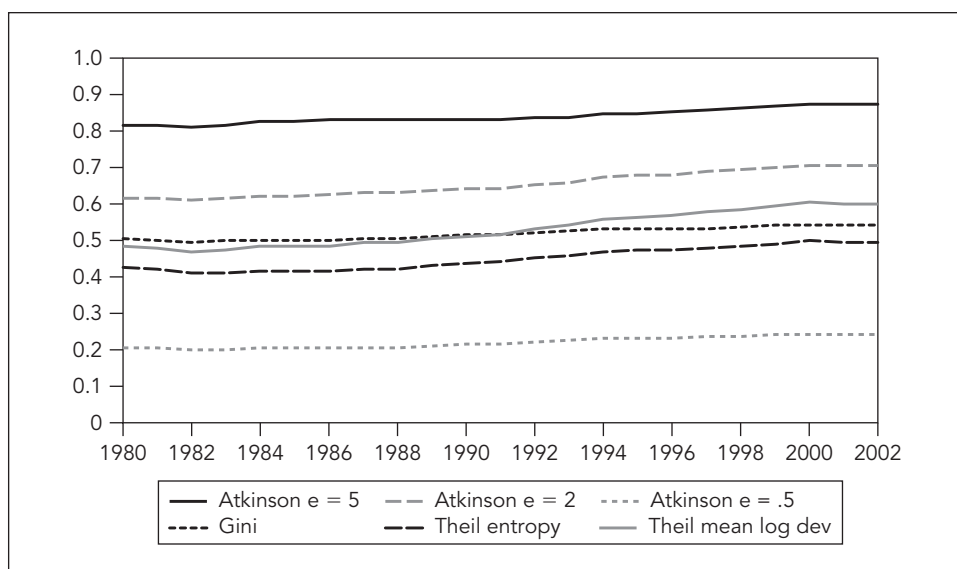
FIGURE 6.
Annual Growth of GNI per Capita by Decile, 1980–2002
(equal weight per country)
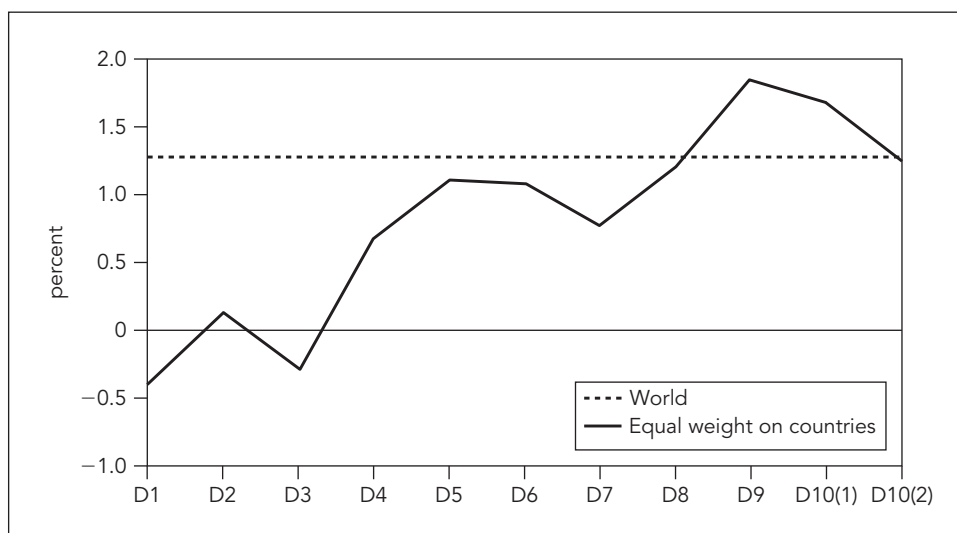
Source: Authors' estimates based on data from *World Development Indicators.*

of result is consistent with the recent literature on macroeconomic convergence, or lack thereof.[8] It is interesting to note, however, that the trend starts to level off during the last three years of the period, as developing countries' growth has improved.

Another representation of this increasing inequality is given in figure 6. Giving countries equal weights, growth in 1980–2002 appears as strongly and positively correlated with income levels. In effect, growth was negative or very close to zero for the bottom three deciles of our 138 countries.

In conclusion, we can say that no Pareto improvement has taken place in the world between 1980 and 2002, which leaves room for different value judgments about the evolution of world welfare, inequality, and relative poverty.[9] The preceding analysis also explains the divergent views expressed in the literature about increasing or decreasing world inequality. In particular, a critical factor in forming those views is whether one takes an "anonymous" approach to the distribution or whether one tracks precisely the actual income trajectories of individual countries.

International Redistribution through Aid

The simplest and most direct instrument of international income redistribution is official development assistance (ODA). ODA flows are a part of a recipient country's gross national income, and one can simply deduct these flows from GNI to determine

the static effect of aid as an instrument of income redistribution; similarly, since donor's GNI is calculated after ODA disbursements, one would simply add the ODA provided by a donor back to the donor's GNI to obtain the counterfactual.

The economic effects of aid transfers—even without considering distributional impacts within countries or externalities—are more subtle. There is a substantial theoretical literature on the general equilibrium effects of transfer flows between countries.[10] The origin emanates from the early papers on German war reparations—aid "in reverse," one might say—where there was an abundant discussion as to whether these financial flows represented a real resource transfer equivalent, greater or less than the nominal financial flow.

Beyond the static general equilibrium effects, a more thorough understanding of aid's effect would include the dynamic impact of aid on relative growth rates of countries. A voluminous empirical literature estimating the impact of aid on growth has developed in recent years.[11] The regression results emerging from these studies oftentimes contradict each other, with no real consensus emerging at the present time; this makes it difficult to estimate with any degree of precision the impact of aid on growth of developing countries, and through growth, on the international distribution of income. However, these dynamic impacts are likely to be significant, and perhaps, much larger than the static impacts discussed here.[12] A solution to all these issues is left for further research. A purely static and partial equilibrium approach is adopted in what follows.

To evaluate the static impact that aid flows have on international income inequality, we conducted a simple exercise, treating grant aid flows and the grant element of concessional loans as *direct income transfers* from donor countries to developing countries. In doing so, we assume that all individuals within the country benefit equally from this transfer in the same way they were assumed to share equally the GNI. Since gross national income includes such transfers as an accounting concept, the exercise consisted of subtracting out aid flows received by the beneficiary countries and adding the value of those flows back to donors' GNI. The OECD/DAC database allowed us to compile a dataset of different types of aid flows by donor-recipient pairs.

The most straightforward type of aid is bilateral grants, which are given by a donor country to a recipient country with no expectations of repayment.[13] To these flows, we added the grant elements of concessional lending, received by developing countries from bilateral donors. To take into account bilateral contributions to multilateral agencies as well as multilateral agencies' distribution of aid to developing countries, we used the data from OECD/DAC on the imputed multilateral contribution.[14] Since multilateral agencies have a wide range of lending concessionality, we performed several sensitivity tests with respect to the grant element of the imputed multilateral component of the aid flows, assuming that this grant element was, in turn, 50/65/80 percent. The total amount of grant-equivalent aid in current U.S. dollars in 2002 used for the simulation is 53.6/56.2/58.8 billion, depending on the grant element of imputed multilateral contribution.[15]

One important issue that arises in this exercise is whether aid flows should be subtracted from countries' gross national incomes, which are expressed in PPP dollars,

in dollar terms or in PPP dollars too. If one converts the aid flows into PPP dollars, then the redistribution of aid is no longer zero sum. This may be appropriate to the extent that donor grants can buy more goods and services in the recipient countries if these grants are used to purchase a typical consumption basket of tradable and nontradable goods and if general equilibrium effects on the real exchange rate are ignored. If, on the other hand, grants were spent only on tradable goods, then the figures in straight dollar terms would more accurately account for the impact of grants on recipient income per capita and no change would take place in the real exchange rate. Since we cannot know if aid income is spent only on tradables or on a typical consumption basket of tradables and non-tradables, we calculated the static impact of aid in both dollar and PPP terms.[16] The first estimate is more justified if most aid is directly spent on imports or if it is spent on a combination of traded and non-traded goods but the supply of non-traded goods is inelastic. I present both sets of estimates below. The real welfare effect is likely to stand between these two extremes.

Not surprisingly, the impact of aid is practically negligible when aid is accounted for in dollars and GNI is expressed in PPP dollars. It can be seen in figure 7 that under this assumption, aid produces a change in the poorest decile's welfare equal to 3 percent and negligible proportions for intermediate deciles. Of course, it represents a loss for the top deciles approximately equal to the share of ODA in the GNI of donor countries, that is approximately .25 percent in 2002.[17] As a consequence, international

FIGURE 7.
Per Capita PPP GNI
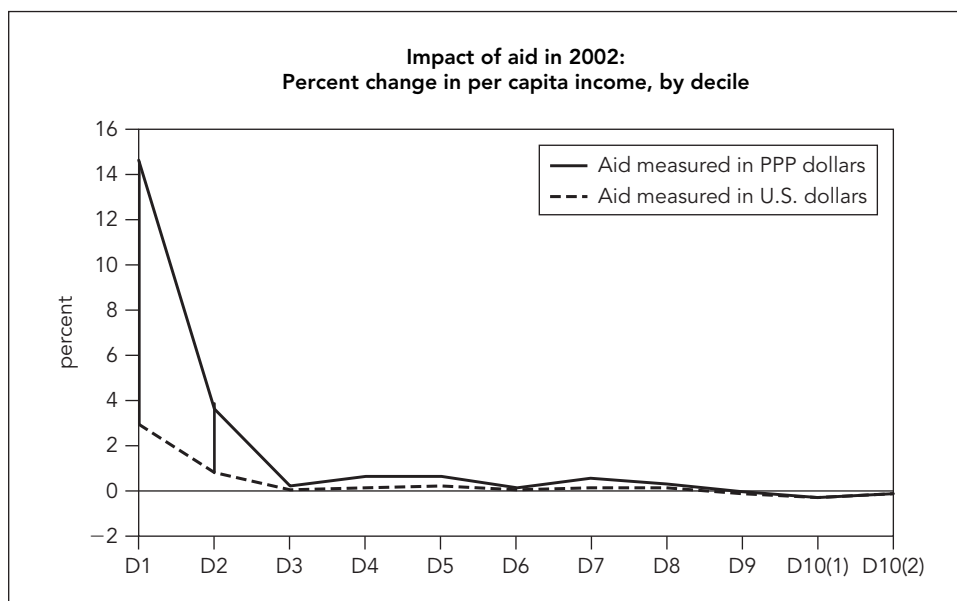
Source: Author's calculations based on data from *World Development Indicators* 2004 and aid flow data from OECD/DAC.

inequality indices are almost insignificantly affected by the international redistribution through aid. For instance, the Gini coefficient would be .539 without aid, whereas it is .538 with it.

The situation is different when aid is evaluated in PPP terms. Then the aid GNI ratio in the poorest decile of the world population is 15 percent, and 4 percent in the second poorest decile. Aid is now making a difference in international inequality. The Gini coefficient would be .42 percentage points higher without aid and poverty sensitive inequality measures like the Atkinson measure with inequality aversion coefficient e = 5 are still more significantly affected.

The preceding result should not be overemphasized, however. First, as mentioned above, due to the assumptions implied in using aid in PPP, this calculation is probably close to the upper bound of a range of estimates, while the lower bound corresponds to close to a negligible international redistribution effect of aid. Second, even the upper estimate is dwarfed when compared to the extent of redistribution that may take place in single countries that are equipped with effective redistribution schemes. Thus, the international redistribution that is taking place in the world through aid appears limited both in absolute terms and in comparison with what may be observed in single high-income and even middle-income countries.[18]

Of course, the preceding conclusion in no way means that aid is ineffective. That it is very effectively targeted towards the poorest means that international poverty measures or inequality measures close to the Rawlsian criterion that emphasizes the relative welfare of the poorest are likely to be much more sensitive to aid.[19] Even so, however, the extent of redistribution through ODA appears limited when compared to what is observed in single countries. There, it is not uncommon to see the income of the poorest decile rise by 50 percent or more because of redistribution instead of a figure somewhere between 3 and 14 percent with ODA.

Redistribution through Trade Protection

To analyze the impact of existing trade protection on international inequality, we used the results from the World Bank Linkage Model prepared for *Global Economic Prospects 2004*.[20] This model calculates the long-term steady-state gains from complete liberalization of trade in goods around the world, simulated to take full effect by 2010. These gains are expressed as the difference in GDP, PPP in 2015 relative to the baseline scenario of no additional trade liberalization. In this exercise, we are discounting these future gains to 2002, the baseline year. Thus, when I refer to "lost potential income from existing trade protection" or the "benefit from future trade liberalization," I am implying the same steady-state gains calculated by the model.

The model works with 23 "country units," some of which are individual countries and others are regional groupings of countries. To remain consistent with the initial framework, it has been assumed that all countries within a country group are affected in the same way by trade liberalization. By comparison with the benchmark

situation, these simulations inform us on the impact of existing protection. The model provides results for the static changes in GDP evaluated in PPP terms and changes in real household income; here, I concentrate on changes in GDP, PPP, since it is closer in definition to our indicator—GNI in PPP. I applied these changes to PPP GNI—assuming that GNI follows the same evolution as GDP—using 2001 as the base year, similar to the Linkage model.

Two points are worth noticing in this exercise. First, the issue of counting changes in income in dollars or in PPP dollars does not arise anymore because the computable general equilibrium Linkage model delivers directly the change in the welfare of a representative citizen of one of the 23 country groups in the model. Second, trade liberalization is increasing the efficiency of the world economy, so that the redistribution that it entails is not a zero-sum game. In effect, the world real GDP increases by approximately 1 percent as a result of this policy.

In terms of inequality, the redistributive effects of protection in world trade are somewhat larger, in absolute value, than those of aid. The gini coefficient of the international distribution would be lower by .5 percentage points in the absence of protection, from .531 to .526. However, the poverty sensitive measures, like the Atkinson index with inequality aversion e = 5, show much less variation, suggesting that the redistribution that would take place through liberalization is not exclusively toward the poorest in the world.

Figure 8 presents the impact of trade liberalization in goods across deciles. As anticipated on the basis of inequality measures, we can see in figure 8 that low- and

FIGURE 8.
Impact of Trade Reform
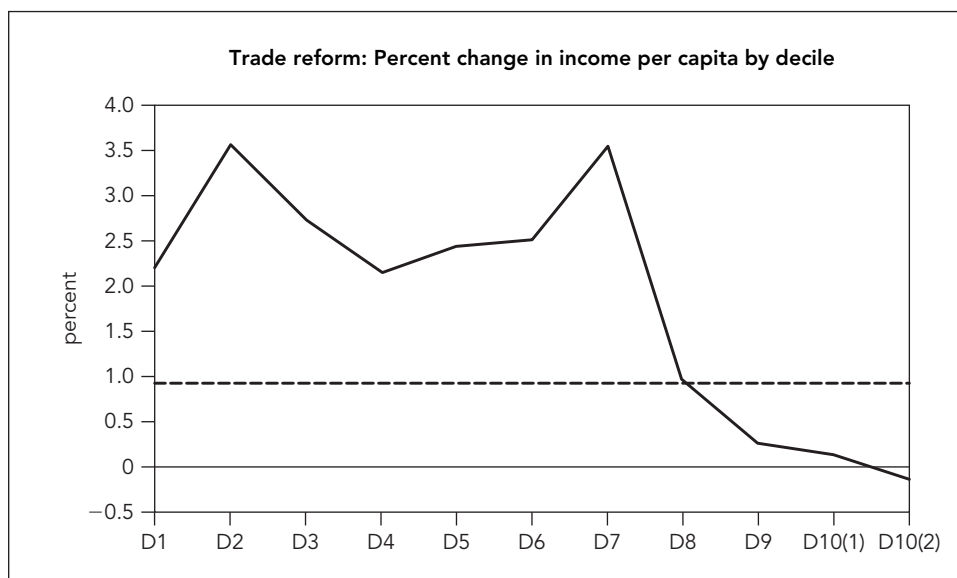
Note: No productivity enhancement. Change in GDP per capita from model. Dotted line represents aggregate world increase.

middle-income deciles benefit the most from merchandise trade reform, while the top vintile loses out not only with respect to the average gain in the world economy but also in absolute terms. This is essentially due to the richest countries' size in the world economy. The protection of their markets contributes to improving their terms of trade. Eliminating all kinds of protection has a negative impact on their welfare, which is not compensated by the benefit they draw from the drop in the protection of developing countries.

It is interesting to bring together the impacts of the two international sources of international redistribution just discussed, aid and trade protection. To do so, we compare the redistribution effected via existing aid levels versus the lost income caused by the existing barriers to trade. In this way, we have a rough approximation of the impact of two international policies (aid and trade) on global distribution of income. In the figure below, the lost potential income due to trade barriers (which is another way to interpret figure 8) is subtracted from the effect due to aid (presented in figure 7). For the bottom decile, part of the gains from aid is negated by the detrimental impact of trade barriers via lost potential income, but the overall effect remains positive. And for other deciles, the lost trade income cancels out and surpasses the income transfers received as aid.

As a consequence, the overall effect of these two policies on inequality is essentially ambiguous. Comparing to the 1 percent average drop in world welfare, it appears as if intermediate deciles were redistributing to *both* the poorest and the richest deciles. Put in another way, one can also say that it is as if rich countries,

FIGURE 9.
Impact of International Policy Mix

Net impact of aid and trade policies
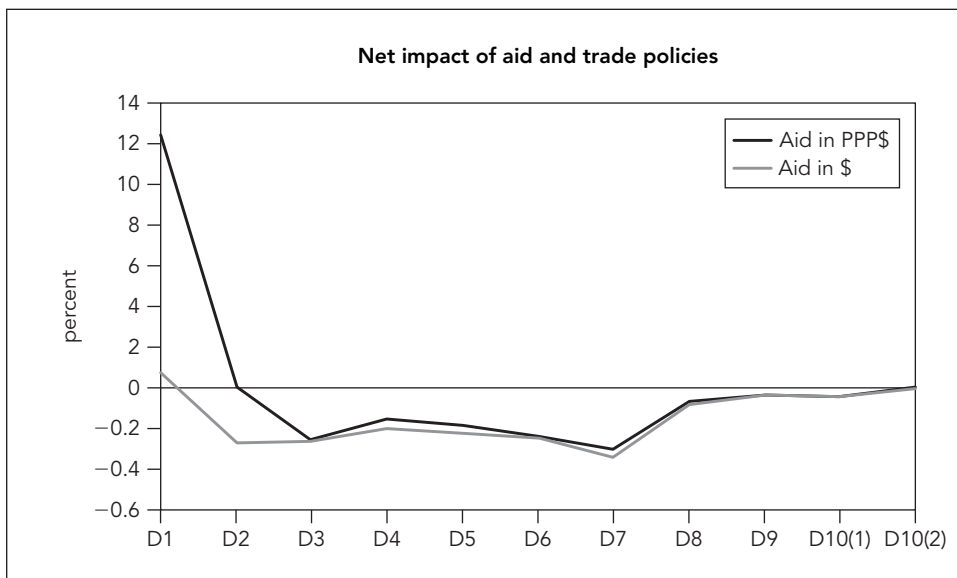
Note: Trade numbers are based on simulations of different country groupings, so the two figures are not strictly comparable. Scenario without productivity increases is used for the above figure.

which bear the burden of aid targeted toward the poorest deciles of the world distribution were essentially recovering it through extracting protection revenues from intermediate deciles.

To sum up, the analysis of the redistribution that takes place through aid and through protection leads to both positive and negative conclusions. The positive conclusion is that the poorest one or two deciles of the international distribution see their welfare increased, although in a limited way and with some ambiguity for the second decile, since this depends on whether aid is accounted for in dollars or in PPP. The negative conclusion is that intermediate deciles are losing welfare due to the protectionist policies implemented both within their borders and in rich countries. Overall, the richest vintile of the international distribution—the richest countries—are neither losers nor gainers in this redistribution game. They essentially gain through protection what they lose through aid.

Conclusions

The current level of global inequality of income is extremely high. If this level of inequality were to exist within a single country, that country would probably experience substantial social strife. A renewed emphasis on increased redistribution from aid and renewed policy reforms to assure that aid enhances growth prospects for low-income countries are both urgently needed.

The historical trends studied in the paper reveal three stylized facts. First, there is a clear decrease in inequality in population-weighted terms over the period, mostly driven by the impact of China and India. Second, there is a clear divergence in income per capita treating countries equally; however, this trend starts to level off at the end of the period. Third, there is a group of low-income countries that are clearly being left behind. The favorable trend in population-weighted terms will start to lose momentum as India and China approach middle- to even upper-middle-income status over the next decade. Hence, there is a need for renewed growth in the countries that now inhabit the bottom decile of the inter-country distribution of income. So the need for greater redistribution does not negate the fact that growth performance will be critical.

Aid flows are small globally but can be large locally. They have a small impact on aggregate indices, but more impact is revealed when one examines the impact on a decile-by-decile basis. One discouraging result is that much of the favorable impact of aid is decimated by the negative impact of trade restrictions on potential income in the developing world. In brief, there is a contradiction in the set of international policies on aid and trade, where the benevolent hand of aid is countered by the malevolent hand of protectionism.

The analysis presented here should be considered a first approximation. It ignores, first of all, the important impact of policies on the within-country distribution of income and how this affects the "true" world distribution of income. Within this caveat, there is also a distributional impact of aid within the recipient country that is

also ignored. Perhaps most importantly, this analysis only takes a preliminary look at some of the general equilibrium effects (in the case of trade). In addition, this analysis treats redistribution of aid and trade as a static one-off phenomenon. Clearly there are dynamic effects of the flow of aid and the factors of production on the distribution of income over time. More work should be done in these areas.

In the increasingly globalized economy, we are moving toward a more global society where the well-being and actions of citizens on different sides of the planet impact one another. Compared with national systems of redistribution and social protection, global systems are woefully inadequate. It is time to repair the contradictions in international policies that I discussed here today. In addition, the development community needs to scale up development assistance to resolve once and for all the tragedy of global poverty.

Notes

1. Based on ongoing research with Victoria Levin (World Bank) and David Rosenblatt (World Bank).
2. See, for example, Baumol 1986, Barro and Sala-i-Martin 1992, Sheehey 1996, Jones 1997, Quah 1996 and 1997.
3. Ram 1989, Theil and Seale 1994, Firebaugh 1999.
4. Berry et al. 1983, Bourguignon and Morrisson 2002, Dikhanov and Ward 2002, Milanovic 2002, Schultz 1998, and Sala-i-Martin 2002.
5. Jones 1997, Pritchett 1996 and 1997, De Long 1988.
6. Bourguignon and Morrison 2002.
7. In population-weighted terms, only part of China was in the bottom decile in 1980. Here we are referring to the countries that replaced that part of China in the bottom decile.
8. Pritchett 1997.
9. Ravallion 2004. On the other hand, absolute poverty rates have declined, and a variety of studies have confirmed this trend. See World Bank 2004b.
10. For aid, see Yano and Nugent 1999.
11. Collier and Dollar 2002; Burnside and Dollar 2000 and 2004; Hansen and Tarp 2000 and 2001; Clemens et al. 2004.
12. Knowledge may be one of the most important economic flows impacting the world redistribution of income; however, this flow is particularly difficult to model, and as such, it is beyond the scope of the exploratory analysis of this address.
13. For each of the flows (bilateral grants, bilateral loans and imputed multilateral contributions), I have distributed "unspecified" and "unallocated" aid (listed as a regional aggregate) back to the region's members, using a pro rata calculation based on specified donor-recipient flows.
14. OECD/DAC calculates the imputed multilateral contribution by distributing each DAC donor country's contribution to multilateral agencies to recipient countries, using pro rata calculation of the percentage the multilateral agency gives in grants or concessional loans to each particular recipient. These figures are then added across multilateral agencies that receive contributions from this particular DAC donor country to arrive at a donor-recipient "imputed multilateral" contribution.

15. To make these data comparable to the base of actual GNI per capita, we deflated these aid volumes to 1995 dollars, using the same procedure used for deflating GNI in Section II.

16. To convert aid in dollars to aid in PPP terms, we multiply the former by a ratio of GDP, PPP to GDP in dollars, all taken from *World Development Indicators* 2004.

17. Note that there is little difference between PPP and dollar values for the rich countries.

18. Please note that this assumes the highest of the three scenarios for concessionality of multilateral lending—0.8. With lower multilateral lending concessionality (0.50 or 0.65), the change in inequality is even smaller.

19. I conducted a similar analysis for 1985. The major difference in the results was that less aid was distributed to the bottom deciles and more aid to the middle deciles.

20. I am thankful to Dominique van der Mensbrugghe for providing the results and advising me on this exercise.

References

Barro, Robert J., and Xavier Sala-i-Martin. 1992. "Convergence." *Journal of Political Economy* 100 (2): 223–51.

Baumol, William J. 1986. "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show." *American Economic Review* 76 (5): 1155–59.

Berry, Albert, François Bourguignon, and Christian Morrison. 1983. "Changes in the World Distribution of Income between 1950 and 1977." *The Economic Journal* 93 (37): 331–350.

Bourguignon, François, and Christian Morrisson. 2002. "Inequality Among World Citizens: 1890–1992." *American Economic Review* 92 (4): 727–744.

Burnside, Craig, and David Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90 (4): 847–68.

Burnside, Craig, and David Dollar. 2004. "Aid, Policies, and Growth: Revisiting the Evidence." World Bank Policy Research Working Paper No. 2834. World Bank, Washington, DC.

Clemens, Michael, Steven Radelet, and Rikhil Bhavnani. 2004. "Counting Chickens When They Hatch: The Short Term Effect of Aid on Growth." CGD Working Paper 44.

Collier, Paul, and David Dollar. 2002. "Aid Allocation and Poverty Reduction," *European Economic Review* 46 (8): 1475–1500.

Development Assistance Committee. 2004. *International Development Statistics*. Paris: OECD. Statistical online database.

Development Co-Operation Directorate/Development Assistance Committee. 2002. *DAC Statistical Reporting Directives*. Paris: OECD.

De Long, J. Bradford. 1988. "Productivity Growth, Convergence, and Welfare: Comment." *American Economic Review* 78 (5): 1138–54.

Dikhanov, Yuri, and Michael Ward. 2002. "Evolution of the Global Distribution of Income in 1970–99." Paper presented at the 5th Conference on Globalization, Growth, and (In)Equality. Warwick, England. March 15–17.

Firebaugh, Glenn. 1999. "Empirics of World Income Inequality." *American Journal of Sociology* 104 (6): 1597–1630.

Hansen, Henrik, and Finn Tarp. 2000. "Aid Effectiveness Disputed." *Journal of International Development* 12 (3): 375–98.

Hansen, Henrik, and Finn Tarp. 2001. "Aid and Growth Regressions." *Journal of Development Economics* 64 (2): 547–70.

Jones, Charles I. 1997. "On the Evolution of the World Income Distribution." *Journal of Economic Perspectives* 11 (3): 19–36.

Milanovic, Branko. 2002. "True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone." *Economic Journal* 112 (476): 51–92.

Pritchett, Lant. 1996. "Forget Convergence: Divergence Past, Present, and Future." *Finance and Development* 33 (2).

Pritchett, Lant. 1997. "Divergence, Big Time." *The Journal of Economic Perspectives 11* (3): 3–17.

Quah, Danny T. 1996. "Twin Peaks: Growth and Convergence in Models of Distribution Dynamics." *Economic Journal* 106 (437): 1045–55.

Quah, Danny T. 1997. "Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs." *Journal of Economic Growth* 2 (1): 27–59.

Ram, Rati. 1989. "Level of Development and Income Inequality: An Extension of Kuznets-Hypothesis to the World-Economy." *Kyklos* 42 (1): 73–88.

Ravallion, Martin. 2004. "Competing Concepts of Inequality in the Globalization Debate." Paper prepared for the Brookings Trade Forum. May 13–14.

Sala-i-Martin, Xavier. 2002. "The World Distribution of Income (estimated from Individual Country Distributions)." NBER Working Paper 8933.

Schultz, T. Paul. 1998. "Inequality in the Distribution of Personal Income in the World: How It Is Changing and Why." *Journal of Population Economics* 11 (3): 307–344.

Sheehey, Edmund J. 1996. "The Growing Gap between Rich and Poor Countries: A Proposed Explanation." *World Development* 24 (8): 1379–84.

Theil, Henri, and James L. Seale, Jr. 1994. "The Geographic Distribution of World Income, 1950–1990." *De Economist* 142 (4): 387–419.

World Bank. 2004a. *Global Economic Prospects*. Washington DC: The World Bank.

World Bank. 2004b. *World Development Indicators*. Washington DC: The World Bank.

Yano, Makoto, and Jeffrey B. Nugent. 1999. "Aid, nontraded goods, and the transfer paradox in small countries." *American Economic Review* 89 (3): 431–449.

Keynote Address
Markets, Capital Markets, and Globalization

VERNON L. SMITH

Manufacturers, gradually shift their places, leaving those countries and provinces which they have already enriched, and flying to others, whither they are allured by the cheapness of provisions and labour; till they have enriched these also, and are again banished by the same causes.

—*David Hume,* "Of Money" (1777)

My message today is about exchange and markets, without which people cannot engage in task and knowledge specialization. It is this specialization that is the secret of all wealth creation. There is no other source of sustainable human betterment. We all function simultaneously in more than one world of exchange. Those worlds overlap, as we live first in a world of personal exchange—trading favors and friendship and building reputations based on trust and trustworthiness in small groups and families; and second, in a world of impersonal exchange through markets—where communication and cooperation gradually emerge in trade with strangers.

I will discuss two kinds of markets—first, markets for commodities and services, which we have studied extensively in the laboratory. Then I will apply the learning from experiments to interpret the nature and function of such markets in the world. These markets are the foundation of wealth creation.

Second, I will talk about markets for capital, or stock markets. Capital and stock markets are far more volatile and unpredictable than are existing commodity and service markets, but this is to be expected, as their function is to anticipate the commodities and services of the future.

I also want to discuss globalization—the worldwide extension of markets—which is really nothing more than a new word for an ancient process of migration and

Vernon L. Smith is Professor of economics at George Mason University in Fairfax, Virginia, and President of the International Foundation for Research in Experimental Economics. He shared the 2002 Nobel Prize in Economics with Daniel Kahneman.

human betterment-seeking that began a long time ago when our common ancestors first walked out of Africa.

Personal Exchange, Reciprocity, and Early Peoples

There are good reasons to support the belief that exchange has its origin in reciprocity and sharing norms within the family, the extended family, and tribes. This personal exchange allowed task specialization between the hunting, gathering, and toolmaking that laid the basis for enhanced productivity and welfare, and in turn enabled early peoples to migrate all over the world. Thus, specialization supported by personal exchange allowed globalization to begin early—long before the emergence of markets.

Since the pioneering work of anthropologist Franz Boas over a century ago, the study of extant hunter-gatherer tribal societies makes plain the antiquity, sophistication, and diversity of property rights in the human career. Of the hundreds of examples that could be cited, I want to quote one of my favorites, from Peter Freuchen's *Book of the Eskimos* (1960), on the "first harpoon" principle as it applied to polar bear hunting teams.

> According to custom, all the hunters present are to get parts in the quarry, in this case both of the meat and skin. There are three pairs of trousers in a bearskin. If there are more than three hunters present, the ones who threw their spears last will usually be generous enough to leave their parts of the skin to the others. The hunter who fixed his spear first in the bear gets the upper part. That is the finest part, for it includes the forelegs with the long mane hairs that are so much desired to border women's kamiks (boots) with. (pp. 53–54)

Note that the Inuit "first harpoon" norm is an incentive rule that rewards, with a greater value, the greater risk and cost of being the first to harpoon this incredibly dangerous prey. It is an equal opportunity rule—not an equal outcome rule—that evolved from ancient prehistory. Any member of the hunting team is free to go first, pay the risk cost, and collect the higher revenue. All others, however, *whose contributions cannot be differentiated,* share equally or more flexibly in the remaining revenue.

Laboratory Evidence

These deep ethical principles surface in laboratory experiments showing that when there is no way to differentiate individual contributions, people support the equal outcome rule. When contributions can be differentiated, people tend to prefer a rule that rewards in proportion to individual contributions—more to those who sacrifice more for the group.

I and my coauthors have studied this ancient norm of reciprocity and trading favors as it emerges in two-person games between anonymously matched subjects, many of whom use trust and trustworthiness to achieve gains from the exchange: cooperative outcomes that maximize joint benefits. Many voluntarily avoid choosing

outcomes that take for themselves without giving something in return to their matched counterpart.

From Personal to Impersonal Exchange and Markets

When the same subjects who consciously cooperate for betterment in elementary two-person interactions come to the laboratory to trade in impersonal experimental markets, what do they do? They strive to maximize their own gain, and in this process maximize the joint benefits of the group without intending to. However, these experimental markets are supported by externally enforced property right rules that prevent taking without giving in return.

The Extended, or Transient, Order of Markets: Achieving Unintended Ends

In established commodity and service markets producers incur recurrent, relatively predictable costs, and consumers experience corresponding recurrent flows of value from consumption. But costs and values are inherently private, and all such information is dispersed. Command and control economies have failed because such information cannot be given to any one mind. But how do we know that the price discovery process in commodity markets yields efficient surplus-maximizing outcomes? We have discovered in controlled laboratory experiments that these recurrent flow markets are incredibly efficient, and these findings have been replicated many hundreds of times. Moreover, the subjects in these experiments are not aware of the group welfare maximizing ends that they produce. Each, in pursuing his personal gain, achieves group maximizing benefits that are not part of his intention.

In debriefings at the end of a market experiment:

1) People deny that any kind of model could predict their final trading prices and exchange volume, yet these outcomes correspond to the demand and supply schedules of values and costs that are dispersed privately among all the participants; market efficiency does not require large numbers, complete information, economic understanding or sophistication, a fact that should not be too surprising since people were trading in markets long before there were economists.

2) People in these market experiments also believe that it ought to have been possible for them to have earned more profit for themselves. Yet in fact they were in equilibrium, each doing as well as was possible given the behavior of all the other participants.

From the Laboratory to the Economy: The Function of Markets and the Power of Diversity

What have we learned about markets? And what is the unseen function that they serve?

Commodity and service markets are the foundation of existing wealth creation. Each of us earns our income from no more than one or two sources. Yet think of all

the hundreds of items that we use or consume throughout the day that are produced by others whom we do not know, and will never know.

The hallmark of commodity and service markets is *diversity:* diversity of tastes, human skill and knowledge, natural resources, soil and climate. These in turn are what account for the differences in the values and costs that we use to define and motivate gains from exchange in the laboratory; that is, we create a controlled economy that exhibits this diversity.

The power of diversity to be extended and to serve human betterment depends vitally on exchange: personal exchange, in our intimate groupings; and impersonal exchange, through markets.

The Origins and Characteristics of Diversity

Initially, diversity was possible and encouraged through sharing and reciprocity norms in the family, the extended family, and tribes. Thus, in stateless hunter-gatherer societies, the women and children gathered fruits, nuts, tubers, and grains; men hunted; and old men advised in the hunt, fashioned tools and weapons, and helped in gathering.

At many times and many places in prehistory, exchange was extended to strangers through barter, and ultimately the use of commodity money.

Indeed, early humans set the stage for a vast expansion of wealth and well-being whenever a tribe discovered that it was better to trade with their neighboring tribes than to kill them. If you kill them, they can't produce something and trade it with you tomorrow—nor can you benefit from their unique skills, learning, art, culture, and experience. Similarly, if you let them live but steal from them they are much less willing to produce more for you tomorrow than if you trade with them today.

Diversity requires freedom, because it is freedom that allows each of us to be as different as we are able and desire to become. Markets in turn support tolerance of freedom. Chile had little political freedom but opened its economy to freer choice, and this spread to political choice and brought democracy.

Similarly, the Chinese government has moved to liberalize its economy over the last several years. Moreover, early in 2004 the constitution was modified to permit individuals to own private property, although the provision added the "catch-22" criterion that the property "be legally acquired." Although this provision disappointed those pushing for liberalization, the move did recognize the operating norm in which people do in fact have private property that is exchanged. The legal modification was motivated by the need to prevent corruption that could occur with payments made to government officials in return for permitting these property trades to occur.

Diversity without the freedom to exchange implies poverty: No human, however abundantly endowed with a single skill or a single resource, can prosper without trade. Daniel Defoe's Robinson Crusoe owned an Island, but he was poor.

We have need of others and the diversity they bring to the table, if we are to rise above bare subsistence. Through markets we depend on others, whom we do not

know, recognize, or understand. We know not how and in what ways others contribute to our welfare, and we contribute to theirs as well. Such are the long, subtle, and largely invisible chains of interdependence through markets connected by prices. The welfare of each of us depends vitally upon the knowledge and skills of others with whom we trade through markets.

Without markets we would indeed be poor, miserable, brutish, and ignorant; if some were less poor, it would be because of conquest, theft, and taking without giving in return, which can be sustained only for as long as there are others to conquer.

Markets and Property Rights Institutions

Diversity is made possible, productive, and permissive of wealth creation through market institutions.

Markets require consensual enforcement of the rules of social and economic exchange. No one has said it better than philosopher David Hume (1739) over 250 years ago, when he said that there are just *three laws of human nature:*

1) the right of possession,

2) its transference by consent, and

3) the performance of promises.

These are the ultimate foundations of order, with or without formal law, that make possible markets and prosperity.

Hume is talking about discovered or emergent law, not made or legislated law. The early "lawgivers" did not make the law they presumed to "give": they studied social traditions, norms, and informal rules—and gave voice to them, as God's, or natural law.

Hume's Laws of Nature are derived from the ancient Judeo commandments: *Thou shalt not steal; thou shalt not covet thy neighbors' possessions; thou shalt not bear false witness.* These principles are evident in other cultural norms as well.

- The game of "steal" consumes wealth without encouraging its reproduction, while the game of "trade" sustains and grows abundance.

- Coveting the possessions of others invites an involuntary state-enforced redistribution of the gains from specialization and trade, endangering incentives to produce tomorrow's harvest—perhaps as surely as its theft.

- To bear false witness is to undermine community, management credibility, investor trust and confidence, long-term profitability, and the personal social exchanges that are most humanizing. Corporate management bears false witness at its peril. Once management loses credibility with shareholders the stock market is unforgiving, as evidenced with the Enron Corporation in the United States, whose stock plummeted when management self-dealing became known.

Capital Markets

This brings me to the topic of stock markets, which are inherently far more uncertain than markets for commodities and services because stock markets must anticipate innovations—the new commodities and services of the *future*. At the time that new innovations emerge, the extent of their subsequent economic success is inherently unpredictable.

Innovations in History, and Stock Markets

If changing knowledge and technologies are to yield new commodities and services, they require capital. Capital markets allow the users and suppliers of capital to be distinct and more specialized; the savers (or suppliers) do not also have to be the entrepreneurs (or users) that can grow new wealth from capital investment. And both can gain by exchanging investment for a share of the return; both also bear the risk of loss.

Stock market bubbles and crashes are not new. Why?

Great stock market booms are fueled by new technologies.

For example, in the 19th century the steam engine allowed the steam ship to replace the square-rigger sailing ship, and the railroad to replace the mule team and stagecoach. Railroad expansion in 19th-century America outran the shipping needs of interregional trade.

Profitability turned to losses, bankruptcies, and consolidations. But out of this 19th-century expansion, long-term value was created and retained for the entire U.S. economy.

At the turn of the 20th century many new technologies emerged. Telephone, electricity, petroleum, and automobiles sustained a wave of investment and development. There was overexpansion in response to high profitability followed by declining margins, losses, bankruptcy, and consolidation; but long-run value was created and not lost to the economy. Bankruptcy allows the assets of failed managers—human and physical—to be reallocated to managers who launch a new attempt at making the business a success.

My hometown of Wichita, Kansas, had 15 airplane manufacturers in 1929: Travel Air, Stearman, Cessna, United, Laird, Swift, Lark, Knoll, Bradley, Yunkers, Wichita, Watkins, Mooney, Sullivan, and Buckley. Two of them were new company names in 1927: Cessna and Stearman. A decade later Stearman's successful operation had been bought by Boeing, and the general aviation survivors were Cessna and Beech, who made Wichita the national center of this new industry. Being start-up companies just two years before the great crash had not prevented success for Cessna and Beech, but the other manufacturers all failed and their salvageable assets were reallocated.

The ballpoint pen is today an almost invisible but classic example of innovation and change: This mid-20th-century invention initially sold for $10. It was a very profitable new product, attracting a rush of entry, falling prices, losses and consolidations,

but the pen stays—yielding continuing long-term value of which we are not aware, except that we are all a little richer as a result. Today when we buy BIC pens for 60 cents we are unaware that these pens are far superior to those $10 upstarts produced more than a half-century ago.

The 1990's Bubble Crash: History Repeats

Picking winners and identifying losers in advance is inherently risky if impossible, except by chance. More than 60 percent of U.S. manufacturing firms leave the industry in their first 5 years; 80 percent in their first 10 years. (Dunne, Roberts, and Samuelson 1988; these data are from the years 1967 to 1982). The 1990s brought an unprecedented volume of initial public offerings (IPOs). I am sure that the history of that decade will record not only an unprecedented failure rate, but also, and much less visible, an unprecedented increase in long-term economic value for the world economy. The current bubble and crash is fueled by new electronic, communication, computer, biological, and pharmacological technologies. The residual long-term value is suggested by the postcrash national income data: output has increased with little increase in employment. We get more for less as manufacturing rapidly goes the way of agriculture.

Diesel truck engines are an example of long-term value created in the old economy by companies that have become severely stressed if not bankrupt. Each cylinder is today controlled by computer, to minimize fuel consumption and meet tough new emissions standards under all operating conditions. Small electronic companies such as Apex Microtech (with 50 employees) developed and sold the power-optimial amplifiers used by Navistar to create the new engines.

It is painful for those who risk investment in new technologies and lose, but the benefits captured by other industries and by the learning and consolidations that leave value for the few winners are retained as new wealth for the economy. This is the substance of growth, betterment, and the ultimate reduction of poverty. This is why almost everybody is wealthier than were their parents and grandparents.

How can the individual pain of investment loss be eliminated, and the long-term value achieved, with a policy fix that avoids the risk of doing more harm than good? We don't know. On this as on many matters there are just two classes of people: those who know they do not know, and those that don't. If someone tells you they can identify new company stock market winners and losers, keep your hands in your pockets.

Here is the problem: If you limit people's decisions to make risky investments in an attempt to keep them from harming themselves, how much will that reduce our capacity to achieve major technological advancements? The hope of great gain by individuals fuels thousands of experiments in an environment of great uncertainty as to which experiments, and which combinations of management and technology, will be successful. The failure of many may be a crucial part of the cost of sorting out the few that will succeed. After a wave of innovation and a bubble bursts, managers know a lot about what did not work, and a little about what did work.

Globalization: A New Word for the Age-Old Process of Migration and Market Expansion

Globalization for us began when our common Cro-Magnon ancestors walked out of Africa about 50,000 years ago: settling Asia and Australia 50 to 40,000 years ago; Europe by 40,000 years ago; Siberia and the Arctic 35 to 20,000 years ago; the Bering gateway to North America 14 to 13,000 years ago; the Americas 12 to 11,000 years ago; and New Zealand and Madagascar only 1,000 years ago. Long before the square-rigger sailing ship, our ancestors had settled every continent except Antarctica, and all the major islands.

Archeology and ethnographic studies suggest that early peoples, long before nation states, traded tools, weapons, and public goods such as symbols, customs, crests, and unmolested rights of access to trade routes and hunting grounds.

Expansion

The first long-distance trade between Europe and the Near East allowed us to escape the static, dead-end, and poverty-ridden Middle Ages. This led to new explorations by stock companies and nation states.

Technology-Driven Change

Exploration was driven by a new technology: the great square-rigger sailing ships.

As commerce spanned the Old and New Worlds, there was a worldwide exchange of plant and fruit products. The Italian gourmet had not a single tomato until the plant was imported from the New World; nor did the Irish have the potato until one of the thousands of varieties was imported from South America. The diversity of nature was the basis of specialization and much wealth creation through exchange. Instead of cutting-edge research and development we had exploration, transportation, and transplantation.

During the 19th century the seas were spanned by steamships, the continents by the steam locomotive railroads.

Whole regions now began to specialize in different industrial and agricultural products, depending on their natural comparative advantages. The diversified subsistence farm reformed into the cattle ranch; the wheat, barley, corn, and rice farms; the milk farm; and the chicken farm.

The recent global Green Revolution stemmed from the new seed varieties developed by Norman Borlaug, who won the Nobel Peace prize in 1970. His work doubled and tripled the yields of wheat, maize, rice, and other crops—enabling China, India, Mexico, and Pakistan to feed their greatly increased populations with only a 1 percent increase in world land cultivation.

The latest great thrust in globalization is driven by innovations in computing power, communications, and transportation logistics. All three serve Internet exchange.

The retail store was once the place where buyers met producers through the intermediary of the merchant who risked the purchase of inventories of what he

hoped people would want to buy, and the buyer risked the quality of the goods produced.

This very high-cost way of matching consumers with producers has been challenged by the Internet, where buyers and sellers are matched at near zero cost, and new institutions are being created for direct shipment and for quality assurance through competition in reputation formation, warranties, and liberal return policies. This new dream world of potential profit led to overexpansion as investors threw investment funds at all the retail "dotcoms," just as their ancestors a century earlier had thrown money at the railroads.

New World of Competition in National Policies

The current wave of globalization is bringing a new discipline to national governments. Budget and monetary excesses by national governments discourage foreign investment, while encouraging domestic nationals to seek foreign, more stable, investment opportunities.

South American countries can better serve their people by asking how they can learn from Chile to bite the bullet, stabilize currencies, control government spending, privatize government-owned industries, and reduce barriers to free trade.

In particular there is the need, emphasized by development theoretician Hernando de Soto, for institutional change that enables the owners of real estate and other assets to hold clear, fungible titles. In this way the use value of existing assets in poor countries can be reflected in exchange value, allowing entrepreneurial access to capital markets to facilitate internal development.

Not only capital but also people move to where there is opportunity; and this is an essential part of creating new wealth and prosperity. This was dramatically illustrated for me in 1978, during a taxicab trip from the Wellington, New Zealand, airport to my hotel. The driver was friendly, and I asked, "Tell me about your country." He replied, "It's really wonderful. I don't like paying half my small income in taxes, but we receive so much that is free: health benefits, prescriptions, free education through college and advanced graduate study. I am just a cab driver, but my son is going to be a medical doctor. He has finished his medical degree and internship, and will begin practicing next year." In recognition of his obvious pride, I said, "How wonderful. You have every right to be proud. Is he going to practice in Wellington?" He replied, "Oh no, he's going to Australia. You can't make any money here." The New Zealand economic crisis hit about two years later. New Zealand could not compete in world markets—and could no longer afford socialism.

Conclusion

Commodity and service markets are the foundation of wealth creation. The fact that stock markets serve by supplying capital for new consumer products explains why they are inherently uncertain, unpredictable, volatile, and—given investor behavior— why they tend to bubble and crash. Stock markets are far more uncertain than

markets for commodities and services because stock markets must anticipate innovations—the new commodities and services of the *future*. Globalization is not new. It is a modern word describing an ancient human movement—a word for humankind's search for betterment, and the worldwide expansion of resource specialization, which is determined by the extent of market development.

I think globalization is a good word—a peaceful word. In the wise pronouncement commonly attributed to the great French economist Frédérich Bastiate:

If goods don't cross borders, soldiers will.

References

Dunne, T., M. J. Roberts, and Larry Samuelson. 1988. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries." *Rand Journal of Economics* 19: 495–515.

Freuchen, Peter. 1960. *Book of the Eskimos.* Cleveland: World Publishing.

Hume, David. 1739. *A Treatise of Human Nature by David Hume,* Book III, Part II, Sections III, IV, and V. Repr. in three volumes and edited, with an analytical index, by L. A. Selby-Bigge. Oxford: Clarendon Press, 1896.

———. 1777. "Of Money." *Essays Moral, Political and Literary.* Rev. ed., edited by E. F. Miller. Indianapolis: Liberty Fund, 1987.

Behavioral Economics

Development Economics through the Lens of Psychology

SENDHIL MULLAINATHAN

Economists conceptualize a world populated by calculating, unemotional maximizers. This view shapes our understanding of many crucial elements of development economics—from how rural villagers save, to how parents decide on whether to send their children to school.

Psychological research, however, has documented the incompleteness of this perspective. Individuals have self-control and time inconsistency problems. They can give in to short-run temptations and later regret it. They can have strong feelings about others that drive them to commit both generous and spiteful acts. They often passively accept defaults rather than make active choices. They let the institutions around them make choices for them. And they may misread new data in a way that fits their beliefs. In short, the rational maximization model may not be a very good approximation of human behavior.

In this paper, I present some of the psychological evidence that I believe helps us to better understand a few core issues in development economics, such as savings, education, and property rights. This gives us new ways to interpret a variety of behaviors in these contexts, and enriches the set of policy tools we should consider. This evidence not only suggests the need for dramatically new tools, but also suggests small cost changes that may dramatically improve their efficacy of existing policies.

Economists often study scarcity. Yet their conception of decisionmaking assumes an abundance of psychological resources. In the standard economic model people are unbounded in their ability to think through problems. Regardless of complexity, they can costlessly figure out the optimal choice. They are unbounded in their self-control. They implement and follow through on whatever plans they set out for themselves.

Whether they want to save a certain amount of money each year or finish a paper on time, they face no internal barriers in accomplishing these goals. They are unbounded in their attention. They think through every problem that comes their way and make a deliberate decision about each one. In this and many other ways, the economic model of human behavior ignores the bounds on choices (Mullainathan and Thaler 2001). Every decision is thoroughly contemplated, perfectly calculated, and easily executed.

A growing body of research interprets economic phenomena with a more modest view of human behavior. In this alternative conception, individuals are bounded in all of these dimensions, and more. In practice, this conception begins with the rich understanding of human behavior that experimental psychologists have developed through lab and field experiments. This view, ironically enough, emphasizes the richness of behavior that arises from scarcities, emphasizing the bounds on cognitive and computation ability, self-control, attention, and self-interest. Theoretical models are now being constructed that help to incorporate these ideas into economic applications. Perhaps even more compelling is the recent empirical work that suggests the importance of these psychological insights for real behavior in contexts that economists care about. In a variety of areas, from asset pricing, to savings behavior, to legal decisionmaking, well-crafted empirical studies are challenging the traditional view of decisionmaking.

This paper attempts to provide an overview of this research to those interested in development economics. I have chosen psychological insights that I believe are helpful in understanding several phenomena in development economics: parents' schooling decisions, savings behavior, choice of financial institutions, bureaucratic corruption, and property rights. For each of these I describe a small piece of the psychology that may be potentially relevant. In this way, I hope to introduce readers to the psychological and associated field evidence and show the practical relevance of this evidence. Given the space considerations, my goals are modest. I am clearly not comprehensive in my review of the relevant areas of psychology; that would take a book at the least. Nor am I comprehensive in describing the various psychological insights that may help in understanding any one topic (savings). As stated earlier, my goal is instead to present only an overview of each topic.

Two important caveats are in order. First, there are many reasons to believe that the psychological factors discussed here may be unimportant in economic contexts. Some could argue that the experiments are "weak" because the people studied are not financially motivated. Others might argue that market competition or arbitrage would guarantee that these "irrational" choices should have no impact on economic outcomes. Yet others might argue that learning would remove these problems. I will not address these objections because they have been dealt with at great length elsewhere.[1] I am more pragmatic in my approach. I do not believe that any set of lab experiments alone can ever provide a firm basis for policy. Even the best experimental evidence will face questions of context specificity, behavioral adaptation, and equilibrium. Instead, these experiments are wonderful because they inspire different perspectives on old problems—and new ideas for economic policy. Their ultimate success, however,

depends on how the experiments fare when tested in the field. So the evidence I provide here is merely to inspire (and not substitute for) careful tests in relevant contexts. The experimental evidence, therefore, need only pass a lower hurdle: Is the bulk of the evidence sound enough to merit future empirical work or policy experimentation? The accumulated evidence, I feel, easily passes this hurdle.

Second, my attempts to incorporate psychology into development should not be confused with pejorative attempts to label the poor as "irrational." This is neither an attempt to blame the poor for their poverty nor to argue that the poor have specific irrationalities. Instead, my goal is to understand how problems in development might be driven by general psychological principles that operate for both poor and rich alike. When I speak of self-control, for example, I am speaking of self-control problems that exist in equal measure around the world. These problems may matter more for the poor because of the context in which they live, but the core of these problems is a common one (Bertrand, Mullainathan, and Shafir 2004).

Immediate Barriers to Education

The rational choice model of schooling is straightforward (Becker 1993). Individuals trade off the costs and benefits of schooling to decide how much schooling to pursue. Benefits come in a variety of forms, such as better jobs or better marriage prospects. Costs could be direct financial costs (fees) as well as any opportunity costs (foregone labor). In the case of children, of course, parents make the actual choices. They do so to maximize some combination of their own and their children's long-run welfare, with the exact weight given to choices dependent on their altruism.

This view of education abstracts from the richness of the hardships faced by parents trying to educate their children in a developing country. Consider a poor father in a village who is eager to send his son to school during the next school year. He recognizes the value of education to his son, which will allow him to get a government job, marry better, or simply exist more comfortably in a rapidly changing world. To ensure that he has money for school fees, textbooks, or perhaps a school uniform, the father begins to save early. But he soon encounters competing demands on the money. His mother falls ill and needs money to buy some analgesics to ease her pain. Though his mother insists that her grandson's education is more important, the father is torn. Enormous willpower is required to let his mother suffer while he continues to save money that he knows could ease her pain. Knowing that he is doing what is best in the long run is small consolation in the moment. The father overcomes this struggle and enrolls his son in school. But after some weeks, his son starts to show disinterest. As for most children everywhere, the son finds that sitting in a classroom (and an unpleasant one at that) is not very appealing, especially since some of his friends are outside playing. Exhausted from tiring physical work and feeling the stresses of everyday life, how will the father handle this extra stress? Will he have the mental energy to convince his son of the value of education? Will he have the energy to follow up with the teacher or other students to see if his son has actually been attending school?

This fictional example merely illustrates one important tension; and even the best of intentions may be very hard to implement in practice, especially in the high-stress settings that the poor inhabit.

Family problems of this type are intimately related to how people view tradeoffs over time, a topic that psychologists and behavioral economists have studied extensively through experiments. I now describe a variety of related evidence and then return to how this evidence may help us to understand the schooling decision.

Would you like to receive $15 today, or $16 in one month? More generally, how much money would I need to give you in one month to make you indifferent to receiving $15 today? What about in one year, or in 10 years? Thaler (1981) presented these questions to subjects and found median answers of $20, $50, and $100. While at first glance these answers may seem somewhat reasonable, they actually imply huge discount rates: 345 percent over one month, 120 percent over a one-year horizon and 19 percent over a 10-year horizon.[2] Subjects most often greatly prefer the present to the future.

These choices also imply that the rate of time preferences *changes* with the horizon. This is made most clear in the following choice problem:

Would you prefer $100 today, or $110 tomorrow?

Would you prefer $100 30 days from now, or $110 31 days from now?

Many subjects give conflicting answers to these two questions. To questions such as the first one they often prefer the immediate reward ($100 today). To questions such as the second one they often prefer the delayed reward ($110 in 31 days).

Such preferences are inconsistent with the standard model. To see this, suppose people discount the future at rate δ. Then the value of $100 today is *u(100)* and its value tomorrow is *u(110)*. On the other hand, in problem two the value is $\delta^{30}u(100)$ versus $\delta^{31}u(110)$. This is the exact same tradeoff. In other words, with the standard constant discounting individuals should choose the same thing in both situations.

Differences in preferences for the immediate versus the future can also be seen in the field. Read, Loewenstein, and Kalyanaraman (1999) asked subjects to choose three rental movies. The subjects either chose one by one, for immediate consumption. Or they chose all at once, for the future. When choosing sequentially for immediate consumption, they tend to pick "low-brow" movies. When picking simultaneously for future consumption, the subjects tend to pick "high-brow" movies. Once again, when planning for the future they are more willing to make choices that have long-run benefits (presumably "high-brow" movies) than when choosing in the present.

The difference in choices at different horizons poses a problem for the individual. Consider a concrete example. Suppose my preference is that next Monday I will begin writing a paper rather than put that off until Tuesday. Of course, today I am busy and would rather put off writing the paper. What happens on Monday? What had been a decision about the distant future (where I exhibited patience) becomes a decision about the present (where I exhibit impatience). My choice may now change. Once again, the option of putting it off for a day seems appealing, as appealing as it did last week when

I made the same decision. In other words, there is a conflict between what I plan to do in the future and what I will actually do when the future arrives.

This type of conflict is only one of the difficulties parents face in getting their children educated. In the example I gave, the father wanted his son to be educated and was willing in the future to put in the effort and money needed to see that happen. Yet in the moment, many immediate pressures impinge on his time, money, and energy, making it hard for him to implement his longer-term plan. This view presumes that parents would like to see their children educated but simply can't find a credible way to stick with that plan. I think this perspective helps improve our understanding of many components of education.

It provides explanation of the gap between parents' stated goals and actual outcomes. The PROBE Report on basic education in India finds that many parents are actually quite interested in education (De and Dreze 1999, pp. 19–26). Even in the poorest states in India, where education is worst, this survey found that over 85 percent of the parents agreed that it was important for children to be educated. In the same survey, 57 percent of parents responded that their sons should study "as far as possible." Another 39 percent of parents said their children should get at least a grade 10 or grade 12 education. Clearly parents in these areas of India value education. Yet these responses contrast with very low educational attainment in these states. This gap is reminiscent of the gap between desired and actual retirement savings in the United States. In one survey 76 percent of Americans believed that they should be saving more for retirement. In fact, 55 percent felt they were behind in their savings, and only 6 percent reported being ahead (Farkas and Johnson 1997). They want to save, but many never make it happen. As noted earlier, immediate pressures are even more powerful in the education context. Putting aside money to pay for schooling requires making costly, immediate sacrifices. Fighting with children who are reluctant to go to school can be especially draining when there are so many other pressures. Walking a young child to a distant school every day requires constant effort in the face of so many pressing tasks. Or stated differently, if middle-class Americans supported by so many institutions cannot save as much as they want, how can Rajasthani parents be expected to consistently and stoically make all the costly, immediate sacrifices needed to implement their goal of educating their children?

This also helps to explain, in part, an interesting phenomenon in many developing countries: sporadic school attendance. In contrast to a simple human capital model, education does not appear to follow a fixed stopping rule, with students attending school consistently until a particular grade. Instead, students go to school for some stretch of time, drop out, and later begin again. This sporadic attendance, though far from optimal, is a characteristic of the dynamically inconsistent preferences described earlier. When faced with particularly hard-to-resist immediate pressures, individuals will succumb to them. When these pressures ease, it becomes easier to implement the original plan of sending their child to school—and they may revert to it. In many related discussions of self-control, the importance of salience is often emphasized (Akerlof 1991). To this end, parents who have "slipped off the wagon" may find some salient moments that encourage them to again try to get their

children to school. One empirical prediction here is that at the beginning of the school year attendance should perhaps be higher than at any other time, as many parents decide to give it another try. As the parents succumb to immediate pressures, attendance would then decline throughout the year.[3]

This perspective also has some policy insights. First, policies that spread immediate pressures over time could be beneficial. For example, school fees that require continuous small payments rather than one large payment may make it easier for parents to finance schooling. It requires far more will power to save up for a big purchase (such as uniforms) than to pay small fees each week or month.[4] Second, this perspective should alter policies that attempt to increase parental demand for education. For example, the success of bonus payments to parents for children's enrollment depends crucially on the payment structure. If payments are made at the end of the school year, they are unlikely to work particularly well. In this model, parents already recognize a long-run reward to education. Adding to that will do little to solve the core problem. In contrast, bonus payments that are made more frequently may help to tilt the tradeoff in the short-run, which is the real barrier. Third, programs that make schooling more attractive to students may provide a low-cost way to make it easier for parents to send children to school. For example, a school meals program may make school attendance attractive to children and ease the pressure on parents to constantly encourage their children to go to school (see Vermeesch 2003 for a discussion of such programs). One could even be creative in designing these programs. For example, school sports, candy, or any number of other cheap inputs that make schooling more attractive to children may have large effects. In fact, under this model such programs could have extremely large benefit-to-cost ratios, much larger than could be justified by the monetary subsidy alone.

In my opinion, this perspective on schooling matches the complexity of life in developing countries. Of course, immediate pressures are not the only problem. Numerous other factors—from liquidity constraints to teacher attendance—surely play a role. Yet, those have been explored and are very much on the radar screen of many development economists. These other forces, while potentially powerful, are not commonly considered and deserve more scrutiny.

Demand for Commitment and Savings

The difficulty of sticking with a course of action in the presence of immediate pressures also has implications for how individuals save. But in the standard economic model of savings, there is no room for such pressures. In that model people instead calculate how much money will be worth to them in the future by taking into account any difficulties they may have in borrowing, and any shocks they may suffer. Based on these calculations, they make a contingent plan of how much to spend in each possible state. They then, as already discussed, implement this plan with no difficulty. As noted earlier, for poor people in many developing countries, implementing such plans is much easier said than done. They face a variety of temptations that might derail their consumption goals.

Behavioral economists have recently begun to better understand the devices that people may use to deal with such temptations. The inter-temporal preferences noted earlier (short-run impatience, long-run patience) are often modeled as discount rates that vary with horizon. People have a very high discount rate for short horizons (decisions about now versus the future) but a very low one for distant horizons. This is often called hyperbolic discounting because the original curve used to produce it was hyperbolic in shape (Strotz 1956, Ainslie1992, Laibson 1997).

A key question in this model is whether people are sophisticated or naive in how they deal with their temporal inconsistency. Sophisticated people would recognize the inconsistency and (recursively) form dynamically consistent plans. In other words, they would only make plans that they would follow through on. Naïve people, however, would not recognize the problem; they would make plans assuming that they will stick to them and abandon their plans only if required, when the time comes. There are reasons to believe both views. On the one hand, individuals appear to consciously demand commitment devices that help them commit to a particular path. On the other hand, they appear to have unrealistic plans. Perhaps the best fit of the evidence is that individuals partly (though not necessarily fully) recognize their time inconsistency.

The important practical feature of this view is that the commitment implicit in institutions is very important for understanding behavior. Institutions can help solve self-control problems by committing people to a particular path of behavior. A common analogy here is with Ulysses, who in Greek mythology ties himself to his ship's mast so that he can listen to the song of the sirens but not be lured out to sea by them. While not so dramatic, similar commitment devices exist in everyday life. Many refer to their gym membership as a commitment device. ("Being forced to pay that much money every month really gets me to go to the gym lest I waste the membership fee.") Or to take another example, Christmas clubs, though now less common than in the past, used to be a powerful commitment tool for some who wanted to save up to buy Christmas gifts.

Relevant evidence on the power of commitment devices is given in Gruber and Mullainathan (2002), which studies smoking behavior. Rational choice models of smoking treat this behavior roughly like any other good. Smokers make rational choices about their smoking, understanding the physiology of addiction that nicotine entails. Behavioral models, however, recognize a self-control problem in the decision to start smoking and in the decision (or rather attempts) to quit. Some survey evidence seems to support the behavioral model. Smokers often report that they would like to quit smoking but are unable to do so. This resembles the temporal pattern above. Looking into the future, smokers would choose to not smoke. But when the future arrives, they are unable to resist the lure of a cigarette today (perhaps by promising themselves that tomorrow they will quit). To differentiate these theories we examined the impact of cigarette taxes. Under the rational model, smokers are made worse off. This is a standard dead-weight loss argument. Smokers who would like to smoke cannot now, because of the higher price. In models with time hyperbolic discounters, however, taxes could make smokers better off. The very same force that is

bad in the rational model—high prices driving smokers to quit—is good in the behavioral model. Because smokers wanted to quit but were unable to, they are now better off. In the parlance of time-inconsistency models, the taxes serve as a commitment device.

To assess well-being we use self-reported happiness data. While such data are far from perfect, they can be especially useful in contexts such as these, where the variable of interest is relatively clean and the mismeasurement is thus simply absorbed in the residual. Using a panel of states in the United States, we find that the happiness of those who tend to smoke increases when cigarette taxes increase. Relative to the equivalent people in other states (and relative to those who tend not to smoke in their own state), these people show actual rises in self-reported well-being. In other words, contrary to the rational model and supportive of the behavioral model, cigarette taxes actually make those prone to smoke *better off*. This kind of effect is exactly the one I alluded to in the introduction: Institutions (or cigarette taxes in this case) have the potential to help solve problems within people as well as among people.

There is also evidence on people actively choosing commitment devices. Wertenbroch (1998) argues that people forego quantity discounts on goods they would be tempted to consume (cookies, for example) in order to avoid temptation. This is a quantification of the often-repeated advice to dieters: Don't keep big bags of cookies at home. If you must buy tempting foods, buy small amounts. Trope and Fischbach (2000) show how people strategically use penalties to spur unwanted actions. They examined people scheduled for small, unpleasant medical procedures—and showed how these people voluntarily chose to take on penalties for not undergoing the procedures. In fact, they cleverly chose these penalties by selecting higher penalties for more aversive procedures. Ariely and Wertenbroch (2002) provide even more direct evidence. They examined whether people use deadlines as a self-control device and whether such deadlines actually work. In an experiment, students in a class at MIT chose their own deadlines for when to submit three papers. The deadlines were binding, so in the absence of self-control problems the students should clearly choose the latest deadlines possible for all three papers. They were told there was neither benefit to an early deadline nor cost to a late one, so they can only benefit from the option value of being able to submit a paper later. In contrast, students chose evenly spaced deadlines for the three papers, presumably to give themselves incentives to complete the papers in a timely manner. Moreover, the deadlines appeared to work. A related study shows that people who are given evenly spaced deadlines do better than those who are given one big deadline at the end.

I think savings in developing countries can also be better understood through this perspective. It provides an alternative view on institutions such as roscas, which are popular in many countries (Gugerty 2001). In a rosca, a group of people meets together at regular intervals. At each meeting, members contribute a prespecified amount of money. The sum of those funds (the "pot" so to speak) is then given to one of the individuals. Eventually, each person in the rosca will get their turn and

thus get back their contributions. Roscas are immensely popular, but what is their attraction? They often pay no interest. In fact, given the potential for default (those who receive the pot early may not continue to pay in), contributors may effectively pay a negative interest rate. One reason for the popularity of roscas may be that they serve as a commitment device in several ways. By making savings a public act, individuals allow social pressure from other rosca members to commit them to their desired savings level (Ardener and Burman 1995). As some rosca participants say, "you can't save alone." Other rosca members have all the incentives to make sure each other member continues to contribute. The groups also enable individuals to save up to larger amounts than they normally could achieve given their own problems with self-control. Imagine someone who wished to make a durables purchase (or pay school fees) of 1,000 rupees. By saving alone and putting aside money each month, the saver faces a growing temptation. When they reach 400 rupees, might not some other purchase or immediate demand appear more attractive? The rosca doesn't allow this temptation to interfere. Individuals get either nothing, or the full 1,000 rupees all at once. This "all or nothing" property may make it easier for some to save enough funds to make large purchases.

This type of scheme also helps to provide a more nuanced view of individuals' demand for liquidity. In the standard logic, the poor unconditionally value liquidity. After all, liquidity allows people to be able to free up cash to attend to immediate needs that arise. If a child gets sick, money is needed to pay for medicine. This might be especially true for the poor. Shocks that are small for the well-off can be big for the poor, and they would need to dip into real savings to address them. But the poor in these models face a tradeoff. They value liquidity for the reasons cited above, but liquidity for them is also a curse: it allows them to too easily dip into savings. Durable goods and illiquid savings vehicles may actually be preferred to liquid savings vehicles. Cash, for example, may be far too tempting and spent too readily. On the other hand, by holding their wealth in items such as jewelry, livestock, and grain, individuals may effectively commit themselves not to give into immediate consumption pressures. In these models, therefore, there is an optimal amount of liquidity. Even when liquidity is provided at zero cost, the poor will choose some mix of illiquid and liquid assets.

Another implication from this perspective is that revealed preference fails as a measure of policy success. Observing that people borrow at a given rate (and pay it back) does not necessarily mean that the loan helps them. A loan may in some cases help them deal with a liquidity shock. But in other cases, it may not help, because the loan assists them in giving way to immediate temptations and leaves them straddled with debts they must repay. This distinction is important for understanding micro-credit in developing countries. Often, the metric of success for such programs is whether they are self-sustainable. Such a metric makes sense if revealed preference makes sense. Profitability would imply that people prefer getting these loans even at a non-subsidized rate; revealed preference then implies their social efficiency. Yet in the presence of time inconsistency, profitability of micro-credit could mean very little about social efficiency. The key question is to what extent the loans exaggerate short-run impatience and to what extent they solve long-run liquidity

constraints.[5] Ultimately one needs a deeper understanding of what drives borrowers. One avenue for this might be data on loan usage. Are loans being spent on long-run investments (as is often touted) or spent on short-run consumption? Of course, some short-run consumption might well be efficient, but this data combined with an understanding of the institution would help to better understand (and improve) the social efficiency of micro-credit.

Policy can also provide cheaper and more efficient commitment devices. After all, even saving in grain is an expensive way to produce a commitment device. Vermin may eat the grain, and the interest rate earned on the grain could be zero or even negative. Moreover, it is important to recognize that even if people demand such commitment devices, the free market may not do enough to provide them. The highly regulated financial markets in developing countries may lead to too little innovation on these dimensions. Monopoly power may also lead to inefficient provision of these commitment devices, depending on whether a monopolistic financial institution can extract more profits by catering to the desire for commitment or to the temptations themselves. In this context governments, nongovernmental organizations, and donor institutions can play a large role by promoting such commitment devices.

Ashraf, Karlan, and Yin (2004) provide a stunning illustration of this. They offered savers at a bank in the Philippines the opportunity to participate in "SEED" accounts, which are like deposit accounts, except that individuals cannot withdraw deposits at will. Instead, the money can be withdrawn only at a predetermined date, or once a predetermined goal has been reached. This account does not pay extra interest and is illiquid. In most economic models, people should turn down this offer in favor of the regular accounts offered by that bank. Yet there is strong demand for the SEED accounts. More than 30 people of those offered the accounts choose them, and banks report that the accounts help these particular individuals to save. Six months later, those offered the accounts show substantially greater savings rates than those not offered the accounts. Experiments such as these will, I feel, eventually help to deepen our understanding of savings decisions and greatly improve development policy.

Defaults and Financial Institutions

Financial institutions do not simply help savings through their commitment value. A very important set of results in behavioral economics suggests that these institutions affect behavior simply through the status quo they produce. Samuelson and Zeckhauser (1988) documented a variety of phenomena known as the status quo bias. Here is a simple example. A group of subjects was given the following choice:

> You are a serious reader of the financial pages but until recently have had few funds to invest. That is when you inherited a large sum of money from your great uncle. You are considering different portfolios. Your choices are:
>
> • Invest in moderate-risk Company A. Over a year's time, the stock has 0.5 chance of increasing 30 percent in value, a 0.2 chance of being unchanged, and a 0.3 chance of declining 20 percent in value.

- Invest in high-risk Company B. Over a year's time, the stock has 0.4 chance of doubling in value, a 0.3 chance of being unchanged, and a 0.3 chance of declining 40 percent in value.
- Invest in treasury bills. Over a year's time, these bills will yield a nearly certain return of 9 percent.
- Invest in municipal bonds. Over a year's time, these bonds will yield a tax-free return of 6 percent.

A second set of subjects is given the same choices, but with one small difference. These subjects are told that they are inheriting a portfolio from their uncle, in which most of the portfolio is invested in moderate-risk Company A. The choice now is subtly different. It is how much of the portfolio to *change* to the options above. Interestingly, the subjects find a large difference between the two treatments: much more of the money is reinvested in Company A when that is the status quo choice.

This bias toward the status quo appears to run quite deep and is not just due to superficial explanations (such as information content of the uncle's investments). Samuelson and Zeckhauser (1988) demonstrated this bias with a very interesting piece of evidence from the field. In the 1980s, Harvard University added several plans to its choice of health plans, thus providing an interesting test of status quo bias: How many of the old faculty chose the new plans, and how many of the newly joined faculty chose the older plan? A stark difference emerged. Existing employees "chose" the older plans at a two to four times higher rate than new employees. In other words, incumbent employees made the easiest choice of all: to do nothing.

This bias toward the status quo could perhaps be motivated by the deeper phenomena of automatic behavior. Psychologists have recently documented numerous instances of the idea that people often make automatic, nonconscious choices. Gilbert, Tafarodi, and Malone (1993) provided an example that illustrates automaticity. Subjects were exposed to false information about a criminal defendant. On some trials subjects were exposed to these false sentences while cognitively loaded with another task—or while under time pressure. In these conditions subjects automatically assumed the (false) statements to be true rather than examining them. This illustrates one of the basic ideas behind this research on automaticity. Unless attention is consciously drawn to a decision, it will be made through some automatic processes. In many practical situations, the likely automatic process is to simply do nothing. Thus, what economists view as a "choice" may not really be an active choice at all. It may instead reflect default behavior combined with the institution underlying that choice.

Madrian and Shea (2001) conducted a particularly telling study along these lines. They studied a firm that altered the choice context for employee participation in their retirement plan. When new employees join the firm, they are given a form that they must fill out in order to participate in the savings plan. Although the plan is quite lucrative, participation is low. Standard economic models might suggest that the subsidy ought to be raised, but this firm instead changed a simple feature of its program. Prior to the change, new employees received a form that said something to the effect of "Check this box if you would like to participate in a 401(k) plan. Indicate how much you'd like to contribute." After the change, however, new employees received

a form that said something to the effect of "Check this box if you would like *not* to have 3 percent of your pay check put into a 401(k) plan." By standard reasoning, this change should have little effect on contribution rates. How hard is it to check off a box? In practice, however, Madrian and Shea (2001) find a large effect. When the default option is to not contribute, only 38 percent of those who were queried contributed. When the default option was contribution, 86 percent contributed. Moreover, even several years later those who were exposed to a contribution default still showed much higher contribution rates.

These results are consistent with (and motivated) those discussed earlier. While we cannot be sure from these data what people are thinking, I would speculate that some combination of procrastination and passivity played a role. Surely many people looked at this form and thought, "I'll decide this later." But later never came. Perhaps the subjects were tempted by activities other than deciding on 401(k) contribution rates (hard to believe, but there are more interesting activities). Perhaps the decision simply slipped from their attention because other factors came to occupy it. In either case, whatever the default was on the form, a majority ended up with this choice. In fact, as other psychology tells us, as time went on these individuals may well have justified their "decision" to themselves by saying, "3 percent is what I wanted anyway," or "that 401(k) plan wasn't so attractive." In this way, their passivity made the decision for them. By making the small, active choice to choose later, these people ended up making a large decision about thousands of dollars in retirement money.

Insights of this type can also help us design whole new institutions. One example is Save More Tomorrow, a program created by Thaler and Benartzi (2003) in an effort to get people to make one active choice—but to have them make it in such a way that if they remain passive afterward, they are still saving. To participate in the program, contributors decide on a target savings level (and we know from before that people actually do want to save). Once they decide on how much they'd like to save, participants agree to small deductions from their paychecks beginning *next year*. And then each year, as they receive pay raises their deductions will increase until reaching their target savings level. Participants can opt out of the program at any time. But the cleverness of the program is that if the savers do nothing and remain passive, they will continue to save (and even increase their savings rate).

The results have been stunning. In one firm, for example, more than 75 percent of those offered the Save More Tomorrow plan participated rather than simply trying to save on their own. Of these, interestingly few of them (less than 20 percent) later opted out. As a result, savings rates increased sharply. By the third pay raise (as the default increases accumulated), individuals had more than tripled their savings rates. But perhaps the greatest success has been the diffusion of this product. Many major firms and pension fund providers are thinking of adopting the plan, and participation in the program will likely soon number in the millions. Save More Tomorrow is an excellent example of what psychologically smart institutional design might look like in the future. It does not solve a problem between people but instead helps solve a problem within people: not saving as much as they would like.[6]

One simple implication of these results is that behavior should not be confused with dispositions (Bertrand, Mullainathan, and Shafir 2004). An economist observing the savings behaviors of both a middle-class American and a rural farmer might be tempted to conclude something about different discount rates. The high savings of the middle-class American surely reflects greater patience. But as we have seen, this need not be the case. Such an inference could be just as wrong as inferring that those who defaulted into their 401(k) plans are more patient than those who did not participate by default. The behavioral difference may be that better institutions facilitate more automatic, default savings by individuals.

Another implication is in the form of banking reform. Some of the lessons learned in the United States could easily be transferred to parts of developing countries. First, protocols such as automatic payroll deposits (as well as the ability to reroute some of this money directly into savings accounts) could be a powerful way to spur savings. Banking innovations such as these could be very inexpensive yet have profound effects on the savings rates of the middle-class in developing countries.

Second, the simple extension of banking to rural areas could in and of itself have a large impact on behavior. While not as powerful a default as having your paycheck automatically deposited, it may very well help to have the money placed out of easy access. The worker then has to make one active decision—putting the money into the account—but then the act of keeping the money becomes a passive one. When money is close at hand, active effort is required to save it. But when money is in the bank account, active effort is required to go and get it in order to spend it. In this sense, a bank account may serve as a very weak commitment device. By keeping the money at a (slight) distance, spending it may be a lot less tempting.

Loss Aversion and Property Rights

Consider the following simple experiment. Half of the students in a room are given mugs, and the other half receive nothing (or a small cash payment roughly equivalent to the value of the mugs). The subjects are then placed in a simulated market where a mechanism determines an aggregate price at which the market clears. How many mugs should change hands? Efficiency dictates that market clearing should allocate the mugs to the 50 percent of the class who value it the most. Since the mugs were initially randomly assigned, roughly half of this group should have started off with mugs, and half should have started off with no mugs. Consequently, trading should have resulted in exactly half the mugs changing hands.

Kahneman, Knetsch, and Thaler (1990) have in fact run this experiment. Contrary to the simple prediction, however, they found a stunningly low number of transactions. Roughly 15 percent of the mugs trade hands. The prediction problem is seen if we look at how students value the mugs. Those who were given the mugs put a reservation price at three *times* that of those who did not receive mugs. Given that, it is no surprise that so few mugs change hands. Numerous follow-up experiments have been run on this so-called *endowment effect,* to rule out the obvious explanations: an

income effect, the value of mug recipients being able to see and feel the mug, or small transaction costs of some form. In the end, the phenomenon is robust. Those who are given objects very quickly appear to value them more than those who were not given the objects.

This phenomenon reflects in part a deeper fact about utility functions: prospect theory. In fact the original experiment was motivated by prospect theory. In prospect theory, people's utility functions are defined in large part on changes. In the traditional model of utility people would value the mug at $u(c+Mug) - u(c)$. That is, their utility is defined in absolute levels of consumption, and the mug adds to that. In the prospect theory approach, utility is defined by a value function that is evaluated locally and in changes. Those who receive the mug consider its loss as a function of $v(-Mug) - v(0)$. Those who do not receive the mug value its gain at $v(Mug) - v(0)$. Notice the symmetry in the original function: both those with and without the mug value it the same (on average). In the second formulation, however, nothing guarantees the symmetry. The difference in valuation between the two depends on whether $v(Mug)$ is bigger or smaller than $-v(-Mug)$. The evidence above is consistent with a variety of evidence from other contexts: losses are felt more sharply than equivalent gains. Thus $v(x) < -v(-x)$. This phenomenon, known as loss aversion, has been seen in many contexts. Perhaps the two cleanest examples are in Odean and Genesove and Mayer. Odean (1998) showed that small investors in the stock market are more willing to sell stocks they have made money on than ones they have lost money on. This fact may seem quite obvious, but it is inconsistent with standard utility theory (he rules out the obvious tax explanations) since gains and losses are symmetric: Investors should merely take the trades they view as best. In fact, Odean finds that this strategy of holding losers and selling winners results in negative abnormal returns. An investor's unwillingness to take on losses, on the other hand, is quite consistent with loss aversion. Another example, familiar to many who have owned housing, is given in Genesove and Mayer (2001), who found that individuals who have taken a loss on their house set far higher prices when it comes time to sell. It appears that they are more willing to gamble to break even, a phenomenon quite consistent with loss aversion.

The insight about loss aversion can also help in understanding why policy change is so difficult in developing countries. Consider market reforms that transfer resources from one group to another with an efficiency gain. For example, suppose privatizing a firm will result in gains for customers while resulting in losses for incumbent workers. Under this perspective, such reforms are fought so vigorously partly because the losses are felt far more sharply by the workers. One implication of loss aversion is, at the margin, to pursue strategies that preserve the rents of incumbents rather than ones that try to buy out incumbents. All other things equal, a strategy that offers a buyout for incumbent workers will be far more costly than one that grandfathers them in. The buyout requires the government to compensate the workers for their loss, and this can be much greater than simple utility calculations suggest. In contrast, a strategy that guarantees incumbent workers a measure of job security would not need to pay this cost.[7] Many situations of institutional change require some form of redistribution. The

recognition of loss aversion suggests that successful policies may require protecting the losses of incumbents.

Loss aversion also reinforces the importance of well-enforced property rights. Consider a situation where there is a single good, such as a piece of land L. Suppose that there are two individuals (A and B) who can engage in force to acquire or protect the land, and that engaging in violence may result in acquisition. In the presence of well-defined property rights (say this land belongs to person A), the decision to engage in force is straightforward. If B engages in force he stands to gain $v(L)$ if his force is successful. A, on the other hand, stands to lose $v(-L)$ if he doesn't engage in force. In this case loss aversion implies that A stands to lose a lot more than B could gain. So with well-defined property rights A would engage in more force than B. Consequently, B may never attempt force. So even in the absence of enforcement, loss aversion may mean that well-defined property rights may deter violence.

Consider now the case of ill-defined property rights. Suppose that both interested parties are unsure who owns a piece of land. Specifically, take the case where they both think they own it. This is an approximation to the situation where ownership with probability one-half already gives a partial endowment effect, or to the situation below of biased beliefs, where both parties may have probability greater than one-half of owning it. In this case, both A and B think they stand to lose $v(-L)$ if they do not fight for the land. In other words, in the absence of well-defined property rights, both parties will put in large amounts of resources to secure what they already believe is theirs. This to me is one of the powerful implications of loss aversion. Appropriately defining property rights prevents two (or more) parties from having an endowment effect on the same object. Conflicting endowments such as this are sure to produce costly attempts at protecting the perceived endowments, and anything ranging from costly territorial activities (fencing and de-fencing) all the way to violence may result.

Social Preferences and Teacher Motivation

In many important development contexts, self-interested behavior is extremely deleterious. Bureaucrats in many countries are corrupt. They enforce regulations sporadically, or take bribes. Another stark example is teacher absenteeism. Numerous studies have found that teacher absenteeism is one of the primary problems of education in developing countries. Teachers simply do not show up for school, and as a result little education can take place. This blatantly selfish behavior stands in contrast to some evidence on social preferences—that individuals may value the utility of others. I will review this literature and describe how social preferences may contribute to the problem but may also serve as part of the solution.

A very simple game called the "ultimatum game" has become an excellent tool for studying social preferences (Güth, Schmittberger, and Schwarze 1982, Thaler 1988). In this game, one player (the "proposer") makes the first move and offers a split of a certain amount, say $10. The second player ("responder") decides whether to

accept or reject this split. If it is accepted, P and R get the proposed split. If the split is rejected, then both players get zero. What makes this game so intriguing is that it clarifies two interesting issues in interpersonal preferences. First, will the responder accept "unfair" offers? In the pure self-interest model the responder should accept any offer greater than zero and be indifferent to even an offer of zero. Second, what kind of offer will the proposer make given the responder's rejection strategy? Is the proposer motivated only by the threat of rejection? In the pure self-interest model he would, of course, offer the responder a tiny bit above zero (or even zero itself) knowing that there's no fear of rejection.

This game has been run in many countries, for stakes that range from a few dollars in the United States to the equivalent of a few months' income in many countries. Yet the pattern of findings is relatively constant.[8] First, responders often reject unfair offers (i.e., those other than 50-50 splits). Second, proposers often make very fair offers, for splits close to 50-50 or 60-40. Moreover, proposers' fair offers are not just driven by fear of rejection. They tend to make offers larger than implied by a simple (risk-neutral) fear of rejection. This is most directly seen in a variant of the ultimatum game, called the "dictator game." Here the proposer makes an "offer" but the responder has no choice but to accept it. In this game, the threat of rejection is removed and one continues to find non-zero offers by the proposer, although the offers are lower than in the ultimatum game.

The ultimatum game illustrates two facts about interpersonal preferences. First, both it and the dictator game suggest (rather prosaically) that people care about others. These are one-shot games with no chance for repetition. Yet people give away rents to others. Such "altruistic" preferences are used to a limited extent in economics (often within a family or perhaps a village). Yet here we see these behaviors as pretty universal. This is, of course, to most people not much of a surprise. The large amount of charitable giving that occurs in most societies, the volunteer activity, and the spending of private time on public goods (recycling, for example) all point to such preferences.

Reciprocity often underpins such preferences, as illustrated in a very nice experiment by Regan (1971). Subjects in this study were asked to rate the quality of some painting along with another person (who is actually a "confederate," or someone who worked for the researcher). Partway through the experiment, during the rest period, the confederate leaves the room. When he returns he has a Coca-Cola for himself, and has also brought one for the subject. In a control condition, the confederate merely leaves the room and comes back (with no Coke for himself or for the subject). So some subjects receive an unsolicited act of kindness, while others do not. At the end of the experiment, as they are parting ways, the confederate mentions to the subject that he's selling raffle tickets and that he'll win a prize if he sells more tickets than anyone else. "Could you help me and buy some tickets?" he asks the subject. This is the outcome of interest in this experiment: How many tickets does the subject buy? Relative to the control condition, the subject buys far more tickets if the confederate has made the small, unsolicited favor of buying the subject a Coke. In fact, so big is the effect that the return on the favor is quite large. The confederate bought a 10-cent can of Coke and ended up selling at least two more raffle tickets at 25 cents each. Consequently, for

a 10-cent "investment" he yielded 50 cents.[9] Such reciprocal fairness is ubiquitous. Survey firms use it by paying people *prior* to filling out their survey because they realize that the norm of reciprocity binds individuals to return the form. Nonprofits send small "gifts" along with their request for donations. The reciprocity norm is one specific and ubiquitous form of altruistic preferences.

Another very important wrinkle to the altruism perspective is provided by experiments in helping behavior. Darley and Latané (1968), for example, conducted a study at Columbia University, where subjects believed they were in a roundtable, virtual conversation. The subjects were seated in a room with a mike and speakers and were told that the conversation was with either one other person, or with six other people, and that the conversation would go in turns, with only one person's mike functioning at any given time. Partway through the "conversation," the subject hears the speaker go through a seizure of some sort and requests help from the experimenter. When the subjects feel they are the only other listener, most (though surprisingly not all) seek help. When they feel there are other listeners, hardly any seek help. Experiments such as this underscore the potential fragility of pro-social behavior: It is by no means universal, and is importantly shaped by context.

Yet the second outcome, rejection by the responder, points to an equally important fact about interpersonal preferences. People will pay costs themselves in order to punish those they feel are being unfair.[10] By rejecting an offer, the responder is passing up money to punish the proposer. This type of behavior illustrates part of the "dark side" of interpersonal preferences. In simple altruistic models, interpersonal preferences are only a good thing: Having one person care in a positive way about another only makes it easier to deal with externalities and so on. The responder's behavior shows, however, that inefficiencies and conflicts might arise.

This possibility is clearest in a classic experiment by Messick and Sentis (1979), who asked subjects to imagine they had completed a job with a partner. The subjects were asked to decide what they considered "fair" pay for their work, but were then divided into two groups. One group was told to imagine that they had worked 7 hours on the task, while the partner had worked 10. The other group is told to imagine that they had worked 10 hours, while the partner had worked 7. Both groups were told that the person who had worked 7 hours had been paid $25 and were asked what the person who had worked 10 hours should be paid. Those who were told that they had worked 7 hours (and paid $25) tended to feel that the 10-hour subject should be paid $30.29. Those who were told that they had worked 10 hours, however, felt they should be paid $35.24. The source of bias in these responses can be seen in the bimodality of the distribution of perceived "fair" wages. One mode was at equal pay ($25 for both), while the other mode was at equal hourly wage (so the 10-hour worker gets paid approximately $35.70). Interestingly, the difference between the two treatments was mainly in the proportion in each mode. Those who had worked 7 hours showed more subjects at the equal pay level mode, while those who had been told they'd worked 10 hours showed more subjects at the equal hourly pay mode. In other words, both groups recognized two compelling norms: equal pay for equal work, and equal pay for equal output. Yet their roles determined (in part) which of the norms they chose.

These results extend beyond choosing between two fairness norms. Such conflicts could easily arise even if there's disagreement about measuring input levels (which often are not fully observed), and they speak to the source of a problem created by fairness. When there is not universal agreement about the fair division of labor or pay, "fairness" preferences can very quickly create conflict.

These experiments as a whole illustrate the complexity of social preferences. Individuals in some contexts do much to help others (at great costs to themselves). Reciprocity in particular appears to be a powerful force. But people will also, at cost to themselves, punish those who they think are being "unfair." The final behavior is especially important since notions of fairness are often driven by self-interest.

Let us return to the case of teacher absenteeism. The PROBE report (De and Dreze 1999) details the results of an extensive survey of teachers in many areas of India. The report, which noted high absenteeism levels, includes comments from many interviews with teachers that are illuminating with regard to their attitudes. For example, it notes

> Having said this, the main issue may not be the low initial motivation of teachers as the fact that many of them lose their motivation over time. Indeed, among recently appointed teachers we often met people with genuine enthusiasm. The honeymoon, however, is usually short-lived, as the morale of young teachers is battered day after day. (pp. 57–58)

Much of this psychological battering can be viewed as a perceived failure of reciprocity. As noted earlier, individuals strongly adhere to the norm of reciprocity. Failures of reciprocity (or perceived failures) can result in punitive or self-interested behavior in response. Teachers may feel a strong social preference early on and be motivated to teach and give much more than they need to. After all, from a pure self-interest motive, they know they can get away with very little teaching. Yet they may be initially motivated to do more, to come to school, to struggle with tougher students, and so on. The teachers may view these contributions as a "gift." One reason for this, of course, is the initial framing of the job (as a "plum job, with good salaries, secure employment, and plenty of time for other activities"). Thus, a young teacher may think, "I am giving a lot to the school." As with any giving, however, the teacher may expect strong reciprocity and see (perhaps in a self-interested way) many outcomes as a lack of reciprocity. For example, the PROBE report notes that:

> The most common complaint is that schools are under-equipped, under-funded, under-staffed, and over-crowded. Poor infrastructural facilities were mentioned by 63 percent of teachers as one of the problems they face. (p. 58)

So teachers may feel that the government is not reciprocating their "gifts." This may be especially exaggerated by the transfer system in India:

> Unwanted postings and arbitrary transfers are seen as a constant threat. Teachers spend a great deal of time and energy trying to avoid undesirable transfers, lobbying for preferred postings, and building up influential connections to play the transfer game. (p. 60)

Thus both the benign neglect of schooling and the active transfers could easily drive teachers to feel that the government does not reciprocate their efforts. They may

also come to feel similarly vis-a-vis the students' parents:

> Teachers are often frustrated by the apathy of parents towards their children's educa-
> tion. They complain that parents do not send their children to school regularly, or with-
> draw them for flimsy reasons. They also see much foot-dragging even when children are
> at school: parents send them late and in tattered clothes, try to dodge the fees, and gen-
> erally fail to watch their children's needs and progress. As teacher[s] perceive it, their
> own efforts to keep the children at school are not reciprocated by the parents. (p. 65)

Thus, even teachers who are at first motivated may soon feel justified in their
apathy. They gave it their best and think that their efforts were not reciprocated.
Are these inferences justified? Perhaps not. As in the Messick and Sentis (1979) study,
teachers may very well make such inferences in a self-interested way. The failure of the
context may be in allowing teachers to make such biased attributions of fairness.
Alternatively, teachers may very well be justified in these attributions. We simply
cannot tell.

In either case, this perspective suggests that the problem of teacher attendance can-
not be studied in isolation. Policies that affect school resources or student attendance
may have a large, indirect effect on teacher attendance. More realistically, the impact
of teacher incentive policies may vary dramatically with the context. In a context of
limited resources where attendance is low, these policies may have only a small or
moderate impact. On the other hand, if teacher incentives are coupled with other
policies to increase both resources as a whole and student attendance, the impact
might be much larger. The teachers would then no longer feel self-justified for their
absence, and the incentives needed to get them to work may be much smaller.[11] Of
course, I suspect that the effects might be greatest for the new teachers. Among exist-
ing teachers, it is harder to tell whether they will anchor on past non-reciprocity or
adapt to the new context. While other factors clearly play a role in driving teacher
absenteeism, a deeper understanding of their social preferences will, I think, also help
to solve the problem.

Norms and Inequality

In 1937, Sherif conducted an interesting psychophysics test. The subjects were seated
in a totally dark room facing a pinpoint of light some distance from them. After
some time when nothing happens, the light appears to "move" and then disappear.
Shortly thereafter, a new point of light appears. It too moves after some time and
then disappears. Interestingly, this movement of the light is a pure psychophysical
phenomenon known as the autokinetic effect. The light does not actually move; the
eye merely makes it appear to move. The subjects were put in this context for
repeated trials (many different resets of the light) and asked to estimate how far the
light had "moved." When the lights were shown to individual subjects, these esti-
mates were variable, ranging from an inch to several feet. However, an interesting
pattern developed when subjects performed this task in groups of two or three. Under
these conditions, the subjects' estimates invariably began to converge on a particular

number. A group norm quickly developed. In one variant, a member of the group was a confederate (someone who worked for the experimenter) who gave a specific number. The subject quickly converged to the confederate's answers. Other researchers have found that norms manipulated in this way persist for quite some time. Even when subjects are brought in up to a year later, they show adherence to that initial norm. Moreover, within the context of the experiment, Jacobs and Campbell (1961) have shown how norms can be transmitted across "generations" of subjects. Suppose subjects 1 and 2 initially converge to a norm, but subject 1 is then replaced by subject 3 for enough trials, and subject 2 is then replaced by subject 4. The final group consisting of totally new subjects 3 and 4 will conform to the norm already established by subjects 1 and 2.[12]

Solomon Asch (1951) expanded on these results through an even simpler task. Subjects were brought into a lab and asked to sit with others and judge the length of lines such as those shown in figure 1. The subject hears the judgment of the others and then makes his own. For several trials, this is a very boring task, as it is pretty obvious which line is longer. But then there is a twist. On one of the trials, the first person makes a wrong choice. A second person then makes the same wrong choice. And so it continues until it is the subject's turn to choose. In Asch's experiment, there were 5 to 12 "conformity" trials out of 10 to 18 total trials. What Asch found was stunning. Between 50 to 80 percent of the subjects yielded to the erroneous majority at least once. Of course, as Asch notes, it is not the subjects' perception of the line length that is altered (unlike, perhaps, in the Sherif experiment). Many subjects (but not all) are simply willing to conform in their behavior.

Other experiments suggest that individuals may conform strongly to their roles (Aronson, Steele, Salinas, and Lustina 1998). A modern-day version of this can be seen in recent work on stereotype threat. In one early and particularly clever study, African-American and American Caucasian subjects in the United States were asked to take the Graduate Record Examination (GRE). In one condition, the subjects are asked to fill out a questionnaire indicating their gender, major area of study, and other demographic variables (but not race). In another condition, they are also asked to fill in their race. This simple manipulation—by evoking the race of the person—elicited conformity to a common stereotype. The African-American students, who are often stereotyped as less intelligent, responded by fulfilling this expectation. In the condition where race was salient their performance was far worse than that of the Whites.' However, in the condition where race was not salient, the African American subjects performed exactly the same as the Whites.

Hoff and Pandey (2004) recently performed a similar experiment on caste in India. Children of lower and upper caste were asked to solve mazes on a piece-rate basis. In some cases caste is made highly salient (through public announcement of the child's caste). When this occurs, the low-caste children solve 25 percent fewer mazes. The researchers go on to provide some evidence for a mechanism in this case. When asked to accept or reject a gamble in which there is *no* scope for judgment by an experimenter, making caste salient does *not* produce a caste gap. Instead, in the case where there is scope for subjective judgment by others, caste appears to have an effect. This

suggests that one of the reasons people fall so easily into caste roles is that they expect others to treat them according to these roles.

As Hoff and Pandey note, these types of findings can be helpful for understanding why institutions and inequalities persist. Norms and institutions can shape what people believe is possible. They can shape people's perceptions of how others will respond to them, and thereby drive behavior. For example, a lower cast child may feel strongly the norms and stereotypes that go along with being lower caste. This can in turn serve as a powerful deterrent to becoming educated or seeking a higher station in life. In this way, inequalities (when defined by well-identified groups) can persist.

Policies attempting to reduce inequalities need to be highly cognizant of the prevailing cultural norms. In the low-caste case, for example, simply giving supply-side incentives or reservations alone may not solve the problem. The tug of the prevailing norms can be stronger than material interests. The flip side of this logic produces a classic "big push" type of argument. If some small group of individuals who are typically discriminated against does manage to break the norms and succeed, the effect can be powerful. They can serve as role models for many others and remove at least the norm-induced barrier. In these models, the key questions are how to promote this initial change, and how to then publicize the resulting successes.

Self-Serving Bias and Evaluation

Hastorf and Cantril (1954) asked two groups of students, one from Princeton and one from Dartmouth, to watch film of a Princeton-Dartmouth football game. Each student was asked to count the number of penalties committed by both teams. Though both groups watched the exact same tape, the counts show that they "saw a different game." Dartmouth students saw an equal number of flagrant and mild penalties committed by both teams. By contrast, the Princeton students counted three times as many flagrant penalties by Dartmouth as by Princeton—and the same number of mild penalties. This experiment illustrates an often-repeated finding in psychology, that the beliefs and perceptions that feed into forming opinions can be biased. In this case, the students' personal affiliations with their schools influenced what they saw. In other cases, it may be prior beliefs or a desire for a particular outcome that leads to biased perceptions and opinions.

Babcock and Loewenstein (1997) provided a particularly stunning example of this bias. Subjects were asked to bargain over how to deal with a particular tort case (which was based on a real trial that occurred in Texas). Each subject was assigned the role of lawyer for either the defendant or plaintiff. The subjects read all the case materials and then bargained with each other over a settlement. If they fail to settle, the award amount will be what the judge decided in the actual case (which is unknown to the subjects at the time of bargaining). Interestingly, subjects are to be paid as a function of how much they manage to get in the settlement; but they will pay a cost if they go to the judge without settling. Subjects are also asked to assess (in private) how much they think the judge will award them. Finally, some pairs of subjects read

the entire description of the case *before* knowing what role they were to play. Others read it afterward. This order of reading the case description has a large effect. Those who read first settled at a rate of 94 percent, without going to the judge. But those who read afterward settled at a rate of only 72 percent. Moreover, as a rule, those who read beforehand tended to exaggerate how much the judge would favor them. In short, these subjects exhibited beliefs that were quite biased, based on their status. Plaintiffs believe the merits of the case support a large award, whereas defendants think it merits a small one. These conflicting beliefs are generated through nothing more than the roles the subjects were assigned. When they read through the case, they selectively interpreted the information they saw in light of their own role. Note that this goes against their material interests in one important way: They must pay to go to court, yet their biased beliefs send them to court much more often. Much like subjects in the Princeton-Dartmouth football game described earlier, these subjects saw very different cases. In some sense, each saw what they "wanted" to see.

Of all the evidence I've presented, I feel this outcome has the most far-reaching implications for how development policy is practiced—and that is why I end with it. I feel this evidence tells us something very important about how development policy ought to be evaluated. A useful example is in the study of Cabot's intervention program for delinquent youth in the towns of Cambridge and Somerville, Massachusetts (Powers and Whitmer 1951). This intervention combined all the best tools available at the time for helping these delinquent youths: from tutoring and psychiatric attention, to interventions in family conflicts. Those involved in the program raved about its success. They all had very positive impressions. What made the program unique, however, was that a true random assignment procedure was used to assign the students. When these data were examined, contrary to the very positive (and likely heartfelt impressions of the caseworkers), there was little measurable effect of the program.

Ross and Nisbett (1991) cited another interesting example: a meta-analysis by Grace, Muench, and Chalmers (1966), who studied all medical research on the "portacaval shunt"—a popular treatment for cirrhosis of the liver, for which 51 studies examined the efficacy. The doctors and scientists conducting these studies all had the same good intent: to determine whether this procedure worked. But the studies differed in one important way: 15 of them used controls but not randomization, while 4 of them used truly randomized strategies. Thirteen of the 15 nonrandomized studies were markedly or moderately enthusiastic about the procedure. Yet only one of the randomized studies was markedly or moderately enthusiastic.

What was going on here? I feel the good intentions of the doctors and scientists got in their way. There is always subjectivity in nonrandomized trials, what controls to include, what controls not to include, which specification to run, and so forth. Such subjectivity leaves room for self-serving bias to rear its head. And it is exactly because the researchers on these topics are well intentioned, exactly because they hope the procedure works, that it is all too easy for them to find a positive result. Much as with the Dartmouth and Princeton students, these scientists saw in some sense what they wanted to see.

As noted earlier, I feel that both of these examples highlight an important fact about evaluation. Especially in the development context where most people working with a

project would like to see it succeed, it is all too easy for self-serving bias to affect evaluations. Beyond the obvious econometric benefits of randomized evaluation, I think this is one of the greatest practical benefits. Randomized trials are a way to minimize (though obviously not eliminate) a researcher's latent biases. They allow us to escape the dangers of biased perception, from which researchers or field workers are no more free than anyone else in the population.

Concluding Observations

Much of recent development economics has stressed the importance of institutions. Property rights must be enforced to provide appropriate incentives for investment. Government workers must be given appropriate incentives to ensure the delivery of high-quality public services. Banking may need to be privatized to ensure a well-functioning credit system that in turn allows for better savings and smoother consumption. The common theme here is that institutions must be improved to help to resolve issues between people. Institutions may reduce externalities, solve asymmetries of information, or help resolve coordination problems. This focus on resolving problems *between* people rather than *within* individuals is natural to economists. The predominant economic model of human behavior leaves little room for individuals themselves to make mistakes. In fact, economists assume that people are unbounded in their cognitive abilities, unbounded in their willpower, and unbounded in their self-interest (Mullainathan and Thaler 2001). And once we admit human complexities, institutional design in development becomes not just about solving problems *between* people. It also becomes about developing institutions in ways that help any one person deal with their own "problems." I hope the small set of examples presented here help illustrate how a deeper understanding of the psychology of people might eventually improve development policy.

Notes

1. See Mullainathan and Thaler (2001) for references and a summary discussion.
2. One reason subjects show such preferences may be that they doubt they will actually receive the money later, leading them to value it at a lower rate. While this may be an effect, the literature on discounting finds similar results—even when these issues of trust are dealt with (Frederick, Loewenstein, and O'Donoghue 2002).
3. This last point provides one way to distinguish this explanation from a rational model with large liquidity shocks. Moreover, in such a rational model, difficulties arise if parents rationally forecast such shocks and there are scale economies to attending for long continuous periods. In this case, parents should build a "buffer stock" early on—to insure against such shocks and then send the child to school for one long (and presumably more productive) stretch.
4. Note that in this framework, unlike in a liquidity constraint framework, this policy would work even if these payments all had to be made *prior* to the beginning of the school year. This would be analogous to the use of lay-away plans at retail stores in the United States.

5. To make this contrast stark, note that in the United States, payday loan companies are a very profitable form of micro-credit.

6. In this short space, I cannot do justice to all the psychological tools that the Save More Tomorrow plan relies on. The full discussion in the original paper is well worth reading as an example of how to use psychological tools to better design policy.

7. Of course, this is a comparative static only. In any given context there may be pressing reasons to favor one policy over the other.

8. For interesting differences in some tribal cultures, see Heinrich et. al. (2002).

9. Of course, the effect may have been smaller had subjects perceived Joe (the confederate) as having bought the Coke for purposes of an investment.

10. One of the debates in the experimental literature in economics is whether this "punishment" view is needed to explain these data. There is enough auxiliary evidence, however, that while the punishment view may not be the full story it is at least part of the story.

11. Part of this implication might be counterintuitive from a pure self-interest point of view. For example, it may be easier to get teachers to come to school if attendance is high than when it is low. This would appear paradoxical if teachers were simply trying to reduce the amount of work they were doing, since higher attendance would precipitate even more work for teachers when they do show up at school.

12. Camerer and Weber (2003) present an interesting examination of how such norms can arise and evolve over time.

References

Ainslie, George. 1992. *Picoeconomics*. Cambridge, U.K.: Cambridge University Press.

Akerlof, George A. 1991. "Procrastination and Obedience." *American Economic Review* 81 (2): 1–19.

Ardener, Shirley, and Sandra Burman, eds. 1995. *Money-Go-Rounds: The Importance of Rotating Savings and Credit Associations for Women.* Washington, D.C.: Berg Publishers.

Ariely, Dan, and Klaus Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." *Psychological Science* 13 (3): 219–24.

Aronson, J., C. M. Steele, M. F. Salinas, and M. J. Lustina. 1998. "The Effects of Stereotype Threat on the Standardized Test Performance of College Students." In E. Aronson, ed., *Readings about the Social Animal* (8th ed.). New York: Freeman.

Asch, Solomon. 1951. "Effects of group pressure upon the modification and distortion of judgement." In H. Guetzkow, ed., *Groups, Leadership, and Men.* Pittsburgh, Penn.: Carnegie Press.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2004. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product." Mimeo. Economics Department, Princeton University, Princeton, N.J. http://www.wws.princeton.edu/~dkarlan/downloads/SEED.pdf.

Babcock, Linda, and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives* 11 (1): 109–26.

Becker, Gary. 1993. *Human Capital* (3rd ed.). Chicago: University of Chicago Press.

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir. 2004. "Behavioral Economics of the Poor." *American Economic Review* 94 (2): 419–23.

Camerer, Colin, and Roberto Weber. 2003. "Cultural Conflict and Merger Failure: An Experimental Approach." *Management Science* 49 (4): 400–15.

Darley, J. M., and B. Latané. 1968. "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology* 8 (4): 377–83.

De, Anuradha, and Jean Dreze. 1999. *Public Report on Basic Education in India*. New York: Oxford University Press.

Farkas, Steve, and Jean Johnson. 1997. *Miles to Go: A Status Report on Americans' Plans for Retirement*. New York: Public Agenda.

Frederick, S., George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* 40 (2): 351–401.

Genesove, David, and Chris Mayer. 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *Quarterly Journal of Economics* 116 (4): 1233–60.

Gilbert, Daniel T., R. W. Tafarodi, and P. S. Malone. 1993. "You Can't Not Believe Everything You Read." *Journal of Personality and Social Psychology* 65 (2): 221–33.

Grace, N. D., H. Muench, and T. C. Chalmers. 1966. "The Present Status of Shunts for Portal Hyperextension in Cirrhosis." *Gastroenterology* 50: 684–91.

Gruber, Jonathan, and Sendhil Mullainathan. 2002. "Do Cigarette Taxes Make Smokers Happier?" Mimeo. Department of Economics, Massachuseets Institute of Technology. Cambridge, Mass. http://econ-www.mit.edu/faculty/gruberj/files/happy81.pdf.

Gugerty, Mary Kay. 2001. "You Can't Save Alone: Testing Theories of Rotating Savings and Credit Organizations." Mimeo. Economics Department, Harvard University, Cambridge, Mass.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3 (4): 367–88.

Hastorf, Albert, and Hadley Cantril. 1954. "They Saw a Game: A Case Study." *Journal of Abnormal and Social Psychology* 49 (1): 129–34.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. 2002. "Cooperation, Reciprocity, and Punishment: Experiments from 15 Small-Scale Societies." Book manuscript, in progress.

Hoff, Karla, and Priyanka Pandey. 2004. "Belief Systems and Durable Inequalities: An Experimental Test of the Effect of Indian Caste on Performance." Working Paper 3351. Investment Climate, Development Research Group, World Bank, Washington, D.C.

Jacobs, R. C., and D. T. Campbell. 1961. "The Perpetuation of an Arbitrary Tradition through Several Generations of a Laboratory Microculture." *Journal of Abnormal and Social Psychology* 34 (4): 385–93.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98 (6): 1325–48.

Laibson David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2): 443–78.

Madrian, B. C., and D. F. Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics* 116 (4): 1149–87.

Messick, D. M., and K. P. Sentis. 1979. "Fairness and Preference." *Journal of Experimental Social Psychology* 15 (4): 418–34.

Mullainathan, S., and Richard Thaler. 2001. "Behavioral Economics." *International Encyclopedia of Social Sciences*. Pergamon Press.

Odean, Terry. 1998. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance* 53 (5): 1775–98.

Powers, E., and H. Whitmer. 1951. *An Experiment in the Prevention of Delinquency.* New York: Columbia University Press.

Read, Dan, George Loewenstein, and S. Kalyanaraman. 1999. "Mixing Virtue and Vice: Combining the Immediacy Effect and the Diversification Heuristic." *Journal of Behavioral Decision Making* 12 (4): 257–73.

Regan, R. T. 1971. "Effects of a Favor and Liking on Compliance." *Journal of Experimental Social Psychology* 7 (6): 627–39.

Ross, Lee., and Richard E. Nisbett. 1991. *The Person and the Situation: Perspectives on Social Psychology.* Philadelphia: Temple University Press.

Samuelson, W., and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk & Uncertainty* 1 (1): 7–59.

Sherif, M. 1937. "An Experimental Approach to the Study of Attitudes." *Sociaometry* 1 (1/2): 90–98.

Strotz, R. 1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies* 23 (2): 165–80.

Thaler, Richard. 1981. "Some Empirical Evidence on Dynamic Inconsistency." *Economic Letters* 8 (4): 201–07.

Thaler, Richard. 1988. "The Ultimatum Game." *Journal of Economic Perspectives* 2 (4): 195–206.

Thaler, Richard., and Shlomo Benartzi. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy*, forthcoming.

Trope, Y., and A. Fishbach. 2000. "Counteractive Self-Control in Overcoming Temptation." *Journal of Personality and Social Psychology* 79 (4): 493–506.

Vermeesch, Christel. 2003. "School Meals, Educational Achievement, and School Competition: Evidence from a Randomized Evaluation." Mimeo. University of Oxford. http://hicks.nuff.ox.ac.uk/users/vermeersch/schoolmeals.pdf.

Wertenbroch, K. 1998. "Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice." *Marketing Science* 17 (4): 317–37.

Comment on "Development Economics through the Lens of Psychology" by Sendhil Mullainathan

COLIN F. CAMERER

Sendhil Mullainathan's paper does a terrific job of acquainting the reader with concepts in behavioral economics. Sendhil is one of the young wizards pioneering behavioral economics in two directions (which those of us working in behavioral economics since the 1980s have been hoping would emerge for years)—formalizing psychology mathematically in a way that makes it directly useful in economic theory, and searching for applications in field data (see Camerer and Loewenstein 2003 for a recent review). Sendhil has laid out the central questions clearly and thoroughly, so my discussion will focus on two topics that go beyond his essay:

1. Other concepts in psychology that can be used to think about development, particularly attribution theory and mental modeling.

2. The prospect for simple experiments that can tell us something about development, illustrated with two examples: a remarkable cross-cultural coordinated field experiment, and a laboratory experiment on simple capital investment economies with poverty traps.

New Psychology and "21st-Century Behavioral Economics"

Sendhil's paper covers what might be called "20th-century behavioral economics." That is, most of these ideas developed during the 1980s, when psychologists like Kahneman and Tversky used the rational choice model as a foil against which to understand limits on rationality; Kahneman shared the Nobel Prize with Vernon Smith in 2002. Their idea was to use deviations from rational principles such as utility-maximization and Bayesian updating, much as optical illusions are used to study perception: The deviations from rationality tell us about the basic mechanisms of judgment and choice.

Of the concepts Sendhil describes, a particularly important one for development is "loss-aversion"—the fact that losses loom larger than equal-sized gains (and probably

Colin F. Camerer is the Rea A. and Lela G. Axline Professor of Business Economics at the California Institute of Technology in Pasadena, California.

activate different regions of the brain). An implication of this principle is that people really dislike giving up what they have gotten used to having or experiencing. This makes it difficult to implement reforms. At lunch with Vernon Smith after this symposium, Vernon mentioned that in implementing economic design (a practice pioneered by him and my Caltech colleague Charles Plott, among others), it is crucial to "grandfather" rules so that people with current entitlements won't fear having them taken away. This dictum is the economic engineer's way of incorporating awareness of loss-aversion to craft reforms that can be implemented from a behavioral point of view.

The fact that the psychologists who laid the groundwork for 20th-century behavioral economics focused on deviations from simple rationality principles constrained the kinds of psychology that they tapped—and that we imported into economics. This leaves a lot of other interesting ideas in psychology. I'll discuss two of these: attribution and mental modeling.

Attribution of Credit and Blame

In most complex systems, we really don't know who is at fault. Psychologists call the problem of determining cause and effect, and credit and blame, "attribution theory." Attribution is important because, if we think oil prices are high because oil sheikhs and OPEC are conspiring, that may lead to one political decision. If, however, we think an incumbent president is to blame, that leads to a different decision.

Another important feature of attribution is that there's lots of scope to disagree about who to blame. Often two groups "self-servingly" blame each other, and there may not be clear evidence that can establish who should be blamed (absent formal institutions such as court proceedings). A good example is economic sanctions and embargoes against countries whose policies we disagree with. It is fair to say that sanctions have not always worked well (the fact that they are often in place for so long might be taken as evidence of failure). Self-serving attribution of credit and blame might help explain why, for example, from an American point of view it seems obvious that the Cubans should blame Castro for the economic loss from American sanctions and push to overthrow him or demand reform that would end the sanctions. But the Cubans may see it differently—they may blame the Americans, which could actually increase Cuban support for Castro. (Being threatened by an outside force often brings people together, a phenomenon called the "common enemy effect" in social psychology.[1]) Of course, I am not taking a stand on who is really to blame. In fact, that's the key point: Because it is difficult to assign blame (the way the legal proceedings may in a negligence case), there is room for disagreement that is self-serving. The disagreement means that the sanctions will not work as well as hoped by those who imposed them.

Mental Models

In developing economies a cognitive scientist would study the mental models that participants in the economy have of what's going on, a point stressed by Douglass North (1994). A mental model is the intuitive set of principles or ideas of how things

work, which governs people's predictions about the effects of change. These mental models are oversimplified, and often illogical or physically incorrect. For example, before special training in physics, people often have ideas about physical systems that are flat wrong, but intuitive (McCloskey 1983). If you ask people what happens to a ball you are twirling on a string in a circular motion, when the string breaks, many of them tell you the ball will continue to arc in a circle away from the broken string—because the velocity of the ball contains "momentum," which includes the circularity from its previous motion. Of course, we now know that this mental model is wrong. When the string breaks, the ball heads off in a straight line tangent to the imaginary perpendicular line between the center of the circular orbit and the ball's position when the string broke. We know the "circular momentum" theory is wrong, but it was an accepted model in physics until about 1500, before the Newtonian revolution.

So it may well be that people in simple economies (or even current theorists!) have oversimplified mental models of political economy that are illogical or empirically incorrect. Yet these mental models often guide voting and protest and, as noted above, attributions of credit or blame. Incorrect mental models are especially important in dynamic systems where the lag time for policies to yield good results is unknown. This point has been clearly established by John Sterman and colleagues (Sterman 2002). Sterman illustrates his point with the problem of getting a hot shower in a hotel you just checked into but have not visited before. You turn on the hot water. If it takes a little while to turn hot, you turn it hotter, wait a little longer, then turn it even hotter. A minute later you step in and are soon blasted by scalding hot water.

The mistake the hapless shower-taker makes is underestimating the amount of hot water in the invisible supply line. You can easily imagine how, in implementing political reforms where it's difficult to promise people when good things will happen, that a misunderstanding of dynamics could be important in causing political impatience.

It is well documented that education is important for economic growth. Literacy, awareness of scientific principles, and marketable human capital are obviously important parts of why education is good for economic growth. But another part of the value of education may be that it supplies people with better mental models and shines light on logical inconsistencies in their beliefs. Education generates a sense of who you can trust, whether you can trust what you read in the paper, and so on, which may in turn disarm naive beliefs, repair people's faulty mental models of political economy, and enable good reforms.

Experiments *about* Development, and *during* Development

Experiments have been crucial in advancing behavioral economics. I'll discuss two directions that may be relevant for development.

Experiments about Development

To an outsider, a striking fact about economic research on growth is the limited dialogue between theorists—those who do statistical analyses (cross-country growth

regressions)—and policymakers with a lot of field experience. Experiments may help bridge this gap by providing evidence from simple artificial domains that correspond to the artificial worlds of simple theory, since experiments can always be enriched to include features policy analysts and statistical analysis suggest are important.

Capra and others (2004) have been conducting simple experiments on economies with capital investment and "poverty traps." In these experiments subjects choose levels of capital investment. Investment produces goods that produce utility. There is a critical mass of investment that boosts productivity, and all the subjects know this. The critical mass means there are two steady-state equilibria: one in which people invest below the critical threshold and produce and consume less (a technological poverty trap); and another in which investment is high, the threshold is crossed so that productivity is boosted, and people consume more and have higher utility (which in turn translates into higher actual money earnings from the experiment). Figure 1 (left graph) shows what happens in a typical baseline session. The y-axis shows utilities from consumption U(C)— the two equilibrium levels are the horizontal lines at 6 (poverty trap) and 18 (efficiency) over many periods of time in the experiment (x-axis). There is some movement up and down, but results basically get stuck near the poverty-trap equilibrium where U(C) = 6.

The middle panel in figure 1 shows what happens when people are allowed to propose capital allocation schemes and vote on them (the votes are binding). This

FIGURE 1.
Utility over Time in "Poverty Trap" Experiments
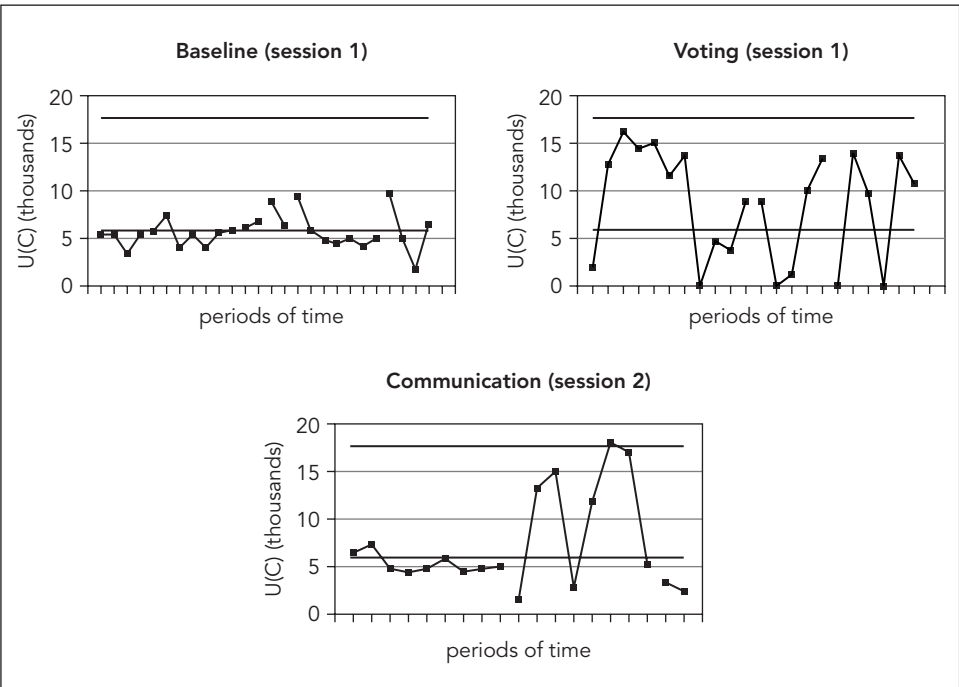
portrays an economy in which the International Monetary Fund or some other coordinating agency (or a national industrial policy) requires certain levels of investment. As shown in this figure, voting often lifts investment up closer to the utility-maximizing level of 18, but votes usually do not stick for long and the experimental economy is often drawn down into the poverty trap result of $U(C) = 6$.

The right panel shows what happens when subjects are allowed to freely communicate using an instant messaging system. As with voting, allowing subjects to talk helps them agree to invest more (in the later stages), but only temporarily.

The message of these experiments is that even when high investment is better for everyone (Pareto-improving, in economic jargon), and when subjects can talk about the fact that everyone benefits from a productivity boost when total investment is high enough, it is hard to reach the best equilibrium and make it stick. Of course, these are simple experiments. But if it is difficult for a small number of college students to reach the good equilibria in these simple domains, one can't help but wonder how a much more complex economy can do so. More importantly, criticisms of the external validity of the experiment can readily be translated into designs for new, richer, experiments and predictions about what would happen if the experiment were changed (played for more money, for more time, or played by actual firms or agency regulators). The point is that the experiments are a platform onto which complications can easily be added.

Experiments on Social Capital during Development

Another kind of experiment that is catching fire is the idea of "packing your laboratory" and conducting controlled experiments in developing countries. There you can study the people whose behavior you want to eventually effect, in domains that are familiar to the subjects (as in Harrison and List, currently in press, and in Karla Hoff's companion discussion to mine, in this volume).

One dramatic field experiment project is a unique collaboration by a dozen anthropologists in 15 small-scale societies, mostly in Africa and the Amazon basin. These are all extremely small-scale societies, typically barter economies with little political structure and some degree of market exchange (such as selling extra crops or cows at market once a week). The anthropologists were interested because these are some of the last places on Earth that resemble hunter-gatherer economies from 100,000 years ago, and in which we think the human brain may have evolved.

The anthropologists in this field experiment conducted a series of simple games. I'll describe only one, an "ultimatum" or take-it-or-leave-it bargaining game. One person is given a sum of money, say $10 (usually a sum with large local purchasing power, worth several days' wages, so subjects are highly motivated). Then the anthropologists offer a fraction of the $10 to another subject (a stranger) in private. If the "responder" subject accepts the offer, both subjects earn the amounts of money agreed upon. If the responder says "No," the offer is rejected and they get nothing. The game is a simple way of measuring norms of sharing, and whether people express "negative reciprocity" by rejecting offers they perceive as unfair. In many

FIGURE 2.
Mean Ultimatum Offers and Market Integration and Cooperation
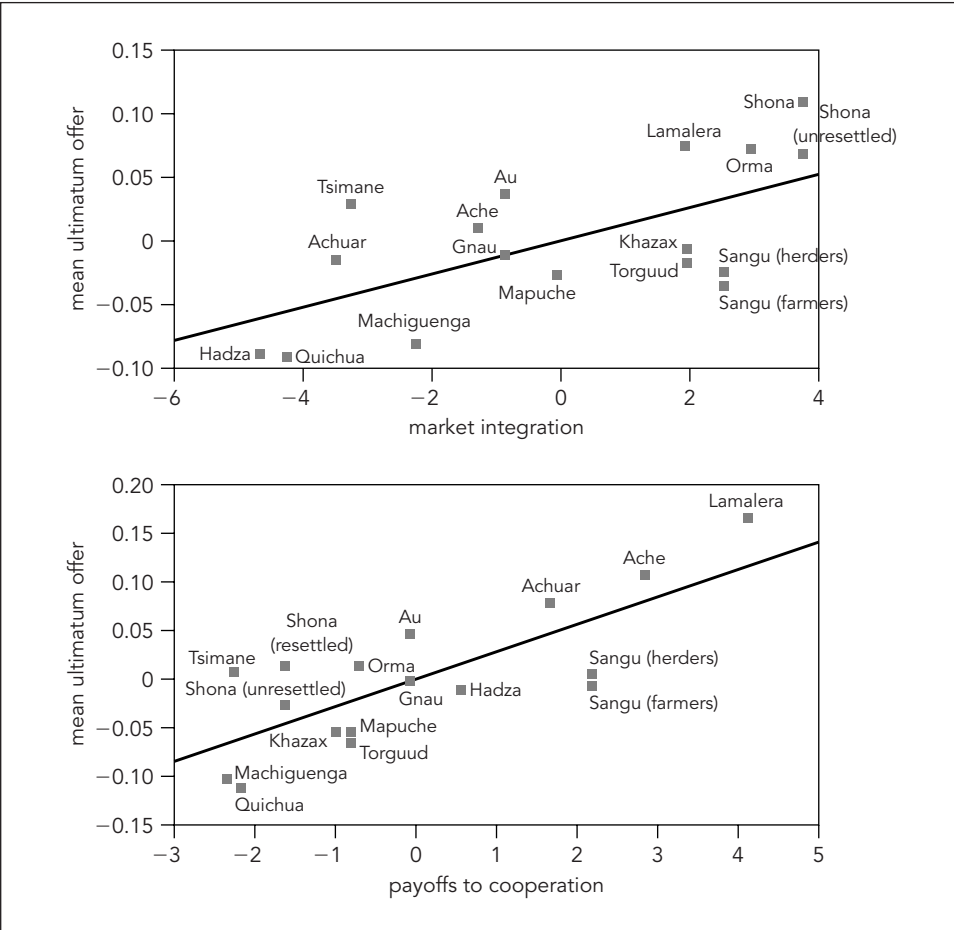
experiments people typically offer a little less than half, and offers of less than $2 are rejected about half the time (see Camerer 2003, chapter 2).

Figure 2 shows the average ultimatum offer in each society on the y-axes. The x-axes represent how the societies rank, from high to low, in terms of market integration (roughly the amount of their consumption that comes from market-exchanged goods, the top graph) and the social payoff to cooperation (e.g., whether they build schools and plow fields together, the bottom graph). The numbers are rescaled so that zero is not a zero offer but is the mean across the groups (around 40 percent, which is typical of Western college students and others).

Ironically, in some of these groups you see something close to the subgame perfect prediction of game theory, which is that people who care only about getting the most money should accept very little, and the person making the offer should anticipate this

and offer very little. Among the Machiguenga in Peru, and the Quicha in Ecuador, the subgame perfect self-interest prediction of low offers is a good approximation. The Machiguenga are highly asocial (they have no proper names to refer to people other than their kin). They don't seem to think a stranger should share with them; and they don't get upset when they are offered very little (only one offer was rejected). Thus, the anthropologists found some places where game theory is alive and well: in remote villages in South America.

The key point of figure 2 is the positive correlation between ultimatum offers and the degrees of market integration and cooperation. A naïve reading of economist Adam Smith (in *The Wealth of Nations*) is that self-interest is sufficient to produce market allocations, because "It is not from the benevolence of the butcher, the brewer or the baker that we expect our dinner, but from their regard to their own interest." This is sometimes interpreted to mean that Smith thought a sense of fairness or justice had nothing to do with effective operation of markets. But in his *Theory of Moral Sentiments* Smith uses the phrase "invisible hand" quite differently. Here he says the wealthy "are led by an *invisible hand* to make nearly the same distribution of the necessities of life which would have been made had the earth been divided into equal portions among all inhabitants." (see Ashraf and others 2004)

Thus, Smith hints that even in market allocations either an implicit or explicit sense of fairness plays a role. The cross-cultural experiments show that fair sharing *is* correlated with participation in markets across some societies, not antithetical to it. Of course, we do not know the direction of causality. Fair sharing of surplus might enable markets to flourish, or trading with strangers might inculcate a sense of sympathy and fairness. In any case, these experiments show how something central to development—a concept of social capital in the form of widely shared norms of surplus-sharing—can be understood with experimental data in a fresh way.

Note

1. A May 2004 *Los Angeles Times* article reported that coalition forces were surprised at how much the often-divisive Sunni and Shiite Muslims had banded together against coalition forces. The article reported a "common saying in Iraq" that encapsulates the common enemy effect: "Me and my brother against my cousin. Me and my cousin against the stranger."

References

Ashraf, Nava, Colin F. Camerer, and George Loewenstein. 2004. "Adam Smith, Behavioral Economist." Unpublished Caltech manuscript. http://www.hss.caltech.edu/~camerer/jepadamsmith15.doc.

Camerer, Colin F. 2003. *Behavioral Game Theory*. Princeton, N.J.: Princeton University Press.

Camerer, Colin F., and George Loewenstein. 2003. "Behavioral Economics: Past, Present and Future." In C. Camerer, G. Loewenstein, and M. Rabin, *Advances in Behavioral Economics*. Princeton, N.J.: Russell Sage Foundation and Princeton University Press.

Capra, C. Monica, Tomomi Tanaka, Lauren Munyan, Colin F. Camerer, Veronica Sovero, Lisa Wang, and Charles Noussair. 2004. "The Effect of Institutions on Poverty Traps in Experiments." Caltech Working Paper. California Institute of Technology, Pasadena, California.

Harrison, Glenn, and John List. Forthcoming. "Field Experiments." *Journal of Economic Literature*.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Herbert Gintis Ernst Fehr, and Colin Camerer, eds. 2004. *Foundations of Human Sociality: Ethnography and Experiments in 15 Small-Scale Societies*. Oxford, U.K.: Oxford University Press.

Los Angeles Times. 2004. "U.S. Learns Art of the Deal in Iraq." May 30, pp. A1, A12.

McCloskey, Michael. 1983. "Intuitive Physics." *Scientific American* 248: 122–30.

North, Douglass C. 1994. "Economic Performance through Time." *American Economic Review* 84: 359–68.

Sterman, John. 2002. "All Models Are Wrong: Reflections on Becoming a Systems Scientist." *System Dynamics Review* 18: 501–31. http://web.mit.edu/jsterman/www/All_Models.html.

Comment on "Development Economics through the Lens of Psychology" by Sendhil Mullainathan

KARLA HOFF

The paper by Sendhil Mullainathan reviews experimental work that challenges the descriptive validity of the tenet that individuals' choices can always be understood as the outcome of a rational, emotion-free balancing of costs and benefits. For example, weakness of will and time-inconsistent preferences may interfere with rational saving, investment, and education decisions. Self-serving biases may lead to (self-defeating) bargaining impasse and overly optimistic assessments of development projects. A better understanding of the ways in which behavior systematically departs from rationality promises to improve the contribution that economists can make to development.

Mullainathan emphasizes the universal aspects of the departures from rational decisionmaking. In this comment I would like to draw from some of the same literature that he surveys a different, but complementary implication. Taking a more realistic view of people's limited understanding of themselves and their environment, as well as of the role of emotions, creates scope for belief systems that are a legacy of history to influence perception and behavior. I will argue that the content of many cognitive biases is shaped by social identity and that such biases can be a source of resistance to change in social systems. "People will be prejudiced so long as they continue to think," as Michael Billig, a leading scholar on the psychological basis of racism, put it.[1] But as he might have added, people in different social groups are prejudiced in different ways. Individuals view the world through the lens of historically created social identities, which may have a pronounced effect on their behavior.

Stereotypes

Psychologists have found evidence that unequal status breeds prejudice (Hoffman and Hurst 1990). Economists have found that even an overtly *arbitrary* classification of individuals into groups and hierarchizing of those groups affects earnings, because people prefer to trade with individuals with high status (Ball et al. 2001). How much

Karla Hoff is a Senior Research Economist in the Development Economics Research Group at the World Bank.

more powerful will status divisions be if they are believed to follow from nature or divinity? In societies that historically were highly unequal, belief systems emerged that attributed to nature or divinity the inequality that in reality resulted from social practices.[2] Such belief systems divide individuals into hierarchized, stereotyped social groups.

Stereotypes affect the way people process information and thus tend to be self-reinforcing. An experimental framework that reveals the effect of stereotypes on the way people process information is found in Stone, Perry, and Darley (1997). They asked all participants (who were American Caucasians, hereafter referred to as White) to listen to the same running account of an athlete's basketball performance on the radio. Half the participants were led to believe that the target player was White, and half that he was African-American. The results indicated that information was less likely to be absorbed if it was discordant with the prevailing U.S. stereotypes that Whites are more academically talented than African Americans and that African Americans are more athletically gifted. The White target player was perceived as exhibiting less natural athletic ability but more "court smarts," whereas the African-American target player was perceived as exhibiting less court smarts but more natural athletic ability. A growing recognition of prejudices of this sort has led economists to pay increasing attention to stereotypes as a mechanism behind the persistent gaps in economic achievement between observably distinguishable social groups long after overt barriers to economic and social mobility have been removed.[3]

Stereotypes can be self-fulfilling not only because they bias perceptions, but also because they influence the behavior of the stereotyped individuals. In another experiment, Stone et al. (1999) asked Princeton undergraduate volunteers to complete a miniature golf course. The students' performance was measured by how many strokes were needed to put the ball in the hole. (Fewer strokes mean a better performance.) The variable that the experimenters manipulated was the description of the task. In one treatment, the task was described as a "standardized test of natural athletic ability." In the other treatment, the task was described as a "standardized test of sports intelligence." Table 1 presents the results. When the task was described as a test of natural athletic ability, the African-American participants performed better than the Whites: They averaged only 23.1 strokes to complete the 10-hole golf course, compared to an average of almost 27.8 for the White participants. But when the task was instead described as a test of intelligence, the race gap was reversed.

TABLE 1.
Effects of Race and the Test Frame on Performance

Student golfers	Test frame manipulation	
	Natural ability	Sports intelligence
Black participants, golf strokes	23.1	27.2
White participants, golf strokes	27.8	23.3

Source: Stone et al. 1999.

One way to interpret the behavior captured in table 1 is that social ideas—stereotypes about the talents of different social groups—impose *bounds from within*. Whereas *homo economicus* changes his behavior only when external constraints change, the behavior of real individuals depends as well on belief systems that society has sedulously impressed on them. Negative stereotypes create anxiety that may interfere with performance; that is why the psychologist Claude Steele termed this kind of behavior "stereotype threat." Positive stereotypes boost self-confidence that may lead individuals to expend greater effort.

Stereotypes influence behavior twice—not only by their impact on individuals' self-confidence, but also by their impact on the way individuals expect to be treated. To try to disentangle the effect of social identity salience on self-confidence from its effect on expectations of biased treatment, I undertook a series of experiments with Priyanka Pandey in rural north India (Hoff and Pandey 2004, forthcoming). In the first experiment, groups composed of three low-caste ("untouchable") and three high-caste children were asked to solve mazes and were paid based on the number of mazes they solved. In some cases caste was made salient by publicly announcing each participant's caste. When caste was not publicly announced, there was no caste difference in behavior. However, as shown in figure 1, increasing the salience of caste led to a significant decline in the average performance of the low caste, regardless of whether the payment scheme was piece rate or tournament. When caste was publicly announced, the low-caste children solved 25 percent fewer mazes on average in the piece rate treatment, compared to the performance of subjects when caste was not

FIGURE 1.
Average Number of Mazes Solved, by Caste, in Five Experimental Treatments
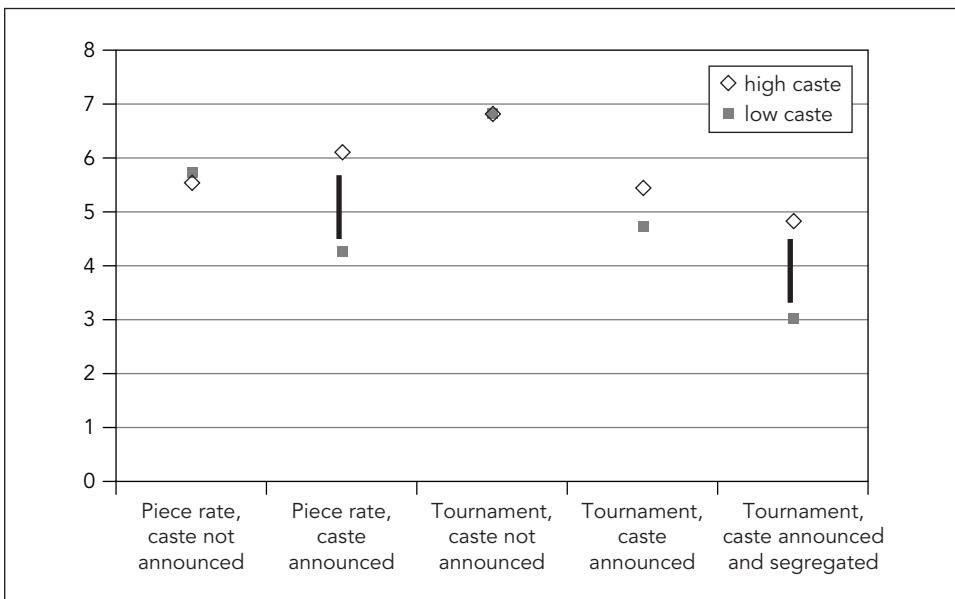
Source: Hoff and Pandey 2004a.

announced. In the tournaments, when caste was announced and groups were composed of six children drawn from *only* the low caste (a pattern of segregation that for the low caste implicitly evoked their traditional outcaste status), the decline in low-caste performance was even greater. A vertical line in the figure illustrates the statistically significant caste gaps.

In order to better understand the mechanism behind our results, in a new study we manipulated the scope for discretion in rewarding performance. In this study, participants were asked to practice solving a puzzle based on the game Rush Hour Traffic Jam. After practicing the game, they made a choice: They could accept a gamble in which they would attempt to solve a similar puzzle, with a payoff of 45 U.S. cents (20 rupees) if they succeeded and 2 cents if they failed; alternatively, they could refuse the gamble and receive 20 U.S. cents. In the control condition, the link between performance in the puzzle and the reward was mechanical;[4] in the experimental condition, there was scope for discretion in rewarding performance because the frame from the game board that kept vehicles on the road grid was removed, and a judge would award the success payoff only if the player did not let any vehicles move beyond the boundary of the game board. The subjects were told the following:

> All around the roads are wheat fields. To solve the puzzle, you have to free the red car from the traffic jam by moving cars around in such a way that no car comes off the road. Otherwise the fields will be damaged. . . . The person in the room who will give you the puzzle will watch you during that time. If he thinks you have solved the puzzle the right way, he will give you 20 rupees. If he thinks you have not solved the puzzle the right way, he will give you one rupee.

Consistent with a causal relationship running from caste salience to expectations of prejudicial treatment, we found that increasing the salience of caste had a large effect on the caste gap when there was scope for discretion in rewarding performance, but not when there was no such scope. The expectation by the low-caste subjects of prejudicial treatment may be rational given the discrimination that low-caste individuals experience in their villages. But the discrimination itself may not be fully rational. Cognitive limitations may prevent others from judging stigmatized individuals objectively.[5]

Biases and Bargaining Impasse

Standard bargaining models attribute bargaining impasse solely to incomplete information. However, experimental evidence indicates that self-serving bias is also a factor in bargaining impasse. Mullainathan describes a role manipulation experiment by Babcock and Loewenstein (1997) that pinpoints the causal influence of self-serving bias. In the experiment, subjects bargained over how to settle a tort case involving a motorcycle accident. Their rewards depended on the settlement or, if they were unable to settle, on the actual court decision less a substantial cost ("court fees"). In the control treatment, individuals were assigned a role (as plaintiff or defendant) *before* reading the 27 pages of materials from the original legal case. In the

experimental treatment, individuals were assigned a role *after* reading the legal case materials. When the subjects knew their role before reading the case materials, they were less likely to settle in face-to-face bargaining; they had formed quite different impressions from reading the case material compared to those who did not know their role in advance, and exhibited biased beliefs based on the position they were in. The greater the bias, the less likely they were to settle. As Mullainathan notes, "[t]hese conflicting beliefs are generated through nothing more than the roles they were assigned."

In the real world, social roles are not arbitrarily assigned but rather are impressed on an individual over a lifetime. A reliable finding in the economics literature is that social fragmentation hurts the delivery of public goods and economic performance.[6] The results of experiments on bargaining impasse provide a micro-foundation for that result. By distorting perceptions of fact and evaluations of fairness, self-serving bias can be "self-defeating" bias, which blocks the achievement of mutually beneficial agreements.

Are Institutions Manipulable?

Are institutions manipulable, or is a society's fate set in stone by its past institutions? The experimental evidence that Mullainathan marshals suggests that stereotypes and biased expectations can be carriers of history, mediating the effect of the past on the present. If beliefs can be changed, then such changes might play a role in bringing about effective institutional reform. The success story of a health care reform in northeast Brazil described in Tendler (1997) can be understood from this perspective. Tendler's case study reveals a three-way dynamic among an activist state government, municipal governments, and civil society. The state government conducted a program of mass advertising on the radio and in print that created in the community an expectation that the *municipio* was capable of delivering good health service. Civil society could then play the role of watchdog over the actions of local government and health care providers. This changed the environment in which the municipal governments functioned, introduced yardstick competition among them, and appears to have contributed to major improvements in public health in Ceará, one of the poorest states of Brazil.

But in cases of institutional reform that are not accompanied by changes in social norms, the prospect for engineering successful institutional change appears limited. Consider the problem of teacher absenteeism, which Mullainathan notes is a central impediment to improving education in developing countries. In India, the problem is particularly severe in areas that have a history of more oppressive past institutions. Pandey (2004) examines the effect of history on differences in teacher absenteeism across villages in north India. She compares districts in north India in which British rule concentrated local power and landownership in a single individual (the "landlord areas") with contiguous districts in which the British gave individual cultivators effective property rights to their land ("non-landlord areas"). In the early

1950s, after Indian independence, many states of India formally abolished landlords or passed tenancy reforms and ceilings on land holdings, and more recently, under a constitutional amendment, all states have adopted democratically elected village councils. However, as Banerjee and Iyer (2003) first documented, the historical legacy of inequality in landlord areas continues to be associated with a markedly poorer performance of local government. Pandey (2004) finds that teacher attendance and teaching activity are lower in landlord than in non-landlord areas. She also finds that mandated reservation for a low-caste individual in the elected position of village council head is associated with 22 percent *lower* teacher attendance. This finding is not explained by lower education or experience of low-caste village council heads in reserved seats. From the villagers' perspective, the opportunity to make a low-caste village council head an effective actor in village government may be a gamble that the low-caste individuals would have to fight for—against the resistance of the high castes. Although it is a great leap to draw from an experiment in a controlled setting an implication about political behavior, one can speculate that the low castes disproportionately reject this gamble, because, expecting to be treated prejudicially and expecting that others also would have that expectation, they are reluctant to challenge the system of high caste domination. The forces analyzed in the stereotype threat experiments that contribute to underperformance of historically subordinate social groups may also contribute in some settings to their political underperformance.

In conclusion, the fact that people are bounded in their ability to absorb and draw inferences from information creates broad scope for belief systems to influence economic behavior. Experimental findings pose two challenges for future research. The first one, as Mullainathan emphasizes in his much broader survey of behavioral economics, is to design field studies to test the relevance of experimental findings to specific development outcomes. The second is to design interventions that change the cognitive frames, and address the biases in judgment, that compromise the opportunities available to people.

Notes

1. Billig (1985), as cited in Fryer and Jackson (2003).
2. See Sunstein (1995).
3. See Loury (2002) for a wide-ranging study, and Bertrand and Mullainathan (2004) for a field experiment on the impact of racial stereotypes in the U.S. labor market.
4. A player solves the puzzle by moving vehicles on a wooden game board in such a way that the "player's car" escapes the gridlock and goes down the exit. The construction of the board makes it impossible to lift the vehicles off the board; they can only slide forward and backward on the tracks in the board. A frame on the game board prevents the vehicles from moving off the board except at the exit point. If a player solves the puzzle, his car exits and he obtains the success payoff, which is visible through the dashboard, from the underside of the car.

5. A survey of the literature on perceptual confirmation of stereotypes is found in Hamilton and Sherman (1994). A recent study of the effect of stereotyping on judgment finds that prison inmates with more Afrocentric features receive harsher sentences than inmates with less Afrocentric features, controlling for race and criminal history (Blair, Judd, and Chapleau 2004). A theory of the effects of "categorical cognition" is presented in Fryer and Jackson (2003).

6. See, for example, Alesina et al. (1999), and Banerjee and Somanathan (2001).

References

The word *processed* describes informally reproduced works that may not be commonly available through library systems.

Alesina, Alberto, Reza Gaqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114 (4): 1243–84.

Babcock, Linda, and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives* 11 (1): 109–26.

Ball, Sheryl, Catherine Eckel, Philip Grossman, and William Zame. 2001. "Status in Markets." *Quarterly Journal of Economics* 116 (1): 161–88.

Banerjee, Abhijit, and Lakshmi Iyer. 2003 "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." Massachusetts Institute of Technology, Department of Economics. Processed.

Banerjee, Abhijit, and Rohini Somanathan. 2001. "Caste, Community, and Collective Action: The Political Economy of Public Good Provision in India." Department of Economics, Massachusetts Institute of Technology, Cambridge, Mass. Processed.

Bertrand, Marriane, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Billig, Michael. 1985. "Prejudice, Categorization, and Particularization: From a Perceptual to a Rhetorical Approach." *European Journal of Social Psychology* 15 (1): 79–103.

Blair, Irene V., Charles M. Judd, and Kristine M. Chapleau. 2004. "The Influence of Afrocentric Facial Features in Criminal Sentencing." *Pscyhological Science* 15 (10): 674–79.

Fryer, Roland, and Matthew Jackson. 2003. "Categorical Cognition: A Psychological Model of Categories and Identification in Decision Making." NBER Working Paper 9579. National Bureau of Economic Research, Cambridge, Mass.

Hamilton, David L., and Jeffrey W. Sherman. 1994. "Stereotypes." In R.S. Wyer, Jr., and T.K. Srull, eds., *Handbook of Social Cognition* (2nd ed., vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Hoff, Karla, and Priyanka Pandey. 2004. "Belief Systems and Durable Inequalities: An Experimental Investigation of Indian Caste." World Bank Research Paper 0-2875. Development Economics Research Group, Washington, D.C.

———. Forthcoming. "Belief Systems and Barriers of Mistrust: Experimental Evidence from India." Special issue, *Economics of Transition.*

Hoffman, Curt, and Nancy Hurst. 1990. "Gender Stereotypes: Perception or Rationalization?" *Journal of Personality and Social Psychology* 58 (2): 197–208.

Loury, Glenn C. 2002. *The Anatomy of Racial Inequality*. Cambridge, Mass.: Harvard University Press.

Pandey, Priyanka. 2004. "Are Institutions Malleable? Effects of History, Mandated Representation and Democratization on Public Schools in North India." Working Paper. Pennsylvania State University and Indian Institute of Management, Lucknow, India.

Stone, Jeff, Christian I. Lynch, Mike Sjomeling, and John M. Darley. 1999. "Stereotype Threat Effects on Black and White Athletic Performance." *Journal of Personality and Social Psychology* 77 (6): 1213–27.

Stone, Jeff, Zachary W. Perry, and John M. Darley. 1997. "'White Men Can't Jump': Evidence for the Perceptual Confirmation of Racial Stereotypes following a Basketball Game." *Basic and Applied Social Psychology* 19 (3): 291–306.

Sunstein, Cass R. 1995. "Gender, Caste, and Law." In Martha Nussbaum and Jonathan Glover, eds., *Women, Culture, and Development*. Oxford, U.K.: Clarendon Press.

Tendler, Judith. 1997. *Good Government in the Tropics*. Baltimore: Johns Hopkins University Press.

Lessons of Experience

A Half-Century of Development

RICHARD N. COOPER

Development as a global policy objective dates from the 1940s. Relative to expectations at that time, the world economy performed outstandingly well during the second half of the 20th century. Worldwide growth in average per capita income exceeded two percent per year (historically unprecedented), many poor countries became rich, infant mortality declined, diets improved, longevity increased, and diseases were contained if not vanquished. Poverty on the World Bank definition of $1 per day (in 1985 U.S. dollars) declined dramatically, and the number of people living in poverty was halved—despite a more than doubling of the world population. Variations occurred over time and space, with rapid growth concentrated in Europe and Japan early in the period, then moving to east Asia, southeast Asia, and south Asia. Economic growth during the 1950s, and especially the 1960s, exceeded that in later decades. Examples of high growth could be found on every continent, but on average sub-Saharan Africa fared much less well than other regions. Declines in national per capita income were rare, but were concentrated in Africa. Civil disorder was a common but not universal cause of low growth. And median world income gained relative to the well-off; but both of these income groups surged ahead of the poorest.

World exports grew more rapidly than output, often leading the way. Many countries gradually shifted their exports away from primary products to labor-intensive manufactured goods, and as development proceeded, to more sophisticated manufactures and services. The fraction of the labor force devoted to agriculture declined significantly. One country after another achieved social stability, created the right incentives for effort and risk-taking, and engaged constructively with the world economy, which facilitated economic growth. Countries that lagged failed to meet one or more of these conditions. Civil and political liberties also spread during the post–World War II period, although less certainly and less securely. On the whole,

however, it was a good half-century for mankind—and the substantial poverty and misery that still exist should not detract from these achievements.

Economic development is a relatively new objective for economic policy, dating from the 1940s. Economists had of course been concerned with the causes and consequences of development since Adam Smith and earlier, and a number of countries in the 19th century—attempting to emulate Britain—strove for industrialization, or what their contemporaries considered development. On becoming Britain's Colonial Secretary in 1895, Joseph Chamberlain (referring to Britain's overseas colonies) pronounced that "it is the duty of the landlord to develop the estate" (Kapur, Lewis, and Webb 1997, p. 95). But development as a global objective for improving the economic well-being of ordinary people, reflected in Franklin Roosevelt's stated desire in 1941 to extend "freedom from want" throughout the world, was first embodied in the United Nations Charter adopted in 1945, which called for "economic and social progress and development." This objective is reflected in the formal name of the World Bank, the International Bank for Reconstruction *and Development* (IBRD), a distinction made at American insistence over initial British reservations (Kapur, Lewis, and Webb, pp. 57–62). It is, however, important to recall here, in view of revisionist history, that this name was formalized in 1943, at the height of World War II and well before the emergence of the Cold War, or roughly 1946–48.

The initial loans of the IBRD were overwhelmingly designated for reconstruction from the devastation of World War II, but a few loans were also made to Latin American countries, and in 1950 the World Bank made its first loan to India. Reconstruction proved too great a task for the IBRD alone, which was soon overtaken in magnitude by the U.S.-financed Marshall Plan for Europe, and U.S. aid to Japan and China. In his 1949 presidential inaugural address Harry Truman announced his "Point Four" program, described as "the first program designed with a truly Third World objective" (Kapur, Lewis, and Webb, p. 151), which laid the basis for America's post–Marshall Plan foreign assistance programs.

Expectations

When policymakers embraced economic development as a desirable objective of postwar economic policy in the late 1940s, what exactly did they have in mind, and what were their expectations? It is difficult to answer these questions in quantitative terms. "Development" was not precisely defined but during the postwar period was taken to mean improving economic opportunity by increasing production of goods and services in a lasting way, through capital formation (e.g., the provision of infrastructure, the early post-reconstruction emphasis of the IBRD), or through improved productivity. In short, development was associated with economic growth but did not include simple income transfers from one country to another. Instead, a recipient country's productive capacity should somehow be increased. It was assumed, however, that economic growth would improve nutrition, reduce mortality and morbidity,

increase longevity, and generally increase living standards, as has indeed generally proven to be the case.

It is worth recalling that while "gross domestic product" (GDP) is today readily familiar to the average informed citizen, national accounts were invented only in the 1930s and were still a relatively new idea beyond nonspecialists in the late 1940s. Quantitative historical work, mainly on the 1920s and 1930s, on aggregate economic output and hence economic growth, was still in its infancy. Policymakers felt they could identify some useful things that needed to be done in an economy without attempting to quantify them.

In *The Theory of Economic Growth,* published in 1955, W. A. Lewis, who later won a Nobel prize for his work on economic development, made only contingent quantitative statements—but they can be taken to reflect his expectations. Lewis states that "raising total output by 2 per cent per annum [in a country whose population is growing annually at 1.5 percent] is no mean feat. It requires considerable expenditure on education and other public services, a doubling of current capital formation, and many changes in beliefs and institutions." Three percent (in a country whose population is growing at 2 to 2.5 percent a year) would be even more difficult. "There is no sign of the less developed countries this side of the Iron Curtain beginning to adopt the sort of heroic measures which a 2 to 3 per cent per annum increase in output would demand" (Lewis, pp. 314–15). Lewis recognizes, however, that Japan doubled its per capita output in 25 years (implying a growth rate of 2.8 percent per year), so the possibilities are there for the rest of Asia and Africa (p. 316). The main requirement, Lewis believed, was to double the rate of capital formation, along with the production of skills required to do that successfully.

I have found only three relevant quantitative projections indicating that at least some of Lewis's contemporaries focused on future growth: Colin Clark (1942), who on a base of the late 1930s projected the world economy and its major components to 1960; the Paley Commission, which projected the period 1951 to 1975 for the U.S. economy, and implicitly for the world; and Woytinsky and Woytinsky (1953), who made quantitative projections of world population and energy use up to the year 2000.

Australian economist and statistician Colin Clark was one of the earliest users of national accounts and other aggregate national statistics to generalize about the process of economic growth, in his *Conditions of Economic Progress,* first published in 1940 (though the revised edition published in 1951 is most often quoted). Clark also wrote a less well-known sequel, *The Economics of 1960,* published in 1942 during the middle of World War II, in which he attempted to project to 1960 growth in population, labor force, productivity, and output for 30 to 35 countries or country groups from a baseline of the period 1935 to 1938. Clark was so interested in the details that he did not even publish world totals, but his projections implied world economic growth (national income measured in "international units" equivalent to U.S. dollars of 1925–34—a precursor to purchasing power parity calculations) of 3.3 percent per year, or 2.2 percent per capita. Interestingly, these projections were well in excess of prewar growth (and of what was probably the consensus view among economists at the time), but Clark assumed significant further recovery from the

Depression of the 1930s, especially in the United States, the world's largest national economy. Clark lacked data for Africa and much of western Asia, but it is interesting that in his projections only nine countries of that era—Finland, Portugal, Italy, the Baltic states, Poland, Czechoslovakia, Hungary, the USSR, and Japan—grow more rapidly than the United States in per worker income. Of course, all these countries were much poorer than the United States, but so were most other countries; China was the poorest among the countries studied by Clark, with about half the per capita income of India by 1960, which in turn had less than half the per capita income of the nine countries listed above.

Clark estimated both the capital requirements for his projected growth and the likely savings, and concluded that a shortage of savings might postpone his 1960 levels of output to 1966, thus lowering the aggregate growth rate to 2.6 percent per year and the annual growth in per capita output to 1.5 percent.

In reality, the world economy excluding the USSR and eastern European countries grew by 3.6 percent from 1938 to 1961, 2.0 percent per capita, higher in the postwar period (United Nations 1963, p. 156). Maddison (2001) estimated annual average GDP growth at 4.7 percent from 1950 to 1960, 2.8 percent per capita, or nearly double Clark's more probable estimate.

Concerned about the sharp rise in materials prices after the outbreak of the Korean War, U.S. President Harry Truman appointed a "materials policy commission" to examine future demand—for the United States and for the world—of natural resources, out to the mid-1970s (President's Materials Policy Commission 1952). This commission, known as the Paley Commission, projected annual growth of the U.S. economy at 2.8 percent from 1950 to 1975, and population growth at 1.0 percent. The outcome for both was significantly higher, 3.6 percent for gross national product (GNP) and 1.4 percent for population, while at the same time the consumption of most natural materials was overestimated, due to underestimation of materials-conserving technical change (for an analysis, see Cooper 1975). More pertinent here, the Paley Commission significantly *under*estimated materials demand in the world; given its overestimation for the United States, this implied a view of prospective world economic growth considerably lower than what actually occurred.

In a massive study of the world economy in the early 1950s, Woytinsky and Woytinsky (1953) made projections of world population to the year 2000, which they estimated would reach 3.25 billion in 2000, up from 2.4 billion in 1950 (an average growth of 0.61 percent per year) (p. 260); and of world primary energy use, which they "hypothetically" projected to be 6.0 billion metric tons of coal equivalent (btce) in 2000, up from 2.9 btce in 1950, or an average growth of 1.41 percent a year[1] (pp. 979–83).

Woytinsky and Woytinsky did not project gross world product (GWP) or any of its near equivalents, but their energy projection gives a rough idea of what they would have expected GWP growth to be over this period, had they projected it. Energy intensity—the amount of primary energy per unit of real output—typically rises in early stages of development (as agriculture is mechanized and manufacturing grows in relative importance), levels off, and then declines as manufacturing and

agriculture recede in relative importance (see Smil 2003, pp. 157–61). In the United States, for example, energy intensity rose sharply after 1880, peaked during the 1920s, declined until 1950, leveled off for two decades as automobiles and house-hold appliances became items of mass consumption, then resumed its decline in the 1970s. Analogous patterns exist for other rich countries. Thus the relationship of energy to GWP growth depends on the stage of development for each country, and for the world, on the relative importance and growth of poor as opposed to rich countries. A rough rule of thumb would be that overall primary energy use increases at the same rate as GDP. Based on this assumption, Woytinsky and Woytinsky implic-itly assumed that GWP would grow by about 1.4 percent over the coming half-century, which when combined with their assumed population growth of 0.6 percent per year implies an increase in world per capita income of 0.8 percent per year, modestly above Lewis's pessimistic expectation of 0.5 percent.

Outcomes

In fact, world fossil fuel plus primary electricity consumption grew by about 3.6 per-cent per year, from 60 exajoules (EJ) in 1950 to 355 EJ in 2000 (calculated from Smil 2003, p. 6), marginally below Maddison's GWP growth of 3.9 percent per year, of which 1.9 percent was population and 2.1 percent was output per person.

By the expectations of Lewis, Clark, the Paley Commission, and Woytinsky and Woytinsky, then, actual increases both in population and in output per capita turned out to be significantly higher than contemporaries expected in the late 1940s. When allowing for the fact that infant mortality declined, longevity increased, nutrition improved, and literacy increased (see table 1, and Thomas and others 2000), we can conclude that the actual performance of the world economy over the past half-century has been nothing short of spectacular relative to expectations at the begin-ning of the period.

Indeed, in the long stream of history, three features of the second half of the 20th century stand out: rapid economic growth; the sharp increase in population, from 2.5 billion in 1950 to over 6 billion at the end of the century; and extensive inflation,

TABLE 1.
Some Indicators of the Human Condition in the Developing World

Indicator	1960	1980	2000
Infant Mortality[a]	140	80	52
Life Expectancy[b]	43	59	64
Illiteracy[c]	53	43	28

Source: World Bank 2001.

a. Deaths per 1,000 births.

b. Years from birth.

c. Percent of adults.

TABLE 2.
Annual Increase in Per Capita GDP
(percent)

Region	1950–60	1960–70	1970–80	1980–90	1990–2001	1950–2001
U.S., Canada, Australia	1.7	2.9	2.1	2.2	2.0	2.2
Western Europe	4.2	4.0	2.5	1.9	1.7	2.8
Eastern Europe	3.8	3.4	3.0	−0.6	0.7	2.0
Former USSR	3.3	3.5	1.5	0.7	−3.5	1.0
Latin America	2.2	2.4	3.0	−0.7	1.3	1.6
Asia	3.8	4.1	2.9	3.2	2.9	3.3
Africa	1.9	2.5	1.2	−0.7	0.2	1.0
World	2.8	3.0	1.9	1.3	1.5	2.1

Source: Calculated from Maddison (2001, p. 330; 2002, p. 39).

with the U.S. GDP deflator increasing by a factor of 6, or 3.7 percent per year. The increase in population was made possible by improvements in material well-being, and in turn contributed to growth insofar as productive lives were both more numerous and longer. To what extent inflation, at least at modest rates, may also have contributed to growth is unclear.

Of course, the spectacular economic success of the postwar period was not uniform over time or across countries. World per capita income, from Maddison (2001, 2002), grew by 2.8 percent per year in the 1950s, rose to 3.0 percent in the 1960s, fell to 1.9 percent in the 1970s, and fell further to 1.3 percent in the 1980s, before rising to 1.5 percent in the 1990s (table 2). Thus while early performance far exceeded early expectations, expectations are presumably revised on the basis of experience; and against the experience of the 1950s and 1960s the last three decades have been disappointing. Indeed, in the late 1960s Herman Kahn and Anthony Wiener projected American per capita income to grow nearly three-fold (3.0 percent per year) from 1965 to 2000 (Kahn and Wiener 1967, quoted in Bell 1999, p. 461).

There were also regional differences. The richest economy, the United States, saw per capita income grow by 2.2 percent per year over the second half of the 20th century. Western Europe grew more rapidly, at 2.8 percent, while Asia grew more rapidly still, at 3.3 percent from a much lower base (table 2). However, Latin America grew "only" at 1.6 percent per year, while Africa produced only 1.0 percent—high by historical standards, but low by the standards we have learned are possible and that some have come to expect. Moreover, during the period 1990 to 2001 annual per capita income in Africa grew at only 0.2 percent, and Latin America at 1.3 percent (calculations from Maddison 2002).[2]

Measurement Issues

We routinely use long-term growth rates as though they are facts, and as though they represent reasonably good measures of improvements in material well-being.

In truth, these rates are problematic on both counts and represent only rough indicators—perhaps the best we have, but rough nonetheless. Three points need to be stressed. First, as environmentalists correctly point out, GDP is a measure of gross current output, and does not deduct for either deterioration in the environment (air and water quality) or for the depletion of easily accessible resources such as high-quality copper ore or hardwood forest. Allowance for such environmental deterioration may, for example, take away the recent modest gains in African per capita income.

Second, measuring output does not allow for changes in the real purchasing power of a country's output, that is, for changes in its terms of trade. It is conceivable that output per capita could rise yet real purchasing power fall if terms of trade have deteriorated enough (called immiserizing growth when such deterioration has been brought about by the growth in output). In fact, this qualification is probably most important for the oil-exporting countries that experienced major improvements in their terms of trade between 1950 and 2000, enough to convert stagnation or even declines in GDP per capita, as in Venezuela and Kuwait, into considerable improvements in living standards. For most countries, however, measured changes in the terms of trade are not sufficient to qualify greatly the changes in per capita output. The terms of trade of non-oil developing countries taken as a group, while showing some variation over time, worsened negligibly, by only 3.7 percent between 1964 (the first year for which such data are available) and 2000 (calculated from the International Monetary Fund's International Financial Statistics database). One reason for the modest deterioration in terms of trade is, perhaps, that during this period of time the export reliance of developing countries on primary products declined significantly, such that by the early 1990s more than half of the exports from these countries were manufactured goods. In contrast, oil prices in real terms rose by 3.6 times—averaging 3.5 percent *a year*—with ups and downs over the same period, increasing income relative to output significantly in countries where oil is a large fraction of exports and GDP.

Third, GDP is composed of thousands of individual products and services, which grow at quite different rates, or even decline while others are growing. GDP growth is a weighted average of these disparate growth rates, and the appropriate weights are likely to change over time, perhaps substantially. As an example, consider a simple economy with only two sectors, agriculture (A) and manufacturing (M). Suppose that A does not grow at all, and M grows uniformly at 10 percent per year. What is the economy's overall growth rate? If manufacturing accounts for 20 percent of national expenditure initially, we can say that GDP grows at 2 percent per year ($= .8a + .2m$, where a and m are the growth rates of the respective sectors). But after 5 years M will make up 29 percent of expenditures (assuming the economy is closed), and after 10 years 39 percent. Thus if we use expenditure weights from the fifth year rather than the initial year, GDP growth will be 2.9 percent per year; and weights from the tenth year will yield a growth rate of 3.9 percent—nearly double the first calculated growth rate!

How is this index number problem solved in practice? Usually by rebasing every five years or so, and then linking the five-year fixed weight rates to create a longer

time series. Indeed, the United States in the late 1990s adopted a chain-weighted measure of GDP growth, meaning that the weights change every year. For growing economies the weights at the end of half a century are very different from the weights at its beginning. Moreover, the weights at any moment in time may be heavily influenced by policy, especially restrictions on imports that raise significantly the domestic prices of protected goods. Thus in the illustration above, suppose that import protection raises M prices to the point at which initially they account for 30 percent of expenditures. The recorded growth rate with initial weights will then be 3 percent rather than the 2 percent reported above, and weights chosen five years later will lead to a recorded growth rate of 3.8 percent at domestic prices, rather than the 2.9 percent at world prices. Of course, protection of the slow-growing sector will have the opposite effect, reducing the measured growth rate.

The question of weights grows in importance the longer the period being examined and the higher the growth rate. Weights are also of great importance in making cross-country comparisons of levels of GDP or growth rates. Maddison (2001, 2002), whose data are used here, undertakes a major effort to achieve cross-country comparisons of both levels and growth rates in per capita GDP. He does this by using a particular variant of purchasing power parity (ppp) comparisons of per capita GDP for 1990, expanding this to GDP by multiplying by population in that year, and then extending the GDP so that it is calculated backward to 1950 (and earlier) and forward to 2001 using national growth rates, sometimes as standardized by international organizations (except for China, where he revises the rates significantly downward).[3] Regional and world totals are then achieved for any year simply by summing the national figures for that year; per capita GDP measured in 1990 international dollars is achieved for any year by dividing the totals by population. Regional and world growth rates in effect weight the national growth rates by 1990 national GDP calculated in 1990 international dollars (i.e., at a variant of ppp), which for example gives China roughly the same weight as Japan, even though at market exchange rates Japan's GDP was over seven times that of China in 1990. Different weights would produce somewhat different regional and world growth rates for, as noted above, countries grew at very different rates during the period, and therefore (for example), weights drawn from early in the period would have produced lower reported aggregate growth rates.

There is another potential problem with the Maddison figures. Since he compares the GDP of countries calculated at international prices of 1990, strictly speaking the national growth rates should also be calculated at international prices for full comparability. Apart from the tremendous amount of work involved, this would yield figures that would be hard to interpret, again because of the weights. Maddison's figures, like all ppp calculations, must choose weights for the individual prices that are being compared across countries. Expenditure weights are most often used, but these differ from country to country, especially between rich and poor countries, where, for instance, poor countries assign much greater weight to food in total expenditures than do rich countries. Convention is to take a geometric average of the expenditure weights of both countries in making bilateral comparisons. This is a purely arbitrary,

if reasonable, convention. Yet it makes a great difference. For instance, ppp for China is based on some 300 price comparisons between China and the United States in 1986, as calculated by Ren (1997) using some strong (and dubious) assumptions to make the goods and services comparable, especially with respect to quality. Using Chinese expenditure weights, which accord more weight to food and less to housing, Chinese per capita income at U.S. prices was $634; using U.S. expenditure weights it was $1,622, nearly three times larger—a significant difference for numbers that purport to measure the same thing. The geometric average of these numbers would be $1,014. Maddison's figure for 1986 is $1,597, in 1990 prices ($1,395 in 1986 prices), after adjustments he made to convert to his 1990 international dollars and to his re-estimation of China's GDP.

For comparing standards of living across countries, some version of ppp exchange rates for conversion of GDP in national currency is necessary. If we focus on world demand, however, market exchange rates need to be used, since they measure the ability of nationals to earn income or buy products or investments in the world market. GDP in terms of "purchasing power" is a meaningless entity, except as an intermediate step to living standards.[4]

Convergence?

A question of general interest is whether, during the past half-century, countries have converged in their standards of living—providing evidence of the "catch-up" hypothesis, whereby countries that were initially far from both the technological frontier and the best economic practice could in principle grow more rapidly than those closer to those frontiers. An extensive econometric literature has developed on this topic, essentially testing whether countries that were relatively poor in 1960 grew more rapidly than those that were relatively rich. This is not the place to review that literature, except to note that the general finding is one of "conditional convergence," that is, initially poor countries grew more rapidly than rich countries, conditional on a number of factors such as life expectancy (as a proxy for general health) or educational attainment (see e.g., Barro 1997).

But I would not have expected to see, over a period as short as 35 years (the rough time frame of these studies), convergence in the sense sought. Many human and indeed biological processes follow a logistics curve, which does imply ultimate convergence but only after initial divergence—and only if the most advanced parties stand still once they have made the adjustment. But the technological frontier, so far anyway, has been constantly expanding, so "best economic practice" is constantly changing—as concretely manifested in the fact that real per capita income in the United States, the richest large country, continues to grow. Thus the "convergence target" is not static but constantly moving. This poses severe challenges for any country to "catch up," even though the potential for catch-up is present. Initially poor countries could still be expected to grow more rapidly, but the catch-up period might indeed be very long.

FIGURE 1.
Country Distribution of Growth in Per Capita Income, 1950–98
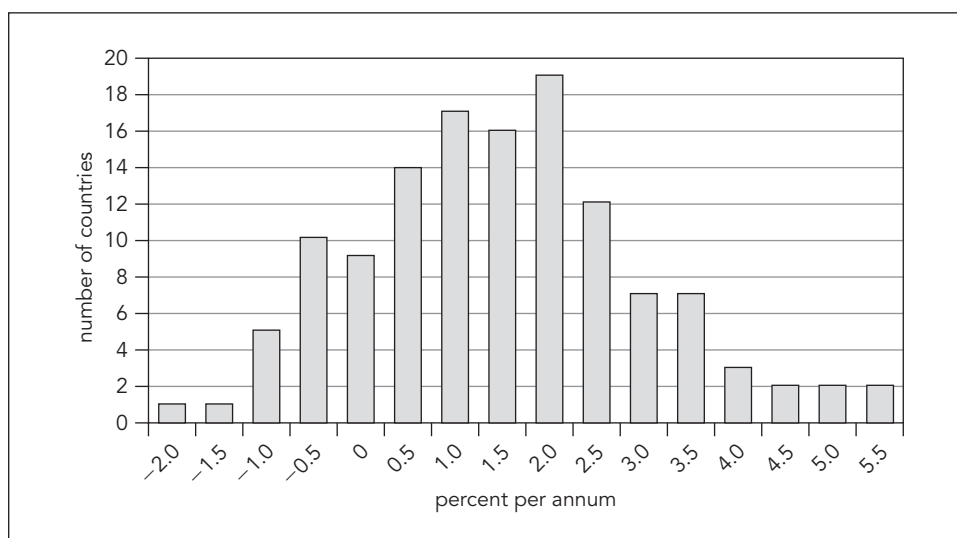
Source: Author's compilation based on Madison (2001).

Figure 1 presents the distribution of national growth rates in per capita income from 1950 to 1998, for 128 countries or groupings of small countries, calculated from Maddison (2001). Per capita income in the United States, initially the wealthiest country (at $9,561 in 1990 international $), grew at 2.2 percent per year, which can be regarded as a benchmark. Many countries grew more rapidly than the United States, but more grew more slowly. Growth in GDP would show more countries growing more rapidly than the United States, because of rapid population growth. But our main concern here is with improvements in material well-being, proxied by growth in per capita output. The growth rates broadly reveal a bell shape, with growth concentrated in the 2.0 to 2.5 percent interval, and tapering away both below and above. Eighteen countries actually showed a decline in per capita income, although the sharpest declines were for Qatar and Kuwait, which as noted earlier experienced large increases in real income due to improvements in their terms of trade. If we exclude the oil-exporting countries, the declines are led by Cuba (−0.9 percent a year), Niger (−0.9), Djibouti (−0.7), Madagascar(−0.7), Haiti (−0.5), and Afghanistan (−0.5). Fourteen countries experienced some decline in per capita income if we exclude the oil exporters, nine of which were in sub-Saharan Africa. Most of these countries experienced considerable internal conflict, sometimes outright civil war.

At the other end of the scale, four economies recorded an extraordinary increase in per capita income in excess of 5 percent per year over this half-century: South Korea (6.0), Taiwan (5.9), Botswana (5.3), and Oman (5.3), all of which were very poor in 1950. Indeed, Botswana was the world's poorest country in 1950, with a per capita income of $349 (in 1990 international $) as calculated by Maddison, followed by Tanzania, Burma (now known as Myanmar), and the behemoth China. But per capita income growth for most countries is clustered in the intervals 1.0 to 3.0 percent.

While "a rising tide lifts all boats" is a nice metaphor, it is not an accurate characterization of most human affairs, which more commonly involve leaders and followers, innovators and imitators, with imitators demonstrating mixed capabilities in following successfully, as well as occasionally altering the path significantly (sometimes productively).

As noted earlier, growth rates declined in the later decades of the 20th century. The distribution of growth rates across regions also shifted. In the 1950s and 1960s it was mainly northern European countries that grew more rapidly than the United States; thereafter their relative growth slowed. By the 1970s the rapid growth shifted to southern Europe and to east Asia, beginning with Japan, then shifting to southeast Asia and more recently to India. A few high-growth countries were spread more widely, including Israel and Palestine, Saudi Arabia (to 1980) and Tunisia in the Middle East, and Mauritius and Swaziland in Africa. Puerto Rico holds the record in the Western Hemisphere, followed by Trinidad and Brazil (mainly to 1980). A host of other countries, such as Egypt, Mexico, Pakistan, Turkey, and many smaller countries, have made steady if less spectacular progress. Eastern Europe and the Soviet Union grew respectably, although at declining rates from 1950 to 1980, but then experienced declining output for a period as their economies were transformed from central planning and control to market-oriented, with a major transformation in the composition of output.

Thus on these figures, there has been both convergence (with the U.S. as benchmark) and divergence. Many countries, especially those in western Europe and east Asia, have reduced greatly the (geometric) gap in GDP per capita. We too often forget how poor some currently rich countries were 50 years ago. On Maddison's comparisons, Japan in 1949 had a lower per capita income than India did in 1998, and Greece, Portugal, and Spain were only modestly richer. South Korea had a per capita income equivalent to that of India in 1965, and only 24 percent higher than India in 1950.

On the other hand, many countries have fallen further behind than they were in 1950. Indeed, African countries, while poor in 1950, were on average nearly twice as rich as east and south Asian countries (excluding Japan). But their low average growth rates have seen them recede considerably, such that by 1998 their per capita income, while 60 percent higher than in 1950, was less than half that of the east and south Asian countries on the same comparison (Maddison 2001, pp. 305, 327). Haiti was 36 percent richer than South Korea in 1950; by 1998 South Korea was 16 *times* richer than Haiti.

Figure 2 records the distribution of national growth rates in per capita income, but weighted by population in 1975 rather than simply by number of countries. There it can again be seen that growth rates averaging between 2 and 2.5 percent per year are the most common, but the distribution is trimodal (due mainly to China and the former Soviet Union), with the preponderance of growth taking place at rates higher than the modal interval.

Using countries as observations, weighting by population, the income of the median world citizen has risen significantly relative to the income of those at the 90th percentile. But those at the 20th percentile have fallen further behind both

FIGURE 2.
Per Capita Growth Weighted by Population
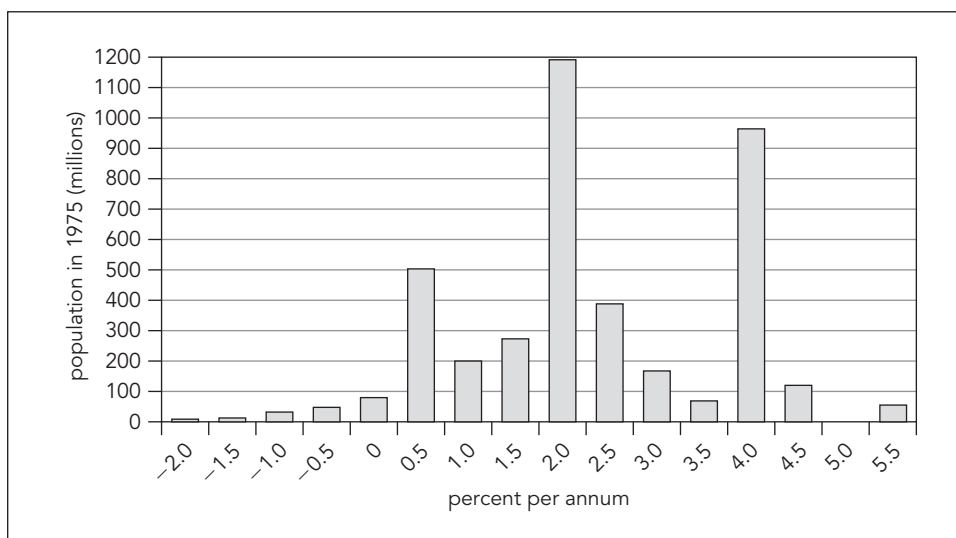
Source: Author's compilation.

(Cooper 2002, pp. 133–34). This calculation is crude, however, compared with those that allow for distribution of income within as well as among countries, as discussed below. To compare the average income in the 20 poorest countries with that in the 20 richest countries, as is sometimes done, is of course deeply misleading, since both groups change in composition, substantially over long periods of time. As noted above, both Botswana and China were among the 10 poorest countries in 1950, but neither would be included in that list today.

It should be noted that recorded GDP growth rates probably understate the true potential growth in rich, especially European, countries, since higher incomes have permitted greater leisure in the form of shorter work weeks and more paid holidays. In this respect Europeans have sensibly elected to take more leisure (which of course is not measured in GDP) than have Americans, who have in turn taken more leisure than the Japanese, who reputedly often do not even take the paid vacations to which they are entitled. By the same token, recorded GDP growth rates probably overstate the true potential growth in poor countries, as many family activities, such as making clothing, are increasingly absorbed into the commercial economy and hence add to recorded GDP, whereas these activities were not counted before. (Food self-sufficiency, in contrast, does in principle enter GDP, although how well that is done in practice is open to some doubt.)

At a very fundamental level, development is moving people out of agriculture into socially more productive activities, as productivity in agriculture also increases. In the poorest countries, those near subsistence, more than 80 percent of the labor force is engaged in agriculture (including fishing). In the richest countries, less than 5 percent is engaged in agriculture (2 percent in the United States)—and sometimes even that

low share produces a surplus for export. In this simple but fundamental dimension the past half-century also experienced great change: Twenty to 30 percent of the (rising) labor force moved out of agriculture between 1965 and 2000 in most developing countries for which data are available. For example, the share of the labor force in agriculture declined over this period, from 84 to 63 percent in Bangladesh (a country of low growth), 49 to 21 percent in Brazil, 81 to 47 percent in China, 55 to 29 percent in Egypt, 71 to 43 percent in Indonesia, 50 to 18 percent in Mexico, and 82 to 49 percent in Thailand, to illustrate with large developing countries; and from 47 to 17 percent in Greece and 26 to 5 percent in Japan, to illustrate with OECD countries. That paragon of growth, South Korea, reduced its share from 55 to 11 percent over the same period, even while strongly protecting agriculture late in the period.

Poverty

In a sense, the whole point of economic growth is to reduce poverty—to create opportunities for development of individual interests and talents that simply cannot be cultivated when nutrition and health are poor, and when one's dominant preoccupation is with providing sufficient food for oneself and family. Yet comparisons of "poverty" across countries, or even over time within countries, are fraught with conceptual and practical difficulties. What exactly do we mean by "poverty," and how sensitive should it be to the general economic (and social) conditions of the society in which one lives? Having decided on a working definition, how do we measure poverty accurately?

In terms of international discourse, the conceptual problem was "solved" by the 1990 World Bank suggestion that poverty in developing countries should be defined as income below $1 per day per person, measured in 1985 purchasing power parity dollars—drawing on the national experience of India, which along with the United States had pioneered the official measurement of poverty in the 1960s. Research was subsequently devoted to discovering the level of poverty that existed in many countries when measured by this standard, how much it had changed over time, and what were the principal determinants of those changes.

Much controversy surrounds the many technical aspects of these comparisons. I will draw heavily on recent work by Surjit Bhalla (2002, 2003), an Indian economist who once worked at the World Bank, and who reviews the history and the technical controversies and provides his own comprehensive estimates of poverty based on the World Bank definition, by region, over the period 1950–2000. A key feature of Bhalla's work is that he focuses (appropriately) on people rather than countries. To do this he needs data on the distribution of income or (preferably) consumption within each country, and changes in such distribution over time. He then needs to aggregate these people-oriented data across regions and the world for a number of years, to discover trends in world poverty.

Table 3 presents poverty rates in what we today call the developing world, as calculated by Bhalla (2002, p. 148), by region for the turn of each decade from 1950 to

TABLE 3.
Poverty in the World, 1950–2000

Region and measure	Poverty line (ppp, US$1.50 per day)					
	1950	1960	1970	1980	1990	2000
Head count ratio (percentage)						
East Asia	86.6	77.5	71.1	67.2	31.3	6.0
South Asia	44.3	37.2	32.1	34.4	18.5	7.8
Sub-Saharan Africa	59.3	53.2	52.2	49.9	55.3	54.8
Middle East and North Africa	26.3	24.3	13.4	4.3	5.2	7.8
Latin America	22.0	16.0	9.4	3.6	5.3	5.2
Eastern Europe	17.8	9.2	3.3	1.7	0	0
Developing world	**63.2**	**52.5**	**46.4**	**43.5**	**25.4**	**13.1**
Number of poor people (millions)						
East Asia	830	729	833	955	521	114
South Asia	208	209	229	310	207	105
Sub-Saharan Africa	104	118	150	188	279	362
Middle East and North Africa	27	32	23	10	16	29
Latin America	36	35	27	13	23	27
Eastern Europe	49	29	12	7	0	0
Developing world	**1,223**	**1,131**	**1,262**	**1,479**	**1,056**	**647**

Source: Bhalla 2002, p. 148.

Note: ppp = purchasing power parity.

2000, along with a head count of the people who live below the poverty line. The poverty line Bhalla uses is $1.50 per day in 1993 ppp dollars ($547 a year, or roughly the equivalent of $505 in Maddison's international $ of 1990), to allow both for inflation that occurred after 1985 and for some undercounting of consumption (or income) of high-income families in the national surveys that everyone relies on for information on income distribution.[5]

In table 3 it can be seen that poverty dropped substantially, even dramatically, over the past half-century—from 63 percent of the population in the developing world in 1950, to 13 percent in 2000. The drop was particularly significant, 30 percentage points, over the last two decades, the period of so-called globalization. Significant declines occurred in every region except sub-Saharan Africa, and even there a modest decline occurred, although an increase was registered after 1980. Furthermore, the *number* of people living in poverty was roughly halved, from 1,223 million to 647 million, despite a more than doubling of world population. Again, the number declined in most regions, except for a modest increase in the Middle East and North Africa, and a dramatic increase of over 300 million in sub-Saharan Africa, where population growth was especially rapid.

The details of Bhalla's work are controversial, particularly with regard to his use of consumption data from each country's national accounts to determine average

consumption levels, which are typically higher than those reported in social surveys. Moreover, the proportionate gap seems to have risen over time. Since at least some of this difference can be attributed to presumed underreporting by relatively wealthy people in the surveys, Bhalla adjusts the poverty line upward from $1.30 per day in 1993 prices (the equivalent of $1 per day in 1985 prices) to $1.50 per day, to allow for his reliance on the national accounts.

Debate has focused particularly on India, an important developing country where the data are relatively rich and well-developed. Both the national social survey (S) and the national accounts (NA) have strong supporters, yet there is a significant and growing discrepancy in average per capita consumption between the two, with the S/NA ratio having fallen from 93 percent in 1973 to 56 percent in 1999, or from 71 percent to 56 percent if allowance is made for rebasing and other revisions in the national accounts (Bhalla 2003, table 4). This is not the place to take a position in this debate, which pertains also to a number of other countries. But if the survey data are valid, such discrepancies raise serious questions about the reported growth rates for the Indian economy over the past three to five decades. If mean consumption in the 1990s is as low as the surveys report, unrealistically low consumption levels several decades earlier are implied if the reported growth rates are accurate. Mean survey data, corrected for Indian inflation, show a growth of only 0.6 percent a year in per capita expenditure over the period 1973–99, whereas per capita real consumption from the national accounts grew by 2.6 percent (calculated from data in IFS).

An alternative explanation is that the growth in consumption India experienced accrued overwhelmingly to the rich, who are not adequately reflected in the survey means. But this would have implications for the distribution of income that are implausible, if not literally impossible. If the top 10 percent accounted for 25 percent of total expenditures in 1973, as they reportedly did in 1993 (Sundaram and Tendulkar 2003a, p. 333), then accuracy in both the survey means and the national accounts consumption would imply that the top 10 percent accounted for 54 percent of total consumption by 1999.[6] We can conclude from this calculation that the survey significantly underestimates the growth in average consumption over the last three decades, or that the recorded GDP growth rates are significantly too high, or that the distribution of income in India has become much more unequal than is generally acknowledged—or some combination of all three. Such conundrums in data consistency have (negative) implications for all cross-country work on growth and inequality, and on poverty.[7]

Whatever the detailed resolution of the technical disputes, it seems likely that poverty in developing countries has dropped significantly, especially in the past two decades. Sundaram and Tendulkar (2003a; 2003b, table 7), strong supporters of the Indian Survey as opposed to the national accounts, nonetheless report a drop in Indian poverty by 8 percentage points between 1983 and 1993, and by a further 5 points between 1993 and 1999. Even more rapid growth occurred in China from 1980 to 2000, and while it was uneven, it was sufficiently widespread to reduce poverty substantially. Indeed, the drop from 67 percent in 1980 to 6 percent in 2000 (shown in table 3 for east Asia) is dominated by China. Given their huge populations

and initially high poverty, rapid growth in China and India alone virtually assures a worldwide reduction in poverty, even with widening income distribution in each. To focus on countries rather than people, and on the poorest countries alone, is deeply misleading with regard to what has happened in the world at large.

A reduction in poverty is generally associated with economic growth, not just in China and India. In fact, it is difficult to find examples of significant reduction in poverty that is not associated with economic growth, and it is difficult to find significant growth that does not reduce poverty (see Dollar and Kraay 2002a, 2002b).

What is quite separate from the issue of poverty, although remotely related, is the claim that inequality in the world distribution of income has increased significantly in recent years. It is usually simply assumed that greater inequality is undesirable, and therefore that an increase in inequality should be a cause for concern. Yet it is difficult to imagine economic growth starting from a stationary condition that does not for awhile, perhaps a long time, increase inequality. As noted earlier, many human activities follow a logistic pattern, which implies an increase in inequality (of whatever), followed by an eventual decline. The circumstances surrounding any increased inequality are all important in evaluating whether it is desirable or undesirable.

Bhalla's work on poverty also permits him to address the issue of inequality, again by focusing on people rather than countries. And his results sharply contradict the conventional findings (based on countries) of growing world inequality. Using Gini coefficients as a measure of inequality (ranging between zero, perfect equality, and one, extreme inequality), Bhalla (2002, p. 178) finds that between 1960 and 1980, when rich countries were growing rapidly, world inequality increased; but between 1980 and 2000 world inequality decreased to below where it was in 1960. This occurred despite an increase in inequality in the industrialized world, as Europe and Japan slowed relative to the United States. If China and India are excluded, inequality was virtually unchanged between 1980 and 2000 in the developing world, but with China and India included (as they should be), inequality declined sharply. This could occur even if, as is usually assumed, income inequality became less equal within China and India.[8]

On a regional basis, inequality was highest in sub-Saharan Africa, and unchanged from 1980 to 2000, whereas it declined in east Asia, the Middle East (incuding North Africa), and Latin America, while increasing modestly in south Asia and sharply in eastern Europe.

An alternative measure of inequality is to compare the median (50th percentile) income in the United States (as the world's richest large country) with the median income in the developing world. According to Bhalla's calculations (2002, p. 179) that ratio declined almost steadily, from 24.1 in 1950 to 11.7 in 2000. In short, the median person in the developing world, while still much poorer than his U.S. counterpart, is nonetheless catching up, having more than halved the (geometric) gap over the past half-century. A similar result holds if the comparison is made between those at the 20th percentile in the United States and in the developing world. However, the median person in Africa, having been modestly better off than his east Asian counterpart in 1960 in comparison with the median American (a ratio of 22.8 versus

25.6), was far worse off by 2000 (38.3 versus 9.0), showing a deterioration both with respect to the median American, and especially with respect to the median east Asian (Bhalla 2002, p. 192).

Capital Inflows

As noted earlier, Arthur Lewis judged the main requirement for raising growth to be a sharp increase in the rate of capital formation, physical and human. Investment ratios did rise in developing countries, but they crept rather than leapt up, rising from 20.6 percent of GDP in the 1960s to 22.8 percent in the 1970s, 25.6 percent in the 1980s, and 26.3 percent in the 1990s (calculated from the International Monetary Fund's International Financial Statistics database: China was added to this database in 1979, Russia in 1994). During the same period of time investment ratios declined modestly in the industrialized countries, from 23.7 percent in the 1960s, to 20.8 percent in the 1990s.

One of the advantages of engagement with the world economy is being able to draw for investment on savings elsewhere in the world; a country does not have to rely on domestic savings alone. The transfer of real resources to any country can be measured through the current account in its balance of payments. In fact, the current account deficit for developing countries as a group, as defined in the 1980s, exceeded 1 percent of their GDP (measured in 1990 international $) only in 1982, a year of world recession, and approached 1 percent in only a few other years (e.g., 1978, 1991, 1993) and a few additional years if the large surpluses of Organization of Petroleum and Exporting Countries (OPEC) members in those years are excluded. The current account deficit was well under 0.5 percent in most years. As a share of domestic investment, these percentages must be increased four- to five-fold, but still remain remarkably small.

Of course, there were large variations from country to country, and even for the same country over time. Net capital inflows were significant, for instance, for Israel and South Korea in the 1960s and 1970s, and represented 2 to 10 percent of GDP for Chile, Costa Rica, and a number of smaller countries in the 1980s and 1990s, and 0.5 to 2.5 percent of GDP for India. Moreover, by virtue of the international capital market a country could engage in intertemporal reallocation of large foreign earnings in some years.

Bilateral and multilateral aid to developing countries was $6.7 billion in 1962, as recorded in the first report of the newly formed Organisation for Economic Co-operation and Development's Development Assistance Committee (1963), or $25.7 billion in 1990 dollars (using the U.S. GDP price deflator), a little more than 1 percent of the GDP of developing countries in that year. It doubled in real terms to the early 1990s, and then receded by 2001 to only 20 percent above the 1962 level, of course representing a much smaller fraction of developing country GDP (World Bank, GDF 2001, p. 87). Foreign assistance was about three times private capital flows (at $2.2 billion) to developing countries in 1962, but by the 1990s was dwarfed by private capital flows, which exceeded $160 billion in 1996, the peak year.

The contribution of foreign capital inflows to development is still a source of uncertainty and controversy (some evidence is summarized in Cooper 2002). Foreign direct investment (FDI) seems to be the most potent in its impact, and even then seems to work best when local education is high. By 2001 FDI to all developing countries exceeded official development assistance (ODA) by nearly five times (World Bank, GEP 2004, p. 292).

Foreign aid has been the most visible vehicle for help to developing countries from rich countries over the past four decades. It is therefore worth asking what we can say about the impact of such assistance on economic growth. A World Bank team supervised by Lyn Squire addressed this issue in *Assessing Aid* (Dollar and Pritchett1998). The study correctly observes that aid is fungible in that it may finance a project or activity that the government would have undertaken with its own funds, so the true marginal expenditure from aid need not be the designated expenditure. Only about 35 percent of aid, on average, increased the investment ratio in a study of 56 countries (42 of which were in Africa or Latin America) over the period 1970–93; none however seemed to finance tax reductions, and some aid is initially targeted on government consumption such as education, agricultural extension, or public health. After controlling for other variables, the 1998 study suggested that foreign assistance as a percent of each recipient's GDP had *no* discernable impact on its economic growth over this period, an appalling result even after allowing for the fact that some aid was not targeted for development. However, aid interacted with economic management (measured as a weighted average of trade openness, inflation, budget surplus, and institutional quality) in such a way that growth in well-managed countries benefited from aid: On average, an increase in aid by 1 percent of GDP increased a well-managed country's growth rate by 0.5 percentage points. And in some well-studied cases foreign assistance was extremely important in launching economic reform and subsequent economic growth (see Haggard, Cooper, and Moon 1993, on South Korea in the 1960s).

Today it is understood that much, perhaps even most, of the benefits of foreign assistance and FDI arise from the technical and managerial knowledge they convey rather than from the provision of capital as such.

Growth, Development, and Freedom

"Development" was earlier associated with increases in per capita GDP, as a rough proxy for improvements in material well-being. A case can be made that this is a necessary—or at least a strongly facilitating—condition, but not a sufficient one. If we conceive, as Sen (1999) does, of "development" as increasing the capability of all human beings to achieve the things that they value most, then development must also cover the ability of citizens to express themselves and to have some voice in the policies that affect them, that is, development involves civil and political liberties as well as improvements in material well-being.

One of the remarkable trends of the past half-century was the extension of functioning democracy to a wider fraction of mankind, including not just northwestern

Europe and North America, but Germany and Japan early in the period, and Portugal and Spain in the mid-1970s. By the year 2000 democracy spread to include eastern Europe, Russia, most of Latin America, much of east, south, and southeast Asia, and parts of Africa. The major remaining gaps were in the Arab world, central Asia, China, and sub-Saharan Africa (where the situation deteriorated significantly following independence in the 1950s and 1960s). The extension of democracy during the 1980s was what political scientist Samuel Huntington called the "third wave," suggesting that the phenomenon is episodic, not linear, and occasionally even involves reversals.

There has been extensive scholarly discussion over the years on the relationship between freedom, of which functioning democracy is a major manifestation, and economic development: whether the latter is a precondition for the former, and whether the former is a serious impediment to the latter. Huntington (1991, p. 311) is a strong proponent of the first view: "Poverty is a principal and probably the principal obstacle to democratic development. The future of democracy depends on the future of economic development. Obstacles to economic development are obstacles to the expansion of democracy." Robert Barro's empirical analysis (1997, ch. 2) supports Huntington's view. Modern India, with a functioning, indeed raucous, democracy since its beginnings in 1947, of course stands out as a leading exception to this generalization, as do a few Latin American countries; but a generalization does not have to be universal to have some validity.

Has democracy been an impediment to economic growth? In a study of 18 developing countries during the 1970s and 1980s, Little and others (1993) found no relationship between economic performance and democracy or the Freedom House ranking on freedom for each country. Helliwell (1994) supports that conclusion on the basis of wider country coverage over a longer period of time. Barro (1997, ch. 2) finds a weak curvilinear relationship, with growth depending positively on what he calls an "index of political rights" up to the level achieved by Malaysia and Mexico in 1994, and negatively for higher levels of rights.

While it is true that examples can be found in which authoritarian government seems to facilitate economic reform and subsequent economic growth—South Korea in the 1960s and 1970s comes to mind—it is also true that many authoritarian governments have neither the desire nor the capacity to undertake economic reforms leading to higher economic growth. Democracies are usually more hesitant and less comprehensive in adopting economic reforms, but also steadier in responding to various forms of turbulence in the process.

Lessons Learned, 1950 to 2000

What conditions are especially conducive to growth, and what conditions are especially detrimental? Geography, institutions, openness, market orientation, rule of law, government micro-engagement, heavy taxation, corruption, and macroeconomic management have all been put forward as serious explanations for inter-country differences in economic growth. Often the debate has focused on dichotomies (e.g.,

market versus planning, or inflation versus price stability), whereas reality was full of grays that suggest the coarse dichotomies of public debate offer little practical guidance.

Being "geographically disadvantaged" has been advanced as a reason for poor economic performance, especially in countries that are landlocked, such as Bolivia, Laos, Nepal, Paraguay, and many African countries. But Botswana, by far the most rapid grower in sub-Saharan Africa, is also landlocked, as are Austria, the former Czechoslovakia, Hungary, Switzerland—and indeed Chicago, if the U.S. Midwest were treated as a separate economy. Perhaps being landlocked creates some modest disadvantage—although being on the sea does not ensure good natural ports—but geography can be overcome through human agency. Chicago can now entertain oceangoing ships (although not the largest) through the St. Lawrence seaway, which once presented huge natural obstacles such as Niagara Falls.

Note also that the Dominican Republic and Haiti share an island, a climate, and equal access to foreign markets; and in 1950 they had equal per capita income of around $1,050 (1990 international $). By 1998, however, income in the Dominican Republic had increased by a factor of three; Haiti's had fallen by 20 percent. Institutions and policies played a major role in the differences.

The best performing economy in Latin America was Puerto Rico, with a growth in per capita income of 3.9 percent per year over 50 years. It started in 1950 with a per capita income somewhat greater than that of many of the smaller Latin American countries, but lower than that of all the larger Spanish-speaking countries. Although Puerto Rico may be dismissed as being part of the United States, it grew much more rapidly than the United States did. And while it is part of the U.S. currency area and customs zone, it has wide autonomy in framing its local policies, including education and taxation. Puerto Rico's superior performance calls for explanation. Neighboring Cuba, once among the richest countries in Latin America, is now one of the poorest, exceeding only Nicaragua and Haiti. Anti-growth policies governed the country. The American embargo of 1960 was an immediate blow but can hardly explain four decades of poor performance, despite Castro's continuing use of it as an excuse—not least because no other country supported the embargo. Comparisons between East and West Germany, and between North and South Korea, make the same point. Political stability prevailed for decades in both pairs of states. What differed were the incentives for individual effort and risk-taking, which are the mainstays of continuing economic growth.

One important dimension of geography is the presence of disease, for agricultural products as well as humans. While small children can be replaced, high child mortality involves a serious opportunity cost to the mothers, and of course any education of the child is lost. Debilitating diseases, representing equilibrium between parasite and host, may be worse than fatal disease from the perspective of economic growth, because of the deleterious effects of debilitation on productivity. Again, human agency can overcome or greatly mitigate the impact of disease; southern Europe was once malarial, as was Singapore. Smallpox was once endemic in Europe, resulting in high mortality. The Panama Canal could not have been built before it was learned

that yellow fever-bearing mosquitoes had to be kept at bay. But the obstacles to be overcome are surely higher in some places than in others.

Hong Kong and Singapore both registered outstanding growth, over 4.5 percent a year for half a century. The relevance of their experience is often dismissed, as mere "city states." But Hong Kong has a population greater than that of many European countries, such as Denmark, whose experience is not casually dismissed; and Hong Kong's GDP registers at 22nd in the world, with the vast majority of economies being far smaller. The main differentiating feature in Hong Kong and Singapore is that agriculture (apart from fishing) was not a dominant part of either economy, even when they were poor. There are, however, other potential "city states," such as Djibouti, that would be much better off if they had performed only half as well Hong Kong or Singapore. And while the transformation of agriculture is an important part of development, it is not the only important part; lessons can be learned even from city states.

It has become conventional wisdom that openness—that is, some form of serious engagement with the world economy—is a significant contributor to growth. Barro (1997) could not find such a relationship for the limited measures he used, and Rodrik and Rodriguez (2000) have criticized on methodological grounds the several studies usually cited in support of this conclusion, including Sachs and Warner (1995). Cooper (2002) argues that there is little theoretical ground for expecting a priori that openness would contribute to growth, as opposed to a once-for-all increase in income, but that in the actual post-1950 world it would be implausible if continuing trade liberalization, in particular, had not contributed to growth. Warner (2003), responding to the methodological criticisms of Rodrik and Rodriguez (2000), has demonstrated persuasively that relative openness, measured in a variety of different ways and formulated with a variety of specifications, was strongly conducive to growth, at least for the 61 to 86 countries he studied over the period 1970–90. Stated differently, those developing countries (with per capita income under $5,000) that remained relatively closed grew far more slowly, on average, than those that engaged more actively with the world economy. GDP growth in the latter group averaged 4.8 percent a year, versus only 0.8 percent in the former group. There were no countries in the open group that experienced negative growth, in contrast to the relatively closed economies. After controlling for many other variables, Warner registers the judgment that being "open" rather than "closed" is worth 3 to 4 percentage points of annual growth in GDP—a tremendous impact.

Even if one accepts this general dichotomous result, each country is different, and the practical details are all important.

Remarkably many small economies, with populations of fewer than one million people, are high-income economies (GDP/P > US$9,266 in 2001). True, many are dependencies, albeit locally autonomous ones, of rich countries: Bermuda, Faeroe Islands, French Polynesia, Greenland, Guam, and Northern Marianas. Some are independent but contiguous to larger rich countries: Andorra, Luxembourg, San Marino, and (arguably) the Bahamas. But many meet neither of these conditions: Barbados, Brunei, Cyprus, Iceland, Macao, and Malta. These economies are not usually included in analyses of economic performance, as being beneath notice. But why? While some are "dependencies," they have a high degree of local autonomy in setting both

economic policies (e.g., taxation) and local laws that are relevant to commerce and finance. Dependency these days usually, in practice, refers to defense and foreign policy.

A few of these economies, such as Bermuda and Iceland, were relatively well-off 50 years ago, but most were poor. They have become rich. Many had a colonial past, usually but not always British. They inherited or developed a legal framework, institutions, and policies that have helped them achieve relative wealth. All are open, and many have developed close economic relations with larger rich countries.

Not all small countries are rich. Among the low-income countries (GDP/P < US$745) are São Tomé & Principe ($280), Comoros ($380), Solomon Islands ($580), Bhutan ($640), and Equatorial Guinea ($700)—all countries in poor neighborhoods. All except Bhutan and Solomon Islands are in Africa. But these very poor small economies are greatly outnumbered by the rich ones. Small countries tend to maintain fixed exchange rates or even to use a larger country's currency and impose import tariffs primarily for revenue, not to protect local industry. Only Brunei (oil) and Iceland (fish) are rich in resources, although São Tomé and Equatorial Guinea seem to be well-endowed with offshore oil, soon to be developed. In short, many small economies have managed to grow significantly despite their small scale, or perhaps because they were too small to entertain pretensions of economic independence.

Inflation is said to be inimical to economic growth (e.g., Fischer 2004, ch. 10, especially p. 336; also Barro 1997, ch. 3). That is shown by much cross-country empirical analysis, and at least for high rates of inflation poses no mystery: Those are usually cases in which the budget has gotten out of control and inflation is the residual equilibrator. High inflation is also closely associated with highly variable inflation, rendering forward-looking financial planning—a key and necessary feature of all modern economies—practically impossible.

But it does not follow from these correct observations, as is often assumed, that moderate and controlled inflation (e.g., in the low double digits) is also inimical to growth. For a country without a well-developed financial market, inflation may in fact be less distorting than many other taxes the government might impose to generate needed revenues, including high taxes on imports; and the "inflation tax" can reach parts of the population that are not otherwise reachable by the tax authorities. Whether this is good or bad for growth of course depends inter alia on what is done with the revenue. South Korea's annual inflation rate averaged 14 percent during its takeoff period 1966–79, for instance. It would require much higher confidence in one's theories than I have to argue that Korea would have grown even more rapidly had it forced its inflation rate into low single digits, as seems to have become the accepted norm, even for developing countries, in official financial circles.[9]

Most analyses of growth, as here, have been based on examination of aggregate economic statistics and their possible determinants. McKinsey Global Institute has since 1990 undertaken a series of "bottom-up" industry studies in selected developing countries—mainly Brazil, India, Korea, and Russia—as well as developed countries, focusing on the detailed obstacles to achieving best practice productivity in the use of labor and capital, recognizing that best economic practice will be sensitive to

relative factor prices. These studies provide the basis for Lewis's (2004) unsurprising finding that there is no single impediment to adoption of best practice, enhancing productivity and hence living standards; instead, a host of often-interacting obstacles is found. He argues, however, that serious competition in domestic markets is usually a necessary condition for adoption of improved techniques, and that serious domestic competition is unlikely to occur without a national mindset that accords high value to consumers as opposed to producers. It is also unlikely to occur without significant engagement with the world economy, where best practice is typically generated and is constantly changing, both with regard to cost and price, and with regard to product quality. Lewis also argues that economically intrusive government is all-but-inevitably captured by special (producer) interests and hence is typically a serious impediment to competition, hence advancing productivity.

What If?

Many events occurred during the past half-century, or did not occur but arguably should have occurred, that have been claimed to impede economic development in poor countries. This field is a vast one, but a few such counterfactuals can be briefly explored:

1) Suppose access to markets of rich countries had been open earlier, especially in agriculture and apparel, as has been urged for the ongoing Doha Round of World Trade Organization negotiations; how much difference might it have made to overall economic performance?

2) Suppose rich countries had actually met the United Nations target of 0.7 percent of donor GDP for foreign aid set in the 1960s; how much difference would it have made to economic growth?

3) Suppose private capital flows had been steadier; how much difference would it have made?

4) Suppose OPEC had not quadrupled oil prices in 1974, causing the deep 1975 recession and subsequent buildup of sovereign debt?

Now for some speculation on these counterfactuals:

1) *Access to markets*. Import restrictions in rich countries have recently been portrayed as serious obstacles to the development of poor countries, denying them both higher income and growing markets. While it is difficult to defend high import restrictions in rich countries, the main impact of these restrictions probably falls on the consumers of the rich countries rather than on developing countries, with a few possible exceptions. It can even be argued that the restrictions in the textile and apparel sector, envisioned in the 1974 Multifiber Agreement (MFA) and invoked by the rich countries (mainly in the European Union and the United States), contributed significantly to development. The reason is straightforward: As each developing country became successful in exporting textiles or apparel to rich markets, further growth from the successful exporting country in the successful products was

restrained through bilateral agreement. While growth was permitted over time, the rate of growth was capped. This both induced the restrained developing countries to move into other, unrestrained products and created an opportunity for other developing countries to market their comparable products successfully, even though their products may not have been competitive with products from the countries under restraint. Thus we saw apparel exports migrate from successful economies such as Hong Kong and Korea to southeast Asia, then to Central America and south Asia. Mauritius owes its successful apparel industry directly to restraints imposed on exports from Hong Kong, which induced producers to find a new, attractive location. Any shopper in the United States, where the origin of imports must be labeled, will be astonished at the large number of countries that export apparel to the United States. This would have been unlikely had it not been for the MFA; without the agreement the process of diffusion would have taken many more years, perhaps decades.

When it comes to agricultural products, an analogous situation can be found.[10] The most competitive countries in the most heavily restricted products are often not the poorest countries. Sugar probably holds the record as the most protected product in the EU and the United States; rice is heavily protected in Japan. Yet the most competitive producers of sugar on any scale are probably Australia and Brazil, a rich and a middle-income country. For rice it would be the United States, Thailand, and perhaps Vietnam: a rich, a middle-income, and a poor country. Removal of U.S. and European subsidies to cotton would undoubtedly benefit some Australian farmers but also some poor farmers in Africa—if their governments allowed the higher prices to reach them, something we have not observed for products such as cocoa and coffee. So the practical issue is whether the governments would successfully devote the proceeds of higher export earnings toward development, something that might happen but cannot be taken for granted. Undistorted trade in cotton could actually hurt the apparel industries of many developing countries, since world cotton prices would presumably be higher.

2) *U.N. targets for foreign aid.* If the U.N. target of 0.7 percent of donor GDP for foreign aid had been met (the U.S. government did not subscribe), total foreign assistance in 2001 would have been on the order of $180 billion, compared with the $38 billion of official development assistance (ODA) actually recorded for 2001. Vastly greater official resources would have been available for development. In view of the growth that actually occurred after the 1960s, transforming some countries from poor to middle-income, this aid would have been available for an ever shortening list of countries (augmented, however, by the former communist countries during the 1990s).

I commented earlier on the apparent ineffectiveness of foreign aid: on average it is not associated with higher growth at all. In well-managed countries, however, foreign assistance evidently increased the rate of growth. Thus the relevant question is whether foreign aid on nearly five times the actual scale would or could have been given mainly to well-managed countries. Or could the prospect of aid on a vastly greater scale have converted some poorly managed countries into well-managed ones? Or, on the contrary, would it have induced rent-seeking within many countries on an even larger scale than actually occurred?[11]

3) *Steady private capital flows.* By the mid-1970s private capital was already moving to developing countries on a scale comparable to ODA, and in the 1990s far overshadowed ODA—first in the form of bank loans to governments, then as bond purchases and (in emerging markets) purchases of private equity. This generally rising trend was punctuated with severe reversals, first in the early 1980s, then in the late 1990s. It would have been less disruptive to these economies if the trend rise had been much smoother. But how could that have been brought about given that private investors are broadly subject to similar incentives and information during each interval? If both borrowers and lenders had been more prudent in the late 1970s and early 1990s, the debt crises of the early 1980s and the mid- to late 1990s might have been avoided, and with that the sharp reduction in output, followed (especially in Latin America) by a prolonged period of relative stagnation. Governments of capital-importing countries could have restrained their borrowing, and arguably could have limited private capital inflows through controls or other inhibitions on the inflows of capital. But that would have presupposed both a higher degree of fiscal discipline and a willingness to buck sentiment in financial circles of the rich countries, which few developing countries have. However, there is little doubt that financial crises, made possible or even inevitable by overindulgence in imports of capital, set back economic growth, briefly in some countries, but for a prolonged period in others.

4) *High oil prices and sovereign debt.* In December 1973 OPEC ministers, hosted in Tehran by the Shah of Iran, decided to increase the posted price of crude—most oil at that time sold on contract at posted prices—by a factor of nearly four, starting in January 1974. This was probably the largest financial shock the world economy has ever experienced in such a short period of time. It led directly to "stagflation"—economic contraction combined with general increases in prices—in the rich countries, and to the recession of 1974–75, the deepest in the post-1945 period. It also led to extensive buildup of debt, as oil-exporting countries initially "recycled" their higher earnings into the world banking system, and many oil importers borrowed to pay their higher import bills. Ironically, several oil-exporting countries also borrowed heavily, excessively it turned out, against their new oil wealth: Indonesia in 1974–75, and Mexico and Nigeria from 1980 to 1982. The depth of the recession and the extensive borrowing, hence the debt crisis of the early 1980s, arguably could have been avoided if oil prices had not been increased so steeply and so abruptly. World economic growth, including in developing countries (despite increased foreign aid from some OPEC members), was undoubtedly set back by this event. It is ironic that officials from many developing countries applauded the increase in oil prices at the time.

The second oil shock of 1979–80 was more complicated in origin, stemming from the revolution in Iran, and some OPEC ministers actually tried to restrain the increase in oil prices. However, the Iranian revolution may itself have had its origin in the large oil price increase of 1974, which was prompted by the Shah in order to generate the revenue he desired to modernize Iran—to make it, on his explicit conception, the Germany of continental Asia. This attempt at rapid modernization antagonized the mullahs and disoriented many Iranians, creating the conditions for

a conservative religious backlash. Human events are complicated, and unintended consequences often thwart ambitious plans.

Conclusion

Economic performance in the period 1950–2000 can only be described as fantastic in terms of the perspective of 1950, in the literal sense that if anyone had forecast what actually happened they would have been dismissed by contemporaries as living in a world of fantasy. One would not get this impression from recent discussion, which (correctly) focuses on unfinished business, but in so doing erroneously gives the impression that actual performance has been poor. Humans are chronically dissatis-fied; expectations rise with success. And some individuals and institutions have a stake in downplaying good performance. There is, to be sure, much work to be done, since too many people still live in poverty. But it is also necessary to acknowledge suc-cess when there has been success, and to avoid drawing erroneous conclusions. The post-1945 international economic system has, in general, served mankind well.

Notes

1. Interestingly, Woytinsky and Woytinsky projected a *decline* in oil and gas consumption from 0.875 to 0.4 btce, presumably reflecting exhaustion of proven reserves.
2. Kahn and Wiener predicted Africa would do relatively badly, but were more optimistic about the prospects for Latin America. See Bell (1999, p. 461).
3. The Geary-Khamis approach used by Maddison is designed to achieve transitivity and additivity, by choosing as weights internationally standardized prices for selected goods and services. For a discussion, see Maddison (1995, p. 163). Maddison judges China's offi-cial figures for growth to be too high for a variety of reasons, including under-estimation of inflation and underestimation of GDP in the earlier years. His revisions place China's 1952–1995 GDP growth at 5.6 percent per year, compared with 6.7 percent on official fig-ures. During the post-1978 transformation period the difference exceeded two percentage points. See Maddison (1998, tables C.8 and C.10).
4. In recent years, "world economic growth" calculated with national GDP at ppp weights has been about 1 percent higher than world economic growth at market exchange rates rates (e.g., 3.0 and 1.9 percent respectively in 2002). See *Global Economic Prospects 2004,* p. 3. The latter concept is appropriate for macroeconomic analysis, measuring the change in effective demand and supply, whereas the former concept is appropriate as a rough indicator of improvement in material well-being in the world.
5. For reasons unexplained, the World Bank adjusted the dollar-a-day standard to US$1.08 per day in 1993 prices (see *World Development Report 2003,* p. 246); but this was well below inflation in any country; the U.S. GDP deflator rose by 27 percent.
    It is noteworthy that Bhalla's poverty line is 26 percent higher than the $400 annual per capita income (in 1990 dollars) that Maddison assumes for most of mankind throughout most of history.

6. Even if income distribution had been perfectly equal in 1973 (it was not), the income of the top decile would have been significantly above 25 percent by 1993, if the two growth rates had both been accurate.

7. In 2001, Ethiopia was the world's poorest country, according to the *World Development Report,* with a per capita gross domestic income (GDI) of $100. At ppp, its per capita income was $710, with five countries being poorer—Sierra Leone the lowest at $480. But only 31.3 percent of Ethiopians were recorded as being below the international poverty line, with a poverty gap, reflecting the depth of poverty, amounting to 8 percent. India, with a per capita income of $460, over four times greater ($2,450 in ppp terms) recorded 44.2 percent of its population below the international poverty line, with a deeper poverty gap of 12 percent. (Comparable relationships hold if the poverty line is taken at $2 per day.) This difference could occur only if India had a substantially less equal distribution of income than did Ethiopia. But the expenditure distribution Gini coefficient recorded for Ethiopia was 40.0 percent, greater (i.e., further from equality) than that recorded for India, 37.8 percent, contrary to what is required to reconcile the other figures.

8. Indeed, it is logically possible that inequality could increase in *every* country, and yet still decline on a global basis, provided poor countries are growing more rapidly than rich ones.

9. Barro (1997, p. 98) finds no statistically significant relationship between inflation and growth for rates of inflation below 15 percent.

10. An excellent discussion of agriculture in the Doha Round can be found in *Global Economic Prospects 2004* (ch. 3).

11. Oil-exporting countries that experienced large financial windfalls from the two large oil price increases of the 1970s typically spent their increased revenues poorly; they encouraged more intense rent-seeking rather than productive entrepreneurial activity, and are not reflected in higher growth rates despite the potential. See Little and others (1993, ch. 10).

References

The word *processed* describes informally produced works that may not be commonly available through library systems.

Barro, Robert. 1997. *Determinants of Economic Growth.* Cambridge, Mass.: MIT Press.

Bates, Robert, and Anne O. Krueger. 1993. eds. *The Politics of Economic Policy Reform.* Oxford, U.K.: Basil Blackwell.

Bell, Daniel. 1999. (First published 1973). *The Coming of the Post-Industrial Society.* New York: Basic Books.

Bhalla, Surjit S. 2002. *Imagine There's No Country: Poverty, Inequality, and Growth in the Era of Globalization.* Washington, D.C.: International Institute of Economics.

———. 2003. "Crying Wolf on Poverty: or How the Millennium Development Goal for Poverty Has already Been Reached." Processed.

Braga de Macedo, Jorge, Colm Foy, and Charles P. Oman. 2002. *Development Is Back.* Paris: OECD Development Centre.

Clark, Colin. 1940 (Revised 1951). *The Conditions of Economic Progress.* London: Macmillan.

———. 1942. *The Economics of 1960.* London: Macmillan.

Cooper, Richard N. 1975. "Resource Needs Revisited." *Brookings Papers on Economic Activity* 1: 238–45.

———. 2002. "Growth and Inequality: The Role of Foreign Trade and Investment." In Boris Pleskovic and Nicholas Stern, eds., *Annual World Bank Conference on Development Economics 2001/2002*. Washington, D.C.: World Bank.

Development Assistance Committee. *Annual Report*. 1963. Paris: Organisation for Economic Co-operation and Development.

Dollar, David, and Aart Kraay. 2002a. "Growth Is Good for the Poor." *Journal of Economic Growth* 7 (September): 195–225.

———. 2002b. "Spreading the Wealth." *Foreign Affairs* 81 (January/February): 120–33.

Dollar, David, and Lant Pritchett. 1998. *Assessing Aid: What Works, What Doesn't, and Why*. World Bank Research Report. New York: Oxford University Press.

Fischer, Stanley. 2004. *IMF Essays from a Time of Crisis*. Cambridge, Mass.: MIT Press.

Haggard, Stephan, Richard N. Cooper, and Chung-in Moon. 1993. "Policy Reform in Korea." In Robert Bates and Anne O. Krueger, eds., *The Politics of Economic Policy Reform*. Oxford, U.K.: Basil Blackwell.

Helliwell, John F. 1994. "Empirical Linkages between Democracy and Economic Growth." *British Journal of Political Science* 24 (August): 225–48.

Huntington, Samuel. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman, Okla.: University of Oklahoma Press.

International Monetary Fund. *International Financial Statistics*. Various years. Washington, D.C.

Kahn, Herman, and Anthony Wiener. 1967. *The Year 2000*. New York: Macmillan.

Kapur, Devesh, John P. Lewis, and Richard Webb. 1997. *The World Bank: Its First Half Century*. Washington, D.C.: Brookings Institution Press.

Kuznets, Simon. 1966. *Modern Economic Growth*. New Haven: Yale University Press.

Lewis, W. Arthur. 1955. *The Theory of Economic Growth*. Homewood, Ill.: Richard D. Irwin.

Lewis, William W. 2004. *The Power of Productivity: Wealth, Poverty, and the Threat to Global Stability*. Chicago: University of Chicago Press.

Little, I.M.D., Richard N. Cooper, W. Max Corden, and Sarath Rajapatirana. 1993. *Boom, Crisis, and Adjustment: Macroeconomic Management in Developing Countries*. New York: Oxford University Press.

Maddison, Angus. 1995. *Monitoring the World Economy, 1820–1992*. Paris: OECD Development Centre.

———.1998. *Chinese Economic Performance in the Long Run*. Paris: OECD Development Centre.

———. 2001. *The World Economy: A Millennial Perspective*. Paris: OECD Development Centre.

———. 2002. "The West and the Rest in the International Economic Order." In Jorge Braga de Macedo, Colm Foy, and Charles P. Oman, eds., *Development Is Back*. Paris: OECD Development Centre.

President's Materials Policy Commission (Paley Commission). 1952. *Resources for Freedom*. Washington: U.S. Government Printing Office.

Ren, Ruoen. 1997. *China's Economic Performance in International Perspective*. Paris: OECD Development Centre.

Reynolds, Lloyd G. 1985. *Economic Growth in the Third World, 1850–1980.* New Haven: Yale University Press.

Rodrik, Dani, and Francisco Rodriguez. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." In Ben Bernanke and Kenneth S. Rogoff, eds., *Macroeconomics Annual 2000.* Cambridge, Mass.: MIT Press for the National Bureau of Economic Research.

Sachs, Jeffrey D., and Andrew M. Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1: 1–118.

Sen, Amartya. 1999. *Development As Freedom.* New York: Alfred A. Knopf.

Smil, Vaclav. 2003. *Energy at the Crossroads: Global Perspectives and Uncertainties.* Cambridge, Mass.: MIT Press.

Sundaram, K., and Suresh D. Tendulkar. 2003a. "NAS-NSS Estimates of Private Consumption for Poverty Estimation." *Economic and Political Weekly* 38 (January 25): 376–84.

———.2003b. "Poverty in India in the 1990s." *Economic and Political Weekly* 38 (November 15): 4865–872.

Thomas, Vinod, Mansoor Dailami, Ashok Dhareshwar, Daniel Kaufmann, Nalin Kishor, Ramon E. Lopez, and Yan Wang. 2000. *The Quality of Growth.* New York: Oxford University Press.

United Nations. 1963. *The Growth of World Industry, 1938–1961: International Analysis and Tables.* New York.

Warner, Andrew M. 2003. "Once More into the Breach: Economic Growth and Global Integration." Processed.

World Bank. 2001. *World Development Indicators.* Washington, D.C.

———. *Global Development Finance.* Various years. Washington, D.C.

———. *Global Economic Prospects.* Various years. Washington, D.C.

———. *World Development Indicators.* Various years. Washington, D.C.

———. *World Development Report.* Various years. Washington, D.C.

Woytinsky, W. S. and E. S. Woytinsky. 1953. *World Population and Production.* New York: Twentieth Century Fund.

The Evolution of Development Thinking: Theory and Policy

GUSTAV RANIS

This paper makes an effort to trace the course of development thinking and associated development policy over the past six decades.

The first section focuses on the early postwar consensus, with theory focused largely on a revival and extensions of classical dualism theory, and policy concentrating on creating the preconditions for development and the severing of colonial ties viewed as market related.

The following section traces the increasing awareness of the role of prices, the rejection of various types of elasticity pessimism, and a diminishing reliance on the developmentalist state as the main actor. This is also the period when the international financial institutions (IFIs) moved toward increased reliance on structural adjustment lending associated with conditionality and reform at both macro and micro levels of policy, as embodied in the Washington Consensus and its extensions.

The third section illuminates the search for "silver bullets" over time, in both the theory and policy arenas. It demonstrates the never-ending search for dimensions of development in both the theory and policy realms that can be identified as critical (or key) to achieving success.

Finally, the concluding section presents the author's rather personal assessment of where we are at the moment and where we will be, or should be, heading in the effort to achieve the Third World's basic development objective of human development fuelled by equitable growth.

I intend to first review the early postwar consensus on development thinking in both its theory and policy dimensions, proceed to take up the period of the Washington Consensus, trace the search for "silver bullets" that has taken place more or less consistently over the past two decades, and conclude with an assessment of where we now are and are likely to be heading in the future.

The Early Postwar Consensus

In the 1950s and 1960s, the previously neglected subfield of development economics was rediscovered. Available economic models seemed to offer only limited insights into the practical problems facing the so-called Third World. The dominant one-sector macro models of the day, from Keynesian to Harrod-Domar (see Harrod 1939, and Domar 1957) to Solow (1956), seemed to have relatively little relevance for societies not primarily concerned with business cycles or steady state properties. Most contemporary growth models, in other words, were seen as advanced country-related, relatively abstract theoretical constructs that were faithful to the dominant assumptions of neoclassical macro-theory: full employment, market clearing, and perfect competition, all of which seemed to have little relevance for the segmented commodity, labor, and credit markets of the poor countries.

Against this background, the concept of dualism attracted considerable attention. Sociological dualism associated with the name of Boeke (1953) emphasized differences between Western and non-Western objectives and cultures; and technological dualism pointed to by Higgins (1956) and Eckaus (1955) focused on the difference between variable factor proportions in traditional and fixed coefficients in modern sectors. A third and increasingly dominant strand focused on the coexistence of sectors that were basically asymmetrical in behavior and thus dualistic in some key analytical dimensions. The first clear manifestation of this third version of dualism undoubtedly appeared in the *tableau economique* of the physiocrats who emphasized the importance of an agricultural surplus to support nonproductive activities elsewhere; but it was classical dualism, coinciding more or less with the advent of what was erroneously termed the industrial revolution in Western Europe, that provided the raw materials for the renewed emphasis on dualism in early post–World War II development theory.

It was classical school concepts, owing much to Ricardo (1951), which focused attention on the coexistence of a still overwhelmingly large agricultural sector subject to diminishing returns to labor on basically fixed land, and nonagricultural activities growing as a consequence mainly of the accumulation of fixed capital and labor drawn out of agriculture that were central to the story. While the classical school did not really model the interaction between these two sectors, it is clear that the main fuel for the reallocation of labor and for the accumulation of industrial capital was seen as coming from the profits of agricultural capitalists. It should, of course, be noted that the assumption of the near-fixity of land was combined with Malthusian population pressures (see Malthus 1815) and with the notion of an institutionally determined agricultural real wage, even though, in contrast to the physiocratic view, the laboring class was now free and in a position to bargain with capitalist landlords in setting the level of that wage. As is well known, Ricardian/Malthusian pessimism with respect to the ultimate stagnation of agriculture in the absence of marked technology change was a dominant feature of their analytical work. Whether innovations in industry, reflecting Adam Smith's relative optimism, would be strong enough to provide sufficient industrial profits to rescue the situation remained a controversial issue.

The first modern theorists to build on classical dualism were undoubtedly Rosenstein-Rodan (1943), Mandelbaum (1945), and Nurkse (1953), each of whom, in their own way, pointed to the existence of surplus labor as a potential resource that, once reallocated from agriculture to higher productivity pursuits in nonagriculture, would constitute a major fuel for development. But it was Arthur Lewis (1954) who, in this seminal article, built on some of the main ingredients of the classical tradition, focusing more precisely on dualism in labor markets (i.e., a competitive wage in non-agriculture tied to a bargaining or institutional wage in agriculture). Lewis, moreover, found himself allied with Smith (1880) in seeing the relatively small nonagricultural commercialized sector as the dynamic partner, expanding and fed by the mobilization of the hidden rural savings that Nurkse and Rosenstein-Rodan had identified. In Lewis's view, the reallocation process would continue until all the surplus agricultural labor (i.e., not necessarily zero marginal product labor but, as emphasized by Fei and Ranis [1961, 1964], all those whose remuneration exceeded their low marginal product) had moved out of agriculture into commercialized nonagriculture, marking a turning point at which time dualism would atrophy and the economy become fully neoclassical.

It is fair to say that the theoretical elements of this early postwar consensus focused on capital scarcity and savings-pushed growth, with relatively minor emphasis on technology change in either sector. Moreover, both Rosenstein-Rodan and Nurkse very much emphasized the need for balanced growth, not only between agriculture and nonagriculture, but also on the need for balance within each sector, so that Say's Law could come into play and not only shoes but also socks would be produced, feeding each other on both the supply and demand sides. It is also noteworthy that there was a good deal of elasticity pessimism in the air during those years, both with respect to agricultural response mechanisms, as already noted, and with respect to the open economy (i.e., export opportunities). The international trade scene, dominated by Prebisch (1962), Singer (1950), and Myrdal (1957), was painted in colors unfriendly to development. There were, of course, some early critics of various aspects of dualism, on the one hand, and of structuralism, on the other, represented by adherents to the neoclassical paradigm. To one degree or another they rejected the notion of labor surplus (Schultz 1964) and the nonresponsiveness to price signals of various actors (Haberler 1988; Bauer 1957). But they were clearly voices in the wilderness.

The prevailing theoretical winds indicated that, on the policy side, there was a strong inclination to turn to the interventionist state as a key instrument of development. The motivation for this trend was at least twofold. One was the desire to cut pre-independence colonial ties which were identified with the market mechanism; and second, there was a felt need to create an economy out of what was often still viewed as an agglomeration of agents and resources requiring, first of all, the creation of the so-called "preconditions of development." At home, the interventionist state accordingly felt the need to create infrastructure, or the institutions required to permit the functioning of a national entity; plus the subsidization in various ways of newly created nonagricultural entrepreneurs, complemented on the international side

by the infamous and indefinite import substitution syndrome protecting these entre-preneurs. Typically, governments thus tended to overcommit by deploying a vast array of direct and indirect policy instruments to shift resources toward themselves and favored private groups, all in the effort to promote growth. These were usually under-the-table transfers that tended to manufacture profits for the state or the favored new entrepreneurial class. The motivation was to promote industry, with relatively less attention paid to what was viewed as a stubbornly stagnant agriculture portrayed as a drag on the economy, and with peasants seen as nonresponsive to prices and profit opportunities. Industrialization was generally viewed as equivalent to development, with policymakers in search of a second industrial revolution.

A logical accompaniment of this view of the world was "planning models" that focused on the flow of resources, domestically financed investment supplemented by foreign capital, and paid relatively little attention to changes in the behavior of the system or the relevance of technology. Such planning models, often based on simple Harrod-Domar foundations, started with exogenous population growth, per capita income targets and focused heavily on how—given certain input-output relations—necessary savings, domestic and foreign, would be sufficient to reach politically required targets. There were, of course, also fancier models, including those of Mahalanobis (1955), modified later by Chenery and associates (1971), all of them relatively silent on price flexibility, exchange rate flexibility, and other dimensions of the market mechanism.

It should be noted that, while there was always some recognition of the impor-tance of distributive issues, the predominant view of policymakers at that time was that growth and efficiency should take priority and that issues of equity, such as income distribution and poverty alleviation, would be taken care of at a later date. Clearly, high profit and savings rates were viewed as paramount instruments and any premature redistribution was viewed as a trade-off with the objective of growth.

The planning school may be characterized by relative formalism in methodology, usually envisioning a multi-sector production function with multiple inputs and inter-national variables, often exogenously postulated. In this way economic plans could be seen to portray the operation and growth of the economy in a wholistic perspec-tive, with all sectors tending to be viewed as homogeneous and symmetrical. A related trait of the planning school was the systematic application of mathematical models in order to determine the magnitude of all the relevant variables consistently through time. Such "planning for resources" was really based on a belief in the appropriateness of the existing policy rails on which the economy found itself.

However, by the 1970s it had become increasingly clear that the development problem was one of transition from one regime to another, during which changes in structure lie at the very heart of the process, coupled with the realization that five-year plans can quickly become political albatrosses around the necks of govern-ments—as exogenous shocks inevitably occur. The real focus of planning conse-quently shifted gradually from a resource focus to devising strategies for policy change to accommodate the changing requirements of transition.

It is undoubtedly correct to say that Solow (1957) and Kuznets (1955) provided the most important transitional mechanisms in the realm of both theory and policy as we move from this postwar consensus into what later became known as the era of the Washington Consensus. Solow's 1957 signal contribution was to emphasize, really for the first time since Schumpeter (1959), the importance of technology in generating growth, spawning a huge literature focused on measuring and quantifying the effects of technology change. This provided a new point of departure for neoclassical growth theory, not only replacing Harrod-Domar with a substitutable production function, but also enthroning exogenous technology change, plus the ensuing effort to whittle down the Solow residual as much as possible. It introduced critical flexibility into the system and spawned a good deal of applied work on the role of research and development (R&D), patents, and other forms of scientific endeavor, leading at a later stage to the so-called "new growth theory" (discussed later) which moved to try to endogenize technology change.

It was, however, Kuznets (1971), though mainly concerned with describing modern growth rather than analyzing the transition process in getting there, who provided another essential ingredient focused precisely on the developing world at the end of the postwar consensus era. Kuznets was interested in why some developing countries were successful and others not, and placed major emphasis on the sources of structural change over time as between agriculture, industry, and services. Chenery and his associates (1974) took up the cudgel, using regression analysis to depict dimensions of average structural change in less developed countries, first by the use of cross-sections, and later through increasing resort to time-series analysis and pooled regressions. The basic question being addressed was how productivity gains and increments in output are allocated among sectors as income per capita rises, and how one explains deviations from average patterns. Kuznets always insisted that such structural changes resulted from the interaction of underlying changes in final demand and capacity conditions, with deviations from any normal pattern largely attributable to differences in the underlying state of nature. He viewed policy as either basically accommodative or obstructive to the play of underlying economic forces, and did not view it as an exogenous variable. This is in contrast to Chenery's inclusion of differences in policy among his typological categories.

Over time there was a growing recognition of the potential relevance of flexibility in factor proportions and of the importance of labor-using or capital-saving technology change. Observers began to realize that distortions in relative factor prices, overvalued exchange rates, low interest rates, and biased internal terms of trade, all instruments of import substitution, not only discouraged agriculture, encouraged industrial capital and import intensity, and limited the generation of employment, but also created windfall profits for favored elites long after such support was no longer necessary for infant industry reasons. The realization that the enhanced use of the market needed to be complemented by institutional reforms (at least to the extent that small-scale rural development actors could obtain an adequate share of credit, foreign exchange, and infrastructural attention) was but one indication of that

gradual change in the development paradigm, applied most pronouncedly at first in East Asia.

The Washington Consensus as Initially Conceived and Subsequently Amended

It is undoubtedly unfair to attribute the realization that policy change is the key ingredient of successful development to the international financial institutions (IFIs). I rather would give credit for the realization that prices matter and that macroeconomic stability matters to Little, Scitovsky, and Scott (1970), as well as to Bhagwati (1978), Krueger (1978), and Cohen and Ranis (1971), among others, who insisted that a restructuring of the rails of development was required.

Once easy import substitution of the nondurable consumer goods type had run out of steam, most developing countries increasingly faced a critical choice: continued import substitution while moving toward more capital and technology intensive output mixes, or export orientation testing competitive international markets. Trade liberalization was generally accepted as an instrument, but its timing was subject to large differences across the developing world. Export promotion often came first, accompanied by a shift from quantitative restrictions to tariffs, the subsequent unification of tariffs and their gradual reduction, even if the timing was differently implemented. But performance lagged almost everywhere except in East Asia, which had moved further than other regions in rejecting the continued import substitution alternative.

There can be little doubt about the important facilitating role of exports, extending beyond the handmaiden role emphasized by Kravis (1970), even if one does not accept the notion that exports constitute the principal engine of growth and that export promotion, especially of nontraditional goods, represents the solution in virtually all circumstances. It should be noted that even in small open economies that have been successful, such as Taiwan, initial development success was determined largely at home, via balanced domestic growth and the subsequent export of, first, traditional (agricultural) goods, before testing the international waters for nontraditional exports. Trade and the associated international movements of technology and capital have increasingly been seen as of potentially great help but still as representing only an assist to the basic domestic development effort. It should again be emphasized that the East Asians encouraged exports long before they opened their domestic economies to competitive imports in a sustained fashion. One causal chain ran from exports to growth by way of enhanced competitiveness as well as the direct impact of imported technology through patents, human capital, and capital goods incorporated in foreign direct investment (FDI). But another important causal chain also runs from domestic growth generated via R&D back toward the enhanced capacity to take advantage of export opportunities.

One basic ingredient of the new emerging consensus was the need for macroeconomic stability, increasingly accepted as a basic necessity by both orthodox and

heterodox observers, whether inflation at 20 percent or 5 percent is viewed as the tolerable limit. Avoidance of large-scale deficits as a percentage of gross domestic product (GDP), along with too rapid monetary expansion, were seen as critical components, with tax reform and the shifting of public expenditure patterns usually part and parcel of the package. With the gradual rejection of structuralism (i.e., the belief in the nonresponsiveness of agriculture), and of export pessimism, attention focused instead on an enhanced reliance on liberalizing markets. The original list of Washington Consensus objectives included other items such as privatization and unified and competitive exchange rates, both still under dispute today, and the simultaneous liberalization of financial markets, both domestic and international, the latter certainly with caveats now attached. What has stood the test of time is the relative openness to FDI and the acceptance of the notion that the gradual deregulation of various control systems is essential for the full mobilization of the private sector.

While not usually included on the Consensus agenda, the realization that technology choice and the choice of direction for technology change could be of major importance for successful development played an increasingly important role (see Stewart 1977; Evenson and Ranis 1990). The importance of public sector research, especially on export-oriented cash crops such as sugar, cotton, and coffee, had long been recognized; but its role in basic food crops, in nontraditional agriculture, and in nonagricultural exports came about only gradually. The Green Revolution, after all, represented an imaginative combination of international and adaptive domestic research (see Griliches 1957; Evenson and Kislev 1975). It became increasingly clear that food-producing agriculture cannot be neglected, that peasants do respond to their economic environment, and that industry cannot pull an economy into modern economic growth if agriculture remains stagnant. It is also interesting to note that R&D in medium- and small-scale firms that typically cannot afford to conduct their own R&D, such as in China's TVEs and Taiwan's small- and medium-scale enterprises, had a large payoff. The productivity of carefully selected public sector research has come to the fore, even as horror stories can be told in reference to the white-elephant aspect of many LDC science and technology institutes setting their own agendas not related to the actual needs of the economy. But such stories do not obviate the point that, when increasingly hard budgets become credible, R&D as a public good can have an important role in permitting the continued realization of domestic balanced growth, combined with an export drive powered by dynamic comparative advantage.

Most R&D, of course, takes place in the private sector. One need only point to the substantial discrepancy among developing countries in terms of levels of total factor productivity or, as some observers prefer, the differential efficiency of investment allocation, to be convinced that an increased emphasis on indigenous applied science and technology is bound to pay off. Tax codes can be modified to encourage greater risk-taking, and increased flexibility in the legal implementation dimensions of intellectual property rights can be paid attention to as a country begins to move up the development ladder. Some countries resort to a different kind of patent, the utility model, with a shorter period of protection and a lower threshold for discovery, one way of

encouraging the potentially important, if not spectacular, adaptive (or blue-collar) type of technological change. This clearly also relates directly to the new growth theory literature (discussed below).

Privatization was part of the macro package generally accepted in the 1980s, partly because of the enhanced efficiency it promised and partly because of the fiscal boost it provided, at least in the short run. On the other hand, critics of privatization have been able to point to the accompanying corruption in some of the transition countries of Eastern Europe as well as the all too frequent exchange of private for public monopoly power (see Fischer, Sahay, and Vegh 1996; Stiglitz 1991).

It is fair to say that, while there was consensus about the basic macroeconomic ingredients of the development policy package needed to ensure economic restructuring, there was also, from the beginning, a considerable difference of views concerning what additional changes were needed at the micro level, which is clearly much more differentiated by country. These included enhanced labor market flexibility, legal, financial, and other institutional reforms. Nevertheless, it is a fact that bilateral agencies, especially the U.S. Agency for International Development (USAID), which termed its 1960's instrument "program lending," and subsequently the World Bank and International Monetary Fund (IMF), which termed similar instruments "structural adjustment lending," combined policy packages incorporating both macro and a variety of micro ingredients with fast-disbursing loans. This device has become the subject of lively debate, ranging from the cost-effectiveness of the resources spent in support of country policy reforms all the way to the implications of extensive conditionality lists infringing on recipients' sensitive internal affairs.

Undoubtedly, today the bloom is off the rose of structural adjustment or program lending. Given the mixed record of aid conditionality cum reform packages compiled by the World Bank's own internal evaluation unit (see also Easterly 2001), the argument currently being made is that now may be the time to abandon the instrument altogether and either return to project lending (including those big bad dams), or move to the poverty reduction strategy papers (PRSPs) currently being fashioned for the poorest LDCs. In theory, policy-based lending can help countries achieve any objective, even if one has to admit that in the case of a multi-cook operation it is extremely difficult to precisely judge the contribution of such packages. The counterfactual is typically unknowable. But before disenchantment takes over completely, we should recall that there are historical USAID cases, such as Pakistan and Taiwan in the 1960s, and a number of World Bank cases, including Chile, Ghana and Poland in the 1970s and 1980s, where such packages worked well.

I would argue, therefore, that, before the policy-based loan instrument is abandoned, it is preferable to see if enhanced decentralization by the World Bank, coupled with an effort to achieve real ownership by recipients, can still rescue it. In my view, the structural adjustment loans of the past and the closely related PRSPs of today continue to be negatively affected by the rush to judgment on both sides, in the attempt to put together packages that can be signed off on so that the money can flow. IFI staff and loan recipients are similarly motivated, the former seeing their rewards and promotions in terms of the volume of commitments made, the latter in

terms of the relief expected from fast-disbursing loans. All the rhetoric about the importance of quality and ownership still lacks bite, with both parties not as concerned as they should be that the reform package is more than superficially a part of the body politic of the recipient.

The IFIs, in other words, all too often don't act like banks; and the countries all too often have a strong incentive simply to go through the motions of borrowing in order to obtain quick economic relief. With the desire to lend still overwhelmingly strong and the attached list of conditions too long and insufficiently differentiated, it is no overstatement to comment that both parties have reached a level of reform fatigue that clearly needs to be addressed. In the wake of the debt crisis of the 1980s this problem became particularly acute. Just as it is impossible for U.S. bilateral aid to the Republic of Egypt, for example, to secure both the support of the so-called peace process with Israel and improved economic performance, it is difficult to use one instrument to achieve both balance of payments (BOPs) crisis support and improved long-term economic performance. There is no doubt that the disenchantment with the structural adjustment experience of the 1980s and 1990s and the nascent disenchantment with the PRSPs on virtually the same grounds has led to a reassessment not only of development thinking, but also of development policy. With old certainties under pressure, the oscillating search for some "silver bullet" continues.

The Oscillating Search for a "Silver Bullet"

With policy-based loans and conditionality under attack, development thought has entered an era of disarray, with a substantial number of competitive concepts in play. Some of these concepts focus on the search for a more appropriate objective of development, others on a reassessment of how to achieve this. Turning once again first to theory, viewing per capita income growth as "the" key objective has been questioned for some time (see for example, Srinivasan 1994, Streeten 1994, Sen 1992, and Sugden 1993). In fact, as early as the 1950s and 1960s, both India and Sri Lanka focused on poverty and employment in their five-year plans. In the 1970s a "basic needs" approach, zeroing in on the direct provision of essential commodities and thus short-circuiting income, made an appearance but was short-lived—partly because it never fashioned firm theoretical links to what else we know about development, and partly because it was never really accepted abroad, where it was seen as a device for explaining away lower levels of aid. But serious mainstream attention to the distribution of income, to the extent to which private income poverty is being reduced, and, more recently, to the extent to which public income poverty (i.e., the distribution of public goods) is being addressed came later, in the late 1970s. During the 1990s, the achievement of improvements in various dimensions of human development (infant mortality, life expectancy, literacy) came to the fore as the appropriate fundamental objective of development. All this, of course, does not mean that income has been dethroned, only that it is now increasingly seen as an essential means to societal ends rather than as an end in itself.

But the concern with distribution has had a long and useful life, ever since Kuznets in the 1950s worried about the possibility that income growth might have to be bought at the cost of an initially worsening distribution (i.e., the basic efficiency-equity trade-off, see Okun 1975). Aside from the large theoretical literature on inequality and growth in developing countries (such as Banerjee and Newman 1993, and Aghion and Bolton 1997), there has been a continuing lively debate ever since on whether or not travel along the so-called inverse U-shaped Kuznets curve was inevitable or avoidable. During his tenure as World Bank president Robert McNamara initially moved the Bank in the direction of discussing distributional issues. And while Dudley Seers talked about "dethroning the GNP," what followed was "Redistribution with Growth" (Chenery and others 1974), a collaboration between the University of Sussex and the World Bank, and a string of research projects including "Growth with Equity" (Fei, Ranis, and Kuo 1979) financed by the World Bank. Current assessments are that, while most countries seem to experience some deterioration in income distribution during rapid growth, this is by no means a necessity and there are quite a few counter-examples, even outside of the well-known East Asian cases. (For example, see Fields 2001, Bourguignon and Pereira da Silva 2003, Deninger and Squire 1997, and Ravallion and Chen 1999). Certainly, we have gotten away from using pooled cross-sections of historical data and are focusing more on country cases that, over time, yield a variety of patterns. Fei, Ranis, and Kuo (1979) illustrate the case of Taiwan with rapid growth associated with improving distribution (see also Persson and Tabellini 1994).

More controversial is the relationship between growth and income poverty allevi-ation. It seems quite clear from the evidence that per capita income growth is a nec-essary but not sufficient condition for poverty reduction (see Ravallion and Datt 1999; Lipton and Ravallion 1995), the necessary rate of growth depending on its character. For example, with respect to the production of primary commodities, what matters is whether they are generated by small farmers on fairly equally distributed plots of land or on large, land-intensive plantations (see Deninger 1999). In nonagri-culture, much depends on technology and output mix choices yielding more or less labor-intensive outcomes (see Evenson and Ranis 1990).

Among theoretical revisionisms has been the recent effort to revive import substi-tution models, supported initially by the "new trade theory" ideas of Paul Krugman (1994), and, more recently, the challenge to openness spearheaded by Stiglitz (2002) and Rodrik (1996, 1999), encouraging a revival of populism in the south. Krugman emphasized the role of economies of scale and externalities in trade which was—in spite of his insistence that the concept was to be applied mainly to trade among rich countries—eagerly taken up by some adherents of a return to the "picking winners for the long run" view. Yet more influential and popular have been the recent attacks on globalization by Stiglitz and Rodrik, in which they question the firmly held posi-tion among Washington Consensus adherents that increased openness correlates pos-itively with higher rates of growth. I acknowledge that infant industry protection has been deployed by every developing country in the post–World War II era, as well as by currently developed countries during their earlier economic history. Contrary to

members of the Chicago School, I believe that such interventionism is necessary in the early stage of a country's development; but it is also clear to me that the regime must be strictly time constrained, providing assurance of a more or less reliable trend in the direction of a gradual reduction of the large interventionist policy paraphernalia.

Stiglitz and Rodrik, along with Wade (1990), Lall (1992), and Amsden (1989), assign the favorable results achieved by Korea and Taiwan, among others, to that large array of government interventions generating hot-house conditions for a new and relatively inexperienced entrepreneurial class; but they fail to pay adequate attention to the seemingly inevitable hardening of protectionist arteries if the signals for a gradual but persistent lowering of these hot-house temperatures are not made transparent and credible. Developed countries are sometimes accused of "kicking the ladder" that brought them developmental success in the past. My reading is that, while this may be true, the more successful countries used a ladder that did not consist of continued and increasingly expensive secondary import substitution policies but was consistent with the expectations of a liberalization trend that enhanced both domestic and international competitiveness over time.

More recently, the emphasis on human development, building on the work of Amartya Sen (1985), Mahbub ul Haq (1992), and the Human Development Reports of the United Nations Development Programme (UNDP), have attracted a good deal of theoretical attention, including, in particular, the two-way relationship between growth as the necessary engine and human development as the bottom-line objective. The relationship between growth and improvements in infant mortality, life expectancy, or literacy—preferable to any necessarily arbitrary index—represents a still somewhat underdeveloped set of production functions (see Behrman 1996; Birdsall 1985). The feedback from increments in human development back to growth comes closer to being captured by the conventional macroeconomic production function as amended over time (i.e., including both conventional Solow-based and unconventional "new growth theory"–related approaches).

This two-way relationship has been studied carefully by Ranis, Stewart, and Ramirez (2000) and in more recent work (Boozer and others 2004). We find convincing evidence across all developing countries over time to the effect that, in order to reach a virtuous cycle of sustained growth, accompanied by continuous improvements in human development, priority attention must be given to the latter. It is difficult, if not impossible, to reach the "promised land" of mutual reinforcement between growth and human development from an asymmetric position favoring growth as a temporal priority.

In the 1980s a new branch of growth theory came into vogue which, based on some well-accepted earlier notions in the literature (Arrow 1962), tried to endogenize technology change through credible models of market externalities to explain some stylized facts in both developing and mature economies. This literature, pioneered by Romer (1990), Lucas (1988), Grossman and Helpman (1991), and, more recently, Aghion and Howitt (1998), shares the Solowian view of technology change as the driving force of output growth, but, while emphasizing constant or even diminishing returns of scale at the individual firm level, sees increasing returns of scale (i.e.,

externalities) at the economy level. Grossman and Helpman analyzed the open economy implications of such endogenous growth theory models and focused largely on R&D that actually serves two functions (i.e., accelerating the introduction of new capital goods and providing spillovers by reducing the cost of manufactured goods). While LDCs undertake relatively little R&D, at least of the formal or white collar variety, the transition to economic maturity in the developing world requires an ever increasing competence to adopt and adapt new technologies (see Pack and Westphal 1986).

On the policy front, guided by the somewhat uncertain search for theoretical advances, we continue to worry about the relative importance of market failure and government failure, while moving from "market friendly" government interventions to focusing increased attention on the institutions needed to repair both inadequate government infrastructure and the functioning of markets. Perhaps the most important change in development thinking in recent years has been a renewed emphasis on the importance of such institutions—ranging all the way from property rights, to civil service reform, to the financial system, with priorities dependent on the pre-existing state of play (i.e., the initial conditions emphasized by Kuznets and others many years ago).

Much current thinking and modeling focuses on the reduction of transactions costs as a result of relevant investments, following the path outlined by North (1990, 1991), Williamson (1975), and others. This renewed emphasis on institutional economics also has relevance for the argument between "big bang" and gradualism approaches attending any developing country reform package. If, for example, institutional changes affecting domestic financial markets have to be put in place before a system can proceed to open itself up to international capital movements, especially of the short-term portfolio variety, both the timing and the sequencing of reforms are clearly materially affected. Early efforts in the transition countries of Eastern Europe to do virtually everything at once, while neglecting the institutional dimensions, have, in fact, led to the conclusion that this is a riskier choice than the gradualism exhibited in East Asia, including Mainland China. A prominent contrast is the way privatization was organized in many other parts of the developing world as well as in the transition countries of Eastern Europe, i.e., in the absence of adequate provision for regulatory institutions to ensure a workably competitive, post-privatization private sector, as well as the reduction of corruption in the very process of transferring public goods into favored private hands.

More recently, in fact mostly in the past decade, there has been a strong emphasis among development economists—academicians as well as those on the policy scene—on the micro foundations of development issues. Development economists and policymakers have become more concerned with micro-level decisions, realizing their role in the growth of an economy. For example, the role of women in household decisionmaking, and the effects of the proportion of household resources controlled by women on the health and nutrition of their children, has been empirically documented in a number of micro studies (Behrman and Wolfe 1987, and Hoddinot and Haddad 1991).

The role of microeconomics in understanding poorly functioning markets has also come to the forefront of development economics research. The importance of poorly

functioning land, labor, and credit markets is being studied extensively. And the role of informal networks and institutions in dealing with such market failure is now the focus of much research, relating directly to the more macroeconomic literature on both the role and the formation of relevant institutions.

A seminal paper in this area was by Townsend (1994), who looked at whether households in India are able to pool risk across space in the presence of poorly functioning capital markets. He directly tested the general equilibrium implications of such a consumption-smoothing model, using household-level data for India, and found that households do indeed pool risk across space. What is fascinating about this paper (and subsequent work) on consumption smoothing is that it is not all that different from the income pooling ideas behind models of dualism that were at the forefront of development economics research a few decades ago. This literature has also subsequently fueled a large microeconomic literature in development economics on credit institutions and their efficiency (see Rosenzweig and Wolpin 1993, Udry 1994, and Deaton and Paxson 1994).

There are also various microeconomic studies on the impacts of differential labor and land markets on bottom line outcomes. The interlinkage of contracts and the two-tiered nature of labor markets in developing economies and their efficiency has been studied extensively (see Eswaran and Kotwal 1985, Mukherjee and Ray 1995, Foster and Rosenzweig 1996, and Rosenzweig 1988). Again, this literature is closely tied to the earlier models of surplus labor and dualism. The aim of this literature has been to understand the implications of market failures, the institutions that may arise at a micro level to cope with such failures, and how best to structure policy to make allowance for these institutions (see, for example, Greif 1993).

Finally, we have recently seen a large increase in the active role played by microcredit institutions and nongovernmental organizations (NGOs) in developing countries, in almost every possible policy sphere, ranging from credit (such as the Grameen Bank and Bangladesh Research Action Committee in Bangladesh), to health (Internationaal Christelijk Steunfonds in Kenya), education, and even intellectual property rights and codes of conduct. Both the macro-and microeconomic impacts of such NGOs are being evaluated. A very active future area of development economics promises to assess the effects of NGO policies and social policy programs. For example, Pitt and Khandker (1998) and Morduch (1999) analyze the Grameen Bank program in Bangladesh; Kremer and Miguel (2001) assess the impacts of deworming health programs in Kenya; and Skoufias (2001) and Schultz (2001) examine the impact of the Progresa schooling initiative in Mexico.

Best Guesses as to the Way Forward

In this concluding section, I intend to unabashedly ride several hobby horses, hopefully moving in the same general direction with respect to where development thinking and policy are (or at least should be) heading.

First, on the methodology or theory front, I think we will be moving away from large *n* Barro-type (see Barro 1991, 1997) cross-sections, which have included more

and more variables, including geography and religion, accompanied by diminishing robustness, and toward a set of small *n* comparative historical studies encompassing typologically "neighboring" countries. Second, we need to pursue much more carefully the aforementioned two-way relationship between growth and improvements in human development, especially with respect to the preferred sequencing, if sustained long-term improvement in both dimensions is to be attained.

Third, I believe we will need to take a much closer look at the pros and cons of decentralization and its relation to democratization and decisionmaking by the broader body politic. Such analysis of decentralization should clearly not only be of the customary vertical type, i.e., focusing on local government and its fiscal and other functions, moving from deconcentration and delegation to the still rare case of devolution, but also horizontal decentralization, moving from the domination by the executive branch of government, especially the ministry of finance, to the legislative and the judiciary, representative of the critical rule of law dimension of democratic governance.

There can be little doubt that greater local control over fiscal resources is bound to lead to larger expenditure on the social sectors as well as on small-scale infrastructure. Given the benefit principle of taxation, it is also likely to lead to larger total resources being available for all purposes, while the promotion of national standards and the support of equity objectives across regions will continue to require central government action. Central resource transfers for health, education and infrastructure, in lieu of actual fiscal devolution, have, nonetheless, significantly enhanced both growth and human development indicators whenever we have observed at least delegation in both unitary and federal government systems (see for example, Ranis and Stewart [1994] for the case of Indonesia, and Habibi and others [2003] for the case of Argentina).

Fourth, I view plentiful natural resources and relatively easy access to foreign capital as extensions of the Dutch Disease problem, more serious than its narrower exchange rate implications, since they focus on decisionmaking with respect to reforms that may be needed but can be avoided. Not being "up against it" represents a prescription for not making the necessary political effort to overcome vested interests. I have observed a marked contrast between the more or less linear trend toward the depoliticization of policymaking in countries with relatively poor natural resource endowments and a consistently more oscillatory pattern of policy evolution in countries with good natural resources, including much of Latin America, Nigeria, and Indonesia. Policymaking in such resource-rich countries illustrates a tendency toward excessive activism during good times, followed by attempts to artificially maintain growth by government action when times are relatively bad. Moreover, there exists substantial evidence that an important indicator of the differential quality of policy response to the inevitable exogenous shock resides in the manner in which growth is financed. In other words, the critical distinguishing characteristic is not just the size of the tax effort relative to the GDP, but also the relative reliance on covert versus overt means of transferring resources.

The typical natural-resource-rich-country case demonstrates that liberalization efforts initiated at the beginning of an upturn cannot be sustained once the improvement in the external environment encourages the government to expand expenditures

additionally through money creation and budget deficits, inevitably leading to inflationary pressures and balance of payments (BOP) crises down the road. The basic point here is that policy evolution over time, while not completely endogenous, is intimately linked to an economic system's initial conditions, a point well supported by both the empirical record of contrasting countries in Latin America and East Asia and the intermediate cases of Southeast Asia (see Ranis and Mahmood 1992). As a long-time observer of Indonesia, for example, I noted that whenever the price of oil was high, state oil firm Pertamina was favored by policymakers, which culminated in policies unfavorable to growth or equity; while, when the price of oil was relatively low, development planning agency Bappenas's advice was accepted and reforms had a much better chance of being pursued and implemented.

Initial conditions, in other words, affect not only income levels but also policy responsiveness and flexibility over time, i.e., the extent to which policies can be seen as accommodating or obstructing the gradual changes that all societies must undergo if they are to have a successful transition into modern growth. It is not only the relative strength or weakness of a system's natural resource endowment but also the strength or weakness of that system's ability to attract long-term foreign capital "for the asking" which is relevant here. The common culprit is the large rents emanating from the primary export sector and/or foreign capital inflows and the resulting animated struggle for these rents among various interest groups. Foreign capital flows, especially those not of the direct investment type, are often strongly correlated with the size of natural resource bonanzas and thus reinforce oscillations rather than acting in a counter-cyclical fashion. My basic argument is that once the relationship between initial conditions and policy responses can be made more transparent the chances for a better understanding and support of the entire development process are enhanced. After all, liberalization in monetary, fiscal, and foreign exchange policies over time is not a function of religious belief but required by the need to maintain dynamic efficiency in an increasingly interdependent global economy.

What we can do about natural resource bonanzas is to recognize their impact, *à la* Norway and Botswana, and try to neutralize them by prudential fiscal and monetary means. What we can do about foreign capital inflows, especially of the public variety, is to insist on much greater passivity by donors, accompanied by real ownership of reform programs on the part of recipients. In spite of protestations to the contrary, the IFIs today still dominate the composition of reform programs and still try to assert their own views and impose some level of conditionality to get them accepted by recipients. Can the recipient really be allowed to take the initiative, with the international donor community willing to respond? Realistically, this would also mean accepting a generally lower level of lending of the "business as usual" type. Passivity on the part of the IFIs and enhanced initiative by the developing countries does not, of course, mean that the international community would sign on the dotted line; but if the credibility of multilateral development bank policy-based lending is to be restored, current cynicism about the annual ritual dance, i.e., demanding conditionality and promising aid flows early in the year and then being driven to disburse later on regardless of what has been delivered, has to be overcome.

The new millennium seems to me a propitious moment to reexamine the way our foreign aid business is done. I do not believe, for example, that the PRSP process focused on the poorest countries is sufficiently recipient-driven, given the fact that it still requires extensive IMF and World Bank tutelage ex ante; in fact, the IMF has prepared a voluminous book of instructions outlining just how a PRSP is to be prepared (see Ranis and Stewart 2001). Nor do I believe that the new window of the U.S. Millennium Challenge Account, intended to reward countries that have already done well, is going to be very helpful to those who most need help. Developing countries have to be able to enter into an adjustment dialogue with an increased sense of initiative, involvement, and ownership, complete with "self-conditionality." And public capital inflows must be commensurate in volume, and especially in time, with the exposure to risk and the threat from veto players in the course of reform.

Fifth, the current trend toward gradual liberalization, of course, implies an increasing substitution of indirect for direct controls and a reduction of such controls generally, as we have seen in the case of foreign trade. With respect to the macro policy picture, the general conclusion from experience is that the state's ability to print money and the compulsory purchase of foreign exchange are more damaging than the temporary retention of high import duties during the development process. The acceptance of a monetary philosophy according to which money supply and foreign exchange reserves are increasingly regarded as mediums of exchange, instead of as purchasing power that can be artificially manipulated to achieve socially desirable goals, is gradually being adopted. At the same time, generally low tax/GDP ratios can be enhanced, in tandem with fiscal reforms, moving the system from indirect border taxes to indirect domestic taxes en route, and, ultimately, to domestic taxes that are more income-elastic and less regressive.

With respect to the related question of the most appropriate exchange rate policy, this is apparently one of those areas on which it is especially difficult to pontificate as one assesses the experience of the past and attempts to look into the future. There seems to be a clear bias toward floating rates, with more or less dirty interventions as a realistic companion. This stance avoids the need to accumulate large foreign exchange reserves in defense of the peg and gives a larger role to market determination—which is, in essence, similar to monetary decontrol moving the system in the direction of an equilibrium interest rate.

Sixth, on the hoary question of the role of the state relative to the role of markets, the above discussion indicates that policy interventions should be focused more on institutional construction and certainly should not follow Amsden's 1989 advice to move purposely against price signals. In short, I find the Acemoglu, Johnson, and Robinson (2002) rejection of geography in favor of institutions, by reference to the reversal of fortunes in different parts of the pre- and post-colonial world, utterly convincing. Secondly, interventions on behalf of a particular industry or even individual firms should undoubtedly be minimized even as selective state intervention to correct market failure undoubtedly played a role in both the historical Japanese and postwar East Asian "miracle" cases. The 1997 *World Development Report* (World Bank) recognized the importance of the state as part of the "new institutional economics."

But favorite episodes of where directed credit policies went right, as in the Pohang steel company in South Korea or the automobile industry in Brazil, are still likely to be swamped by the admittedly large herd of industrial white elephants that are trampling the small folks and potential newcomers in all parts of the developing world.

The point is sometimes made, by Rodrik and others, that Latin America, following Washington Consensus prescriptions, has done poorly, while interventionist East Asia has done well. This conclusion is challengeable on several counts. First of all, as we have already noted, natural-resource-rich Latin America has been famous for a continuous stop/go pattern as between enhanced liberalization and increased interventionism, probably yielding the worst of both worlds, private and public, on average. Secondly, the secular trend toward the reduction of political power in various markets has been much clearer in East Asia, even if, on average, tariffs for example, might have been higher. Gradualism in East Asia seems to have done better than shock therapy in Eastern Europe. What also matters, I believe, in comparing development experiences over the past 30 to 40 years, is the relative extent of isolation or cohabitation between industry and a meritocratic public service, the basic competence of a bureaucracy capable of pragmatic course reversal, as well as the extent to which the government is concerned about ensuring workably competitive conditions in the private sector.

The current emphasis on organizational and institutional change decidedly does not imply a diminished role for government or some sort of laissez-faire prescription, but a different and undoubtedly more effective role for policymakers. Gradual but persistent depoliticization in the monetary, fiscal, and foreign exchange arenas requires sensitivity to differing local conditions and the willingness and ability to stay the course.

Last, on the international scene, current negotiations on trade, intellectual property rights, and other issues, both multilateral (WTO-related) and regional, are under intense scrutiny. To examine each of these in detail in terms of current thinking and policy would take us too far afield; but it seems clear that overloading the World Trade Organization circuit with issues such as labor and environmental standards, just because it is the only organization with teeth, will only result in immobilizing the organization. A clear and present danger to trade emanates from the mounting "spaghetti bowl" of free trade agreements, as well as the current escalating debate on outsourcing. On the latter, I would enter a plea for urgent international action providing for vastly improved national adjustment assistance programs, possibly financed out of foreign aid budgets, which would establish consistent rules for countries facing the inevitable adjustment costs accompanying trade liberalization. Adjustment assistance, of course, has been tried before in many countries, but in very few cases has it really focused not so much on extending unemployment insurance to those affected by job losses as on ensuring the portability of benefits and encouraging emerging new industries, including non-traded and export sectors, to hire displaced workers with the help of temporary subsidies. The mutual benefits of freer trade are indisputable, but if "trade not aid" is to become more than a catchy slogan, we urgently need to imaginatively address the political fallout from displaced workers.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 47 (4): 1231–94.

Aghion, Philippe, and Patrick Bolton. 1997. "A Trickle Down Theory of Growth and Development." *Review of Economic Studies* 64 (2): 151–72.

Aghion, Philippe, and Peter Howitt. 1998. *Endogenous Growth Theory*. Cambridge, Mass.: MIT Press.

Amsden, Alice H. 1989. *Asia's Next Giant: South Korea and Late Industrialization*. New York: Oxford University Press.

Arrow, Kenneth. 1962. "The Economic Implications of Learning-by-Doing." *Review of Economic Studies* 29 (3): 155–73.

Banerjee, Abhijit, and Andrew Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101 (2): 274–98.

Barro, Robert J. 1991. "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics* 106 (2): 407–43.

———. 1997. *Determinants of Economic Growth: A Cross-Country Empirical Study*. Lionel Robbins Lectures. Cambridge, Mass.: MIT Press.

Bauer, Peter T. 1957. *Economic Analysis and Policy in Underdeveloped Countries*. Westport, Conn.: Greenwood Press Reprint.

Behrman, Jere R. 1996. "The Impact of Distributive Policies, Governmental Expenditure Patterns, and Decentralization on Human Resources." Philadelphia, Penn.: University of Pennsylvania, Background Paper for United Nations Development Programme (UNDP), *Human Resources Development Report, 1996.*

Behrman, Jere, and Barbara L. Wolfe. 1987. "How Does Mother's Schooling Affect the Family's Health, Nutrition, Medical Care Usage and Household?" *Journal of Econometrics* 36 (1–2): 185–204.

Bhagwati, Jagdish. 1978. *Foreign Trade Regimes and Economic Development: Anatomy and Consequences of Exchange Control Regimes*. Cambridge, Mass.: Ballinger.

Birdsall, Nancy. 1985. "Public Inputs and Child Schooling in Brazil." *Journal of Development Economics* 18 (1): 67–86.

Boeke, Julius Herman. 1953. *Economics and Economic Policy of Dual Societies*. New York: Institute of Pacific Relations.

Boozer, Michael, Gustav Ranis, Frances Stewart, and Tavneet Suri. 2004. "Paths to Success: The Relationship Between Human Development and Economic Growth." Economic Growth Center Working Paper 874. Yale University, New Haven, Conn.

Bourguignon, François, and Luiz A. Pereira da Silva. 2003. *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*. Washington, D.C.: World Bank and Oxford University Press.

Chenery, Hollis B. 1971. *Studies in Development Planning*. Cambridge, Mass.: Harvard University Press.

Chenery, Hollis, Montek S. Ahluwalia, C. L. G. Bell, John H. Duloy, and Richard Jolly. 1974. *Redistribution with Growth: Policies to Improve Income Distribution in Developing Countries in the Context of Economic Growth*. New York: Oxford University Press.

Cohen, Benjamin, and Gustav Ranis. 1971. "Import Liberalization and Growth: The Second Post-War Restructuring." In Gustav Ranis, ed., *Government and Economic Development*. New Haven, Conn.: Yale University Press.

Deaton, Angus, and Christina Paxson. 1994. "Intertemporal Choice and Inequality." *Journal of Political Economy* 102 (3): 437–67.

Deninger, Klaus. 1999. "Making Negotiated Land Reform Work: Initial Evidence from Colombia, Brazil, and South Africa." *World Development* 27 (4): 651–72.

Deninger, Klaus, and Lyn Squire. 1997. "Economic Growth and Income Inequality: Reexamining the Links." *Finance and Development* 34 (1): 38–41.

Domar, Evsey. 1957. *Essays in the Theory of Economic Growth*. Oxford, U.K.: Oxford University Press.

Easterly, William. 2001. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. Cambridge, Mass.: MIT Press.

Eckaus, R. S. 1955. "The Factor Proportions Problem in Underdeveloped Countries." *American Economic Review* 45 (4): 539–65.

Eswaran, Mukesh, and Ashok Kotwal. 1985. "A Theory of Two-Tiered Labor Markets in Agrarian Economies." *American Economic Review* 75 (1): 162–77.

Evenson, Robert E., and Y. Kislev. 1975. *Agricultural Research and Productivity*. New Haven, Conn.: Yale University Press.

Evenson, Robert E., and Gustav Ranis. 1990. *Science and Technology: Lessons for Development Policy*. Boulder, Colo.: Westview.

Fei, John C. H., and Gustav Ranis. 1961. "A Theory of Economic Development." *American Economic Review* 51 (4): 533–65.

———. 1964. *Development of the Labor Surplus Economy: Theory and Policy*. Homewood, Ill.: Richard A. Irwin, Inc.

Fei, John C. H., Gustav Ranis, and S. W. Y. Kuo. 1979. *Growth with Equity: The Taiwan Case*. New York: Oxford University Press.

Fields, Gary S. 2001. *Distribution and Development: A New Look at the Developing World*. New York: Russell Sage Foundation and MIT Press.

Fischer, Stanley, Ratna Sahay, and Carlos A. Vegh. 1996. "Stabilization and Growth in Transition Economies: The Early Experience." *Journal of Economic Perspectives* 10 (2): 45–66.

Foster, Andrew, and Mark Rosenzweig. 1996. "Comparative Advantage, Information, and the Allocation of Workers to Tasks: Evidence from an Agricultural Labor Market." *Review of Economic Studies* 63 (3): 347–74.

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review* 83 (3): 525–48.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technical Change." *Econometrica* 25 (4): 501–22.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press.

Haberler, Gottfried. 1988. *International Trade and Economic Development*. San Francisco: International Center for Economic Growth.

Habibi, Nadir, Cindy Huang, Diego Miranda, Victoria Murillo, Gustav Ranis, Mainak Sarkar, and Frances Stewart. 2003. "Decentralization and Human Development in Argentina." *Journal of Human Development* 4 (1): 73–101.

Harrod, Roy F. 1939. "An Essay in Dynamic Theory." *Economic Journal* 49 (193): 14–33.

Higgins, Benjamin. 1956. "The 'Dualistic Theory' of Underdeveloped Areas." *Economic Development and Cultural Change* 4 (2): 99–115.

Hoddinott, John, and Lawrence Haddad. 1991. "Household Expenditures, Child Anthropometric Status, and the Intrahousehold Division of Income: Evidence from the Cote d'Ivoire," Research Program in Development Studies, Woodrow Wilson School, Discussion Paper 155, Princeton University.

Kravis, Irving. 1970. "Trade as a Handmaiden of Growth: Similarities Between the Nineteenth and Twentieth Centuries." *The Economic Journal* 80 (320): 850–72.

Kremer, Michael, and Edward Miguel. 2001. "Worms: Education and Health Externalities in Kenya." NBER Working Paper 8481, National Bureau of Economic Research, Cambridge, Mass.

Krueger, Anne. 1978. *Foreign Trade Regimes and Economic Development: Liberalization Attempts and Consequences*. Cambridge, Mass.: Ballinger.

Krugman, Paul R. 1994. *Rethinking International Trade*. Cambridge, Mass.: MIT Press.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45 (1): 1–28.

———. 1971. *Modern Economic Growth: Rate, Structure, and Spread*. New Haven, Conn.: Yale University Press.

Lall, Sanjaya. 1992. "Technological Capabilities and Industrialization." *World Development* 20 (2): 165–86.

Lewis, Arthur. 1954. "Economic Development with Unlimited Supplies of Labour." *Manchester School of Economic and Social Studies* 22 (May): 139–91.

Lipton, Michael, and Martin Ravallion. 1995. "Poverty and Policy." In Jere R. Behrman and T. N. Srinivasan, eds., *Handbook of Development Economics, Volume 3*. Amsterdam: North Holland.

Little, Ian Malcolm David, Tibor de Scitovsky, and M.F.G. Scott. 1970. *Industry and Trade in Some Developing Countries*. Oxford, U.K.: Oxford University Press.

Lucas, Robert E. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (1): 3–42.

Mahalanobis, P.C. 1955. "The Approach of Operational Research to Planning in India." *Sankhya: The Indian Journal of Statistics* 16 (1,2): 3–62.

Malthus, Thomas Robert. 1815. *An Inquiry into the Nature and Progress of Rent and the Principles by Which It Is Regulated*. London: John Murray.

Mandelbaum, K. 1945. *The Industrialization of Backward Areas*. Oxford, U.K.: Blackwell.

Morduch, Jonathan. 1999. "The Microfinance Promise." *Journal of Economic Literature* 37 (4): 1569–1614.

Mukherjee, Anindita, and Debraj Ray. 1995. "Labor Tying." *Journal of Development Economics* 47 (2): 207–39.

Myrdal, Gunnar. 1957. *Economic Theory and Underdeveloped Regions*. London: G. Duckworth.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. New York: Cambridge University Press.

———. 1991. "Institutions." *Journal of Economic Perspectives* 5 (1): 97–112.

Nurkse, Ragnar. 1953. *Problems of Capital Formation in Underdeveloped Countries*. New York: Oxford University Press.

Okun, Arthur. 1975. *Equality and Efficiency.* Washington, D.C.: The Brookings Institution.

Pack, Howard, and Westphal, Larry E. 1986. "Industrial Strategy and Technological Change." *Journal of Development Economics* 22 (1): 87–128.

Persson, Torsten, and Guido Tabellini. 1994. "Is Inequality Harmful for Growth?" *American Economic Review* 84 (3): 600–21.

Pitt, Mark, and Shahidur R. Khandker. 1998. "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of the Participant Matter?" *Journal of Political Economy* 106: 958–96.

Prebisch, Raúl. 1962. "The Economic Development of Latin America and Its Principal Problems." *Economic Bulletin for Latin America* 7 (1): 1–22.

Ranis Gustav, and S. A. Mahmood. 1992. *The Political Economy of Development Policy Change.* Cambridge, Mass.: Blackwell.

Ranis, Gustav, and Frances Stewart. 1994. "Decentralization in Indonesia." *Bulletin of Indonesian Economic Studies* 30 (1).

———. 2001. "The Debt-Relief Initiative for Poor Countries: Good News for the Poor?" *World Economics* 2 (3): 111–24.

Ranis, Gustav, Frances Stewart, and Alejandro Ramirez. 2000. "Economic Growth and Human Development." *World Development* 28 (2): 197–219.

Ravallion, Martin, and Shaohua Chen. 1999. "When Economic Reform Is Faster Than Statistical Reform: Measuring and Explaining Inequality in Rural China." *Oxford Bulletin of Economics and Statistics* 61: 33–56.

Ravallion, Martin, and G. Datt. 1999. "When Is Growth Pro-Poor? Evidence from the Diverse Experience of India's States." Policy Research Working Paper WPS 2263, World Bank, Washington, D.C.

Ricardo, David. 1951. "On the Principles of Political Economy and Taxation." In Piero Sraffa, ed., with M. H. Dobb, *The Works and Correspondence of David Ricardo*, *Volume I.* Cambridge, U.K.: Cambridge University Press.

Rodrik, Dani. 1996. "Understanding Economic Policy Reform." *Journal of Economic Literature* 34 (1): 9–41.

———. 1999. *Making Openness Work: The New Global Economy and the Developing Countries.* Washington, D.C.: The Overseas Development Council.

Romer, Paul. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71–S102.

Rosenstein-Rodan, Paul. 1943. "Problems of Industrialization of Eastern and South-Eastern Europe." *Economic Journal* 53 (210/211): 202–11. Reprinted in A. N. Agarwala and S. P. Singh, eds., *The Economics of Underdevelopment*. Bombay: Oxford University Press.

Rosenzweig, Mark. 1988. "Labor Markets in Low Income Countries." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics, Volume 1.* Amsterdam: North Holland.

Rosenzweig, Mark, and Kenneth Wolpin. 1993. "Credit Market Constraints, Consumption Smoothing and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India." *Journal of Political Economy* 101 (2): 223–44.

Schultz, T. Paul. 2001. "School Subsidies for the Poor: Evaluating a Mexican Strategy for Reducing Poverty." Discussion Paper 102, Food Consumption Nutrition Division, International Food Policy Research Institute (IFPRI), Washington D.C.

Schultz, Theodore W. 1964. *Transforming Traditional Agriculture*. New Haven, Conn.: Yale University Press.

Schumpeter, Joseph A. 1959. *The Theory of Economic Development*. Cambridge, Mass.: Harvard University Press.

Sen, Amartya Kumar. 1985. *Commodities and Capabilities*. Amsterdam: North-Holland.

———. 1992. *Inequality Reexamined*. Cambridge, Mass.: Harvard University Press for the Russell Sage Foundation.

Singer, Hans Walter. 1950. "U.S. Foreign Investment in Underdeveloped Areas: The Distribution of Gains between Investing and Borrowing Countries." *American Economic Review, Papers and Proceedings* 40 (2): 473–85.

Skoufias, Emmanuel. 2001. "PROGRESA and Its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI." International Food Policy Research Institute, Washington D.C.

Smith, Adam. 1880. *The Wealth of Nations*. Oxford, U.K.: Clarendon Press.

Solow, Robert. 1956. "A Contribution to the Theory of Economic Growth." *The Quarterly Journal of Economics* 70 (1): 65–94.

———. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39 (4): 312–20.

Srinivasan, T. N. 1994. "Human Development: A New Paradigm or Reinvention of the Wheel?" *American Economic Review, Papers and Proceedings* 84 (2): 238–43.

Stewart, Frances. 1977. *Technology and Underdevelopment*. London: MacMillan.

Stiglitz, Joseph E. 1991. "Some Theoretical Aspects of the Privatization: Application to Eastern Europe." Paper presented for the Third Villa Mondragone International Conference, Rome, June 1991.

———. 2002. *Globalization and Its Discontents*. New York: W. W. Norton and Co.

Streeten, Paul. 1994. "Human Development: Means and Ends." *American Economic Review, Papers and Proceedings* 84 (2): 232–37.

Sugden, Robert. 1993. "Welfare, Resources, and Capabilities: A Review of Inequality Reexamined by Amartya Sen." *Journal of Economic Literature* 31 (4): 1947–62.

Townsend, Robert. 1994. "Risk and Insurance in Village India." *Econometrica* 62 (3): 539–91.

Udry, Christopher. 1994. "Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria." *Review of Economic Studies* 61 (3): 495–526.

ul Haq, Mahbub. 1992. "Human Development in a Changing World." UNDP Occasional Paper No. 1. United National Development Programme, New York.

UNDP. *Human Development Report*. Various Issues. United National Development Programme, New York.

Wade, Robert. 1990. *Governing the Market*. Princeton, N.J.: Princeton University Press.

Williamson, Oliver E. 1975. *Markets and Hierarchies*. New York: Free Press.

World Bank. 1997. World Development Report. *The State in a Changing World*. New York: Oxford University Press.

Comment on "A Half-Century of Development" by Richard N. Cooper and "The Evolution of Development Thinking: Theory and Policy" by Gustav Ranis

ABHIJIT V. BANERJEE

I have the privilege of commenting on two enormously rich essays: one by Richard Cooper, on what has happened to developing countries during the past 50 years; and the other by Gus Ranis, on what has happened to the economics of developing countries. Both essays draw from the authors' ringside views of the "unfolding" that they each describe, in their roles as leading thinkers about development policy and as advisers to the governments that were implementing the policies discussed. What they have produced are unique documents that are part intellectual autobiography, part personal testament.

With something so personal, I do not believe that it is even appropriate for me to do what a commentator usually does, which is to try to make the papers more "correct." Instead, my goal here is to retell the history they describe, but only keeping what I see as the essentials—and converging on what I hope are the key points for thinking about what is to come. In other words, I offer what one might call a cartoon history of the past 50 years of development.

A Cartoon History of Economic Development

In the beginning (circa 1950), there was import substitution and planning. The presumption was that markets did not really work in poor countries, and therefore state action was needed in order to mobilize dormant resources.

Development economists wrote about dual economies, where economic activity somehow slowed down once we crossed the boundary between the modern and traditional sectors.

Multi-sector plans based on giant input-output models designed to drag the recalcitrant peasants into the production of steel, electricity, and other wonderful things, were the rage of the day.

Abhijit V. Banerjee is the Ford Foundation International Professor of Economics, and Director of the Poverty Action Lab, both at the Massachusetts Institute of Technology in Cambridge, Massachusetts.

A few voices in the wilderness spoke out for the rationality of the peasant and the relevance of conventional economics in the developing world, but these were largely unheeded.

Growth was good, at least until the first oil crisis, and even afterward, but starting in the late 1960s, there were murmurs that things were not quite going as planned. India experiences its first major crisis in 1966–67; Bangladesh is born in 1971 and soon pronounced a basket case; hyperinflation hits parts of Africa during the 1970s.

The 1980s

The party really ends somewhere around 1981, with debt crises and default by a number of countries.

Since it was widely held that government profligacy had ruined the party, it was decided that from now on, good policy required that markets be protected from governments, rather than governments making markets work better.

Development economics became all about fiscal discipline and currency management, and the Washington Consensus was born.

Brave central bankers fought and tamed inflation—and eventually brought some semblance of stability to national currencies.

Unfortunately, somewhere along the way, we managed to mislay growth: Economies in Africa and Latin America began to shrink.

The 1990s

Growth recovered slightly during the 1990s, but remained patchy and unpredictable—at least based on what we knew from development economics circa 1990.

In particular, China and India—with state-controlled banking sectors and some of the largest budget deficits in the world—dominated the growth scene, while the largely well-behaved economies of East and Southeast Asia went into crises. More generally, many countries felt that their fiscal rectitude was not adequately rewarded by increased growth.

And so, development economists began talking about the post–Washington Consensus consensus, which was the idea that the economy flourishes when we succeed in creating an environment that will allow individuals to make the most of their talents.

Moreover, bad things are bad: Arbitrary confiscations, civil wars, and printing money at will do not work.

And good things are good: delivering education to the masses, providing an efficient, stable business climate; developing well-functioning asset markets and court systems; building effective safety nets; peace . . .

Which just leaves us to sort out the details:

- What do you mean by delivering education?

- What do you mean by education?

- Who are you?

And to answer the process questions:

- How do you deliver education where there are no teachers?

- How do you build a stable business climate where people are starving?

- How do you create an efficient court system where corruption is a way of life?

In other words, we are not far from where we started, but we now have a better idea of what we are looking for.

Conclusion

The essays by Cooper and Ranis show their shared sense of progress: Cooper's for the developing world, and Ranis's for the world of development economics. Both development and development economics have come a long way from where they were 50 years ago.

At one level, this cartoon history is a history of progress. The world is a much richer place today than it was in the immediate postwar period, and quite a bit of that growth has taken place in countries that used to be poor. As a result, there is less poverty in the world today, at least as poverty is conventionally measured.

However, as Cooper notes, a lot of this progress, especially in the past two decades, is a result of sustained, fast growth in China and India. From the point of view of the poor in those two countries, this is obviously good news. But from the point of view of worldwide development, this is hardly news at all. These are just two data points among many—two successes to be weighed against so many failures. The fact that these countries are large should not really make a difference, because we should probably think of each of them as a single policy experiment. If you are a random poor country looking to grow, the fact that China is growing at 9 percent per year (with India not far behind) may be inspiring, but it is unlikely to make you feel confident about your future. Indeed, you might even be slightly discouraged given that your exporters now need to compete with the galloping Chinese and resurgent Indians.

This pessimism is obviously not unrelated to the fact that while development economics has also come a long way, we have unlearned more than we have learned. We are no closer to a useable blueprint for development than we were 50 years ago. It is true that we now think we know which blueprints to avoid, but it is not always clear that we know why. Like so many others, we have preferred to focus on the next great idea rather than try to understand where the last idea went wrong.

This is, of course, where history can be useful. So if there is one thing that disappointed me in these two essays, it is the fact the authors chose to take the past 50 years of economic history at face value. For example, both essays note the early success of *dirigisme* and its subsequent decline, but neither asks why, for so long and in so many places, this approach did so well. Was this purely accidental, or were there in fact underused resources, as economists of the day believed. And if there were indeed underused resources, have they since been depleted? And if not, should we find ways to utilize them? If we do want to do something, what? And if not by central *diktat,* how?

Comment on "A Half-Century of Development" by Richard N. Cooper and "The Evolution of Development Thinking: Theory and Policy" by Gustav Ranis

PEDRO S. MALAN

> *There is a tide in the affairs of men*
> *Which, taken at the flood, leads on to fortune;*
> *Omitted, all the voyage of their life*
> *Is bound in shallows and in miseries.*
> *On such a full sea we are now afloat,*
> *And we must take the current when it serves,*
> *Or lose our ventures.*
>
> Shakespeare, *Julius Caesar*, Act IV, Scene 3

Professors Richard Cooper and Gustav Ranis wrote interesting papers, well in the tradition of relevance to current discussion and long established by the World Bank's Annual Conference on Development Economics. These two authors show that the world has become far more complex since World War II. The process of change over the past half-century is far more impressive than it may seem; and a better understanding of these changes and complexities is essential to envisioning possible futures.

The Starting Point

There are indeed tides in the affairs of men. One such tide was there, in the early postwar years described by professors Cooper and Ranis. This is the right starting point for their papers because, for a variety of reasons, the tides of the 1930s and the 1940s had led to what appeared to be a new and exciting field of unlimited promise in the social sciences: the economics of development.

The new field appeared so promising in the early postwar years because it seemed to draw on, or result from, a variety of historical changes and theoretical

---

Pedro S. Malan, formerly Minister of Finance for Brazil, is now Chairman of the Board of Unibanco. He also teaches economics at the Catholic University of Rio de Janeiro.

breakthroughs in thinking about the conditions for economic progress. The experiences of the 1930s had shown the scope for governmental activism: in Nazi Germany; in socialist Russia; and even in the United States with Roosevelt's New Deal, the war years, and later the Marshall Plan. The 1940s and 1950s saw the creation of many national planning commissions, five-year development plans, and the establishment of many public corporations in sectors deemed critical for the structural transformation of an economy.

The "Keynesian revolution" in macroeconomics has given many young entrants into the new field the reassuring sense that there were truly respectable analytical foundations on which to move forward both the theory and the policy of development economics. National accounting gave quantitative expression to macroeconomic concepts.

A whole new literature emerged on the terms of trade between agriculture and manufactures—and on the general subject of development strategies. Confidence was high: Development was within the grasp of the truly committed.

Therefore, it is not surprising that many eloquent speeches of today express, in a sense, a desire to revive the spirit of the early postwar years when it comes to expectations about turning hopes and dreams into realities.

Two recent examples of such speeches are the opening of this ABCDE conference, delivered by Jim Wolfensohn; and the speech delivered by Larry Summers here at the Bank earlier this year.[1] Both called for a more effectively functioning constituency for development—especially in rich countries—as well as a deepening of both the understanding of actual problems and, in particular, the motivation to address them in the real world. This motivation was clearly there in the early postwar period, from the place where both Cooper and Ranis begin their stories of the tides that have marked the past half-century of economic development.

Ranis: Growing Complexities

In the draft of Ranis's paper presented at the Conference, the first section was entitled "The Early Postwar Consensus." This was followed by a second section, on "The Washington Consensus as Subsequently Amended;" and a third, on "The Search for Silver Bullets." The conclusion was "Best Guesses on the Way Forward."

I have serious doubts about the appropriateness of using the word "consensus" to describe the prevailing situation in the early postwar years—or again more than 40 years later, when the somewhat unfortunate expression "Washington Consensus" was used for the first time by John Williamson (1990).

In the early postwar years there was in general a climate of confidence in the future, but surely no consensus on policies. I believe, like Albert Hirschman, that the subdiscipline of development economics came into being "as a result of an a priori unlikely conjunction of distinct ideological currents . . . (which) proved to be extraordinarily productive but also created problems for the future . . . because (given) the circumstances under which it arose, development economics became overloaded with

unreasonable hopes and ambitions" (Hirschman 1981, p. 2). I will return to this point later in this comment, given its importance to current discussions.

The "literature" about the Washington Consensus and its amendments, extensions, corrections, and so forth has proved to be one of the most unproductive, simplistic, rhetorical, and politically motivated nonissues of recent debates about economic policies for development.

The author of the term "Washington Consensus," himself a respected economist, recognizes that the brand name was a misnomer and led to discussions that had little to do with the purpose of the original article—which, by the way, is an interesting piece that few critics bother to read and understand (Williamson 1990).

The expression "consensus" was seen by many as a defense of a single, unique model from which there could be no departures. The word "Washington" associated with it was seen as proof that something was being imposed unilaterally on the rest of the world by both the capital of the "empire" and the international financial institutions based in the city. As Williamson sadly wrote recently, ". . . it has become an obstacle to clear thought." It is time to drop the expression from the vocabulary, "as (hopefully) we leave behind the stale ideological rhetoric of the 1990's." (Williamson 2003, p. 13)

Ranis's section on the search for a silver bullet is relevant to this discussion. The conclusion is that there is no silver bullet (i.e., there is no single way, there are no simple prescriptions nor single models). It is not that economic theory does not apply. It does. But basic economic principles do not automatically translate into specific, detailed, and effective macro, and, especially, microeconomic policymaking in different institutional, political, and social environments (Rodrik 2003).

In this section Ranis describes some of the tides that have marked the past 50 years or so, especially those that moved to center stage as broader indicators of social and human development, and for which GDP growth is a necessary but not sufficient condition.

Ranis's concluding section, on "best guesses on the way forward," suggests that (among other things) we should be moving away from large samples of cross-section analyses that incorporate an increasingly large number of qualitative and quantitative variables, and toward a smaller set of comparative historical studies of countries with similar characteristics, typologies, and initial conditions. I agree.

Cooper: The Importance of Perspective

Cooper's very interesting paper follows a more pragmatic, empirical approach: He begins with what were the available quantitative growth projections of the future as of the 1940s (he found only three) and compares these projections ex post with the observed data. In the process Cooper discusses, in an illuminating way, some important measurement problems that are involved in the thorny issues of "convergence" of standards of living and, especially, of comparative analysis of relative poverty levels both within countries and across countries. Before concluding with lessons from the past half-century and an intriguing counterfactual ("what if") section, Cooper

addresses some controversial issues related to capital inflows and makes brief comments on development and freedom.

Cooper shows that actual performance turned out to be significantly higher relative to expectations and prevailing projections of the 1940s. Despite an increase in world population from 2.5 billion in 1950 to 6.0 billion in 2000, world income per capita increased significantly (by nearly 3 percent per year in the 1950s and '60s, and by about half that amount during the last two decades of the century).

It is true, as Cooper notes, that aggregate figures such as these (or for the broad, regional groupings presented in his paper) may conceal as much as they reveal given the tremendous diversity in growth performance and social development both across countries and over time in the history of different countries.

Cooper rightly criticizes an old metaphor ("a rising tide lifts all boats") as "not an accurate characterization of most human affairs which more commonly involve leaders and followers, innovators and imitators, with imitators demonstrating very mixed capabilities in following successfully, as well as occasionally altering the path significantly. . . ." In more than one sense, Shakespeare's beautifully worded characterization of tides in the affairs of men is far more appropriate and balanced than the old metaphor: Some may "lose their ventures" in the way, and could find themselves "bound in shallows and in miseries."

Indeed, it is sadly true that in many parts of the world development remains an elusive goal and a daunting task, amid the deprivation of far too many people living in countries that lag behind. But Cooper is surely right in noting that while current discussion should focus on either unfinished business or the much that remains to be done, the discussion should not (and could not) lose the historical perspective entirely, downplaying the overall performance of the world in the past half-century as unimpressive or proclaiming that the last few decades have been "lost for development." Simply because empirical evidence shows that, as a general statement, this assertion is just plain wrong.

In this connection, it is worth noting that Eric Hobsbawn, the well-known British historian, wrote in his recent autobiography: "In the thirty years after the second World War, the world and what it was like to live in it changed more rapidly and fundamentally than in any other period of comparable length in human history. . . . If you think you can go back, we can tell you, it can't be done" (Hobsbawn 2002, p. 414).

It may well be that the last 30 years, from the mid-1970s to the first half of the current decade, will come to be seen as another period of fundamental qualitative and quantitative change in the affairs of men, with no parallel except for the previous 30-year period (1944–74).

For the young, always forward-looking, who have not lived as adults through these periods, the observations above may seem vague and platitudinous, because the young take these past developments for granted.

But the key point is that he or she who does not understand how fast, or how deeply and how irreversibly, the world changed from the 1940s to the 1990s will not be able to envision possible futures—or to engage in actions that bring about a better future in other "possible worlds."

The concluding section of Cooper's paper contains a "necessarily speculative" set of four counterfactuals that are worth reading and thinking about. They refer to trade (better market access to rich countries); aid (meeting U.N. targets); private capital flows (less overborrowing, less overlending); and oil prices (what if they had not been quadrupled in 1974?). All of these are worth pursuing further in future World Bank conferences on development economics.

Past and Future: The Never-Ending Dialogue

Forty-four countries were assigned membership quotas at the 1944 Bretton-Woods conference that formally created the World Bank and the International Monetary Fund. Only four member countries were in Africa, only three in Asia, and only two in the Middle East (Horsefield 1969, p. 96). Today there are about 110 member countries at the World Bank and the Fund—from these three regions alone. The populations of these countries have higher hopes, dreams, and expectations of a better future for themselves, their children, and their countries. These people may not care to know what GDP is all about, but they know something about how their living conditions compare with those of others within their own countries and elsewhere in the world.

It took several thousand years for the world population to reach the impressive figure of 2.5 billion in 1950, but it took merely 50 years more for this figure to reach 6 billion people, 80 percent of whom live in so-called developing countries. A few billion more people will be added over the next 50 years, almost all of them in the so-called developing world—and all of them with renewed hopes and dreams of a better life.

These are just two simple-minded observations made to illustrate the extent to which the world has become a far more complex place than it was a half-century ago. And the world will likely become even more complex during the next 50 years.

One could well argue that if the world has become much more complex than it was, it is also better prepared now in the sense that technical knowledge, understanding of the problems, and the resources needed to address these problems are increasingly available—largely due to the achievements of the last half-century.

But these achievements have also generated a far higher level of expectation about the future and prompted the anguished demands for urgent action, or for at least urgent, action-oriented thought.

However, experience has shown, again and again, that the true challenge lies in establishing appropriate links between understanding, resources, and, especially, motivation cum technical and organizational capacity—to address the relevant, context-specific issues that are at stake.

It is futile, for example, to design grandiose, comprehensive development strategies, or to lay down preconditions for economic and social development. As Albert Hirschman noted: "they would merely serve to spell out a wholly utopian scheme for changing, from the outside, everything that has characterized a given reality—and would therefore amount to wishing away that reality" (Hirschman 1986, p. 176).

Instead of looking for necessary and sufficient conditions for change, insists Hirschman, "we must train ourselves to be on the look out for unusual historical developments, rare constellations of favourable events, narrow paths, partial advances that may be conceivably followed by others . . ." (pp. 176–77).

I think that this is what Cooper and Ranis have attempted to do in their papers—which I have enjoyed reading, and to which my comments surely do not do justice.

Some readers may feel that these apparently backward-looking studies have limited use given the current overloading of the agenda, the nervous search for urgent "solutions," and for the urgent recovery of an apparently lost excitement.

But the early excitement brilliantly described by Hirschman in "The Rise and Fall of Development Economics," was largely based "on the implicit idea that development economics could slay the dragon of backwardness virtually by itself or, at least, that its contribution to that task was central" (Hirschman 1981, p. 23).

Hirschman concludes: "We now know that this is not so: a consoling thought is that we may have gained in maturity what we have lost in excitement" (p. 23).

This being said, let me add in conclusion: Maturity does not mean surrender to the forces of complacency, inaction, cynicism, or despair. Maturity means finding ways to be constructively engaged in processes of change for the better. Knowing that, despite the fundamental importance of links to the international environment, the truly relevant development battles are fought, won, or lost on the domestic front.

Note

1. President Jim Wolfensohn's opening speech at this year's Annual Bank Conference on Development Economics (published in this volume); and the speech Larry Summers delivered at the World Bank Practitioners of Development series, on February 2, 2004.

References

The word *processed* describes informally reproduced works that may not be commonly available through library systems.

Hirschman, Albert. 1981. "The Rise and Fall of Development Economics." *Essays in Trespassing: Economics to Politics and Beyond*. Cambridge, U.K.: Cambridge University Press.

———. 1986. *Rival Views of Market Society*. New York: Viking.

Hobsbawn, Eric. 2002. *Interesting Time—A Twentieth Century Life*. New York: Pantheon.

Horsefield, J. Keith. 1969. *The International Monetary Fund 1945–1965: Twenty Years of Monetary Cooperation*. Washington, D.C.: International Monetary Fund.

Rodrik, Dani. 2003. "Growth Strategies." John F. Kennedy School of Government. Harvard University, Cambridge, Mass. Processed.

Williamson, John. 1990. "What Washington Means by Policy Reform." In John Williamson, ed., *Latin American Adjustment: How Much Has Happened?* Washington, D.C.: Institute for International Economics.

———.2003. "From Reform Agenda to Damaged Brand Name." *Finance and Development* 40 (3): 10–13.

Infrastructure
and Development

Infrastructure and Development

RÉMY PRUD'HOMME

The notion of infrastructure is presented here as a subset of the notion of capital. Several definitional characteristics of infrastructure are identified and discussed. Curiously, for two centuries, infrastructure as an analytic concept was virtually absent from the economist's tool box.

By contrast, during the 1990s, a vast body of literature introduced infrastructure as a determinant of production functions, with a view to estimate its contribution to economic growth. This paper reviews the difficulties associated with this enterprise, and the not too clear conclusions that emerge from it. The heterogeneity of the concept is emphasized. Unlike productive capital, which is homogenized by market forces, politically driven infrastructure may—and often does—consist of white elephants as well as of highly useful roads.

Why and how does infrastructure contribute to development? It is a space shrinker, it enlarges markets, and it operates like the lowering of trade barriers. In urban areas, it can be shown that infrastructure contributes to enlarge the effective size of the labor market and of the goods or ideas markets, thus increasing productivity and output.

Institutional and financial regimes have a direct impact upon the socioeconomic efficiency of infrastructure. Because infrastructure always has a government dimension and can also have a private dimension, the menu of institutional options available is quite large: from direct government provision (with or without tolls and prices), to unsubsidized concessions with various forms of public-private partnerships, such as subsidized concessions or shadow tolls. Three mechanisms have to be taken into account: (1) the welfare loss often (not always) associated with tolls and prices, which implies that in such cases, all other things equal, non-tolled infrastructure is better than tolled ones; (2) the cost advantage usually associated with private production, which implies that, all other things equal, privately managed infrastructure is better; and (3) the distortionary impact of taxes, which implies that, all other

Rémy Prud'homme is Professor Emeritus at the University of Paris 12.

things equal, a toll-financed infrastructure is better than a tax-financed one. A small model combining these three mechanisms is developed. A simulation using reasonable values for the main parameters is presented. It suggests that the more private options, in particular the shadow-toll option, are economically superior to the more government-oriented options. The problem is complicated, however, when one takes into consideration the public finance dimension of the various options.

Forecasting errors and associated risks are characteristic of infrastructure projects. Costs are generally underestimated and patronage is overestimated, by large amounts. Errors of 50 percent or more seem to be the rule rather than the exception. An understanding of the various reasons that explain such errors is useful to allocate the related risks between government bodies and private partners. Substantive risks (those linked to changes in project design) as well as pure economic risks (those associated with the macroeconomic environment), which are not insurable, should be borne by the public entity. Technical risks (errors in forecasting costs and usage) should be borne by private enterprises. But institutional errors resulting from a strategic behavior of public and even private agents can only be reduced by changes in institutional design and contracts.

"Infrastructure" projects are many and diverse: roads, tunnels, bridges, railways, airports, harbors, canals, subways and tramways, dams, irrigation networks, water pipes, water purification plants, sewers, water treatment plants, dumps and incinerators, power plants, power lines and distribution networks, oil and gas pipelines, telephone exchanges and networks, and district heating equipment, among others.

Infrastructure and infrastructure-related services have always been with us, but the word itself is relatively recent, particularly in English. Although *The American Heritage Dictionary of the English Language* writes that "the term *infrastructure* has been used since 1927 to refer collectively to […] roads, bridges rail lines, and similar public works," it does not appear in the 1952 *Concise Oxford Dictionary,* or in the 1950 *Real Academia Espanola Diccionario*. The word does not appear in the works of the "pioneers in development" (Meier and Seers 1984) writing in the postwar period. It is, for example, absent from the standard treatises of Lewis (1955), Higgins (1959), and Bauer (1957). The term simply was not used at that time.[1]

This contrasts with the formidable success of the word in the 1980s and 1990s, when it invaded United Nations institutions, World Bank organizational charts, academic journals, and daily newspapers. The process has clearly been inflationary. The meaning of the word has been extended so much that it no longer means much. As the *American Heritage Dictionary* puts it: "Today, we may hear that conservatism has an infrastructure of think tanks […] or that terrorist organizations have an infrastructure of people sympathetic to their cause." In this paper the word is used to describe objects such as the ones listed above, which have in common all or most of the following attributes.

TABLE 1.
Infrastructure and Associated Services

Service	Associated infrastructure
Transportation	Roads, bridges, tunnels, rail tracks, harbors, etc.
Water supply	Dams, reservoirs, pipes, treatment plants, etc.
Water disposal	Sewers, used water treatment plants, etc.
Irrigation	Dams, canals
Garbage disposal	Dumps, incinerators, compost units
District heating	Plant, network
Telecommunication	Telephone exchanges, telephone lines, etc.
Power	Power plants, transmission and distribution lines

Source: Author's compilation.

First, they are capital goods. They are not consumed directly. Rather, in combination with labor, and possibly other inputs, they provide services. Table 1 shows the relationship between infrastructure and the associated services.

Indeed, what matters is the service, much more than the infrastructure used or needed to produce it. Policies should focus on the end (service provision), not on the means (infrastructure endowment). The confusion that often arises between service and infrastructure reflects the fact that, in many cases, the role of the infrastructure is predominant in the production of the service, or, to put it otherwise, that these services are very capital intensive.

Second, infrastructure is often lumpy, as opposed to incremental. The usefulness of a dam or a bridge that is eight-tenths built is zero. Since the demand for infrastructure services usually increases gradually, adjusting supply and demand over the course of time is difficult, if not impossible. Lumpiness also implies that siting and construction often take years.

Third, infrastructure is usually long lasting. Its lifespan is often measured in decades, if not in centuries. In Europe, there are still roads and sewers in use that date from the Roman Empire. Aside from infrastructure, there are other long-lasting goods: housing, and some ordinary capital goods can also have very long lives. Nevertheless, this characteristic has major implications, in terms of financing or maintenance, for example.

Fourth, infrastructure is space-specific. Unlike most goods, it is generally immobile. A pair of shoes in A is very much like a pair of shoes in B, because shoes can easily be moved (at a small transportation cost) from A to B. It is therefore quite meaningful to add up the total production of shoes in a country. However, a sewer in A can in no way render services in B. Adding sewers in a country can be misleading if sewers have not been located optimally. In addition, the combination of immobility with long life duration means that infrastructure investments will shape the economic geography, or regional policy, of a country for decades.[2]

A fifth characteristic is that infrastructure, or rather the service it renders, is associated with market failures, in the traditional forms of public goods, externalities

TABLE 2.
Infrastructure-Related Services and Alleged Market Failures

Infrastructure-related service	Alleged market failure
Power, gas	Natural monopolies
Water supply and treatment	Natural monopolies, externalities
Telephone	Natural monopolies, externalities
Rail transport	Natural monopoly, merit good
District heating	Natural monopoly
Garbage collection and disposal	Pure public good, externalities
Cable	Natural monopoly, merit good
Roads	Quasi-public good, externality

Source: Author's compilation.

(including network externalities), decreasing costs (leading to natural monopolies), or merit goods, as shown in table 2.

This is usually considered to imply some form of public intervention. Infrastructure, and infrastructure services, cannot be left to pure market forces alone. This important policy conclusion, which is generally true, must be approached with caution, however.

Many market failures are not as clear-cut as is often claimed. The notion of decreasing costs leading to natural monopolies, for example, might make sense for some parts of a service but not for others. In the case of power, for instance, it makes more sense for transportation or distribution than it does for production. In the case of telephone, this notion is wiped out by technological progress.

Then, the existence of market failures is not an automatic justification of government intervention. The opposite view—which has long been dominant—is akin to the attitude of the jury of a beauty contest, who would look at the first candidate and declare the other candidate a winner. The existence of market failures only provides a presumption of the need for government intervention. But in practice, one must take into account possible government failures and compare the costs and benefits of both options.

Finally, of course, government intervention, when required or desirable, can take many forms. Direct public provision is only one option, and not necessarily the best one.

Sixth, infrastructure, or the service it provides, is usually consumed by both households and enterprises. It is at the same time a final consumption item, and an intermediate consumption item. It increases welfare (directly), and it increases output. The relative importance of these two types of consumption varies with each type of infrastructure, and over space and time, but in general, the consumption of enterprises seems to be somewhat greater than that of households.[3]

These attributes might be used to define, albeit loosely, the notion of infrastructure. They exclude the so-called "social infrastructure," such as schools, universities, clinics, hospitals, and so forth. It does not mean that schools and clinics are not important, but rather that they do not share some of the characteristics mentioned. These "social" entities are not always long lasting, and the service they provide owes generally more to labor input than to infrastructure input.

It is not easy to assess the relative importance of infrastructure capital in our economies. This is in part because of uncertainties attached to the notion, and in part because data on the value of capital stock (as opposed to data on flows) is difficult to estimate everywhere and is scarce in many countries. Easterly and Rebelo (1993) produced estimates of "public investment" in a large number of countries, including decade averages for 1960, 1970, and 1980. Consolidated public investment, consisting of investments by governments and by public enterprises, represented 43 percent of total investments (and 9 percent of GDP). Because the life length of such public investments is likely longer that that of private investments, this would suggest that the stock of "public investments," thus defined, represents about half the total capital stock.

Some countries, including France, publish estimates of the capital stock by type. Government capital stock in 2002 represented 15 percent of total capital stock, and 47 percent of gross domestic product (GDP). Government capital stock is different from infrastructure. It includes administrative buildings, schools, and hospitals—but ignores the capital stock of public enterprises, in many cases a component of infrastructure. Assuming that these two items cancel each other out, this gives some idea of the relative importance of infrastructure in France, which appears to be much smaller than the Easterly and Rebelo estimates. Table 3 presents this data and extends it to Brazil and Mexico. The ratio of flow to stock calculated for France has been applied to flow data in order to produce stock data for Brazil and Mexico. The methodology is crude, but it produces estimates of the relative importance of infrastructure, which for Brazil is similar to French numbers (15 percent of total stock of capital, 50 percent of GDP), and for Mexico much smaller (9 percent of total stock of capital, 7 percent of GDP).

The concept of infrastructure, as opposed to the word itself, has largely, and surprisingly, been absent from the history of economic analysis. Infrastructure, particularly transport infrastructure, plays a key role in Adam Smith's vision of economic development. No roads, no transport, no trade, no specialization, no

TABLE 3.
Infrastructure, Other Capital, and GDP: France, Brazil, and Mexico, 2001–2

	France 2002		Brazil 2002		Mexico 2001	
	Flow	Stock	Flow	Stock	Flow	Stock
	G euros	G euros	G reais	G reais	G pesos	G pesos
Infrastructure	46[a]	718	46	728	24	374
Housing	73	2,101	73	2,097	67	1,920
Productive capital	164	1,884	164	1,883	179	2,058
Total	296	4,706	296	4,708	270	4,352
GDP	1,521	1,521	1,346	1,346	5,286	5,286

Sources: France, flow and stock: INSEE (2003); Brazil, flows: ibge.gov.br; Mexico, flows: inegi.gov.mex; Brazil and Mexico, stock: flow figure multiplied by the stock/flow ratios for France.

a. Capital flow and stock of "government," excluding capital of publicly owned enterprises such as SNCF, PEMEX, or Petrobras.

economies of scale, no productivity progress—and no development. Yet during the 19th century and much of the 20th century, infrastructure virtually disappears from economics. In Marx, Walras, Marshall, Keynes, and Domar, output is produced only by labor and capital, and the capital these economists have in mind is mostly, or only, the so-called "productive" capital of private enterprises. This is strange because, during the 19th century, governments in the then-developing countries—today's developed economies—did invest heavily in infrastructure, especially in urban areas. This somehow largely escaped the attention of dominant, mainstream macroeconomists.

Even during the post–World War II period, when development economics appeared as a branch of economics, references to infrastructure and its role are scarce. "Capital" plays a key role in most growth theories and analysis, but "capital" is undifferentiated. Roads and factories are lumped together in the common concept of capital. The obvious differences described earlier were ignored. And because factories were weighted more heavily than roads, the discussion of "capital" turned out to be a discussion of factories. Some pioneers, such as Rosenstein-Rodan or Singer, were more perceptive than others, and made timid references to infrastructure. Thus, Rosenstein-Rodan, discussing in 1984 his wartime views, writes: "The third new idea was that before building consumer goods factories, a major indivisible block of social overhead capital or infrastructure must be built and sponsored because private market initiatives will not create it in time" (Meier and Seers 1984, p. 208). But this is an exception. Until the 1970s, infrastructure, even by a different name, hardly existed as an analytic concept or category in economic theory and policy.

In the meantime, however, governments were busy building roads or sewers. They felt the need for principles and tools to improve their infrastructure investments, which led to the development of cost-benefit analysis. The intellectual foundations of this type of analysis date back to the mid-19th century, with the seminal article by Dupuit on the utility of a non-tolled bridge, and the concept of "surplus." But the key roles in the development of cost-benefit analysis—which was mostly applied to infrastructure investments—were played by the New Deal and by the World Bank. In the late 1930s, the U.S. federal government financed massive infrastructure investments, but the U.S. Congress prescribed that only the projects with sufficient social utility could be undertaken. The Keynesian digging and filling of holes would not qualify. The U.S. Army Corps of Engineers, and economists such as Robert Dorfman, carried out the required studies and tried to give a content to the notion of "sufficient utility." Similarly, after the war, the World Bank—and with it many other international, bilateral, and national institutions primarily involved in infrastructure financing—was required to undertake only projects that would meet the test of a cost-benefit analysis. This led to the development and refinement of project appraisal methodologies that still continue today.

The literature on infrastructure, although recent, is enormous. On transportation infrastructure alone, Stough et al. (2002) published a reader, supposedly limited to the "classics," which comprises 650 pages of fine print. The World Bank itself has published extensively on this topic, and its 1994 World Development Report on *Infrastructure for Development,* prepared under the leadership of Greg Ingram, remains a major contribution. Indeed, financing infrastructure for development could be defined

as one of the main activities, if not the main business, of the Bank. In fact, presenting a paper on this topic at the World Bank sounds like bringing coal to Newcastle.

This paper obviously will not attempt to cover all the important dimensions of the subject. It will deal only marginally with the issues of privatization and regulation, and will largely ignore the key question of pricing. Relatively few infrastructure projects are pure public goods that cannot be priced. Most are chargeable—and the World Bank, among others, has actively argued in favor of charges, for the sake of replicability. But replicability does not say much about the structure of charges: Is marginal social-cost pricing really the only and most efficient pricing method? The paper will also not address the qualitative dimensions of infrastructure supply. Most studies have considered infrastructure endowment in quantitative and dichotomic terms: as present or absent. In reality, in many cases, the problem is not so much to provide the infrastructure as to improve the quality of its service.

Instead, the paper will focus on three issues: the contribution of infrastructure to economic growth, the relationship between financing options for infrastructure investments and economic efficiency, and the magnitude of forecasting errors in infrastructure projects and what they mean in terms of uncertainty and risks.

Contribution of Infrastructure to Economic Growth

How Much?

What is the contribution of infrastructure to economic growth? The topic, largely neglected until the late 1980s, suddenly became fashionable after a seminal (although later much criticized) paper by Aschauer (1989). Dozens of contributions were produced in the following decade.[4] In a survey article Gramlich (1994) goes as far as talking of "research bubbles here" (p. 1189). All of these studies have one point in common: They relate to infrastructure capital and ignore infrastructure services. The main line of this research uses an extended production function, in which output Y is not merely a function of labor L and capital K, but also of infrastructure G:

$$Y = f(L, K, G)$$

Various functional forms were used, particularly Cobb-Douglas type functions, and translog functions. Various notions of "infrastructure" were utilized, more often dictated by data availability than theoretical arguments. And various data sets were used: time-series, cross-section data, and panel data. The pitfalls of such analysis are, however, formidable.

First, there is the issue of reverse causality. Even if it appears that infrastructure G and output Y are correlated, it does not mean that more infrastructure necessarily produces more output. It can also be argued that more output makes it possible to finance more infrastructure. There is a chicken and egg problem here.

Second, infrastructure investment is a component of output. An increase in infrastructure investment mechanically raises aggregate demand and output, even if it does not contribute to increased productivity and output.

Third, many infrastructure projects are undertaken in order to increase welfare, and welfare is only a relatively distant cousin of output or GDP. Many welfare improvements are not (or are poorly reflected) in increased GDP. Time savings, the main justification of most transport investment projects, are a case in point. If a new road makes it possible for me to drive faster to visit my aging mother, this is good for me, and for her, but it does not contribute much to the GDP of France. The same is true of many of the services provided by infrastructure and consumed by households.

Fourth, data on the U.S. dollar value of the stock of infrastructure are scarce and questionable. This is absent in most developing countries, which is one reason why so few of the many studies on this topic deal with those countries. Infrastructure, as mentioned earlier, is often long lasting, and permanent inventory methods do not fare well for 100-year-old infrastructure investments. What is the value of the Suez Canal? Is it its historic cost, assuming that is known? How should we treat depreciation, and repairs? Or, is its value what it would cost to build the canal anew? Any number placed on the value of the Suez Canal or any similar infrastructure will be highly dependent upon the answers given to such questions, and will therefore be questionable.

Fifth, infrastructure is heterogeneous from the standpoint of its relationship to economic development. It includes elements that are likely to contribute a great deal, and white elephants that are unlikely to contribute at all. It will be argued that private capital and even labor are also heterogeneous. But this is not true (at least not to the same extent), in the sense that market mechanisms can homogenize these inputs, precisely from the standpoint of their contribution to output. The marginal utility of a dollar of capital in one form is in principle equal, and in practice not very different from the marginal utility of a dollar of capital in a completely different form. These market mechanisms do not operate in the case of infrastructure projects, which are politically decided. Cost-benefit analysis is supposed to offer an alternative equalizing mechanism, though few practitioners would argue that it fulfills that function perfectly.

Finally, what matters for economic development is infrastructure usage, but what we observe is infrastructure supply. And the latter is a poor proxy for the former. Between the two, there are demand schedules and pricing policies, which are ignored by production function analysis. Let us consider a given infrastructure, such as a bridge. Its contribution to economic development will obviously not be the same if it is free rather than priced, and if it is overutilized (i.e., congested) or underutilized. Here again, the analogy with private sector capital does not hold. In the private sector, overinvestment carries heavy cost penalties and underinvestment means benefits forgone. Overinvestment and underinvestment usually do not last for very long. Infrastructure is not subject to similar market discipline. In addition, infrastructure, as mentioned earlier, is usually more lumpy and intrinsically prone to overinvestment, followed by underinvestment. The divorce between supply and usage is therefore quite common.

Researchers have of course been aware of these pitfalls and have done their best to avoid them. They have used sophisticated econometric techniques or independent

data to deal with the chicken and egg problem and the common trend issue. They have tried to use physical indicators (such as road length or road space) rather than monetary indicators to bypass the infrastructure valuation difficulty. And they have introduced infrastructure usage variables when feasible (see for example Nadiri and Mamuneas 1994). Variations in prices, however, have apparently been ignored.

The output of this production function industry is therefore somewhat inconclusive. Most studies—not all[5]—suggest that infrastructure contributes to economic development. But the magnitude of this contribution varies from one study to another. The elasticities of GDP to infrastructure differ greatly, but this could merely reflect the different notions of infrastructure used in the studies. More worrisome is the fact that the rates of return that can easily be associated with these elasticities also vary significantly, from 0 to 50 or 60 percent. The overall conclusion that emerges from this important line of research is that infrastructure seems to have a relatively high rate of return—something like 15 percent—comparable to or even higher than the rate of return of private "productive" capital. The verb "seems" emphasizes the prudence with which this conclusion should be taken.

When one thinks of the many infrastructure investments that clearly do not contribute much to economic development—empty roads, luxury administrative buildings, and so forth—an average of 15 percent (or of even 10 percent) is quite encouraging. This suggests that about half of all infrastructure investments have rates of return that exceed 15 percent. Because the variation of the distribution of such rates of return is large—certainly much larger than that of the rates of return for private "productive" investments—this even suggests that a significant share of infrastructure investments have rates of return higher than 20 or 25 percent. This would imply that there is underinvestment for certain types of infrastructure in certain areas, as well as overinvestment in other cases.

How?

Rather than "how much," it may be more useful to ask "how" infrastructure contributes to economic development. Production function analyses are black boxes, with infrastructure as one of the inputs and GDP as an output. Let us try to open the lid of these black boxes. In so doing, we shift from infrastructure capital to infrastructure services.

As figure 1 indicates, infrastructure affects the "development" of both households and enterprises—the consumers of infrastructure-related services—and does so through three main mechanisms.

First, infrastructure-related services improve the welfare of households, often in dramatic fashion: Water supply and sanitation, power supply, and transportation change the lives of beneficiaries, especially in cities. *Stricto sensu,* these welfare improvements do not contribute much to GDP, although one can argue that they improve the quality of the labor force and hence its productivity. Many economists, of course, consider improving welfare as part of "development," even if it does not contribute to "growth."

FIGURE 1.
How Infrastructure Contributes to Development
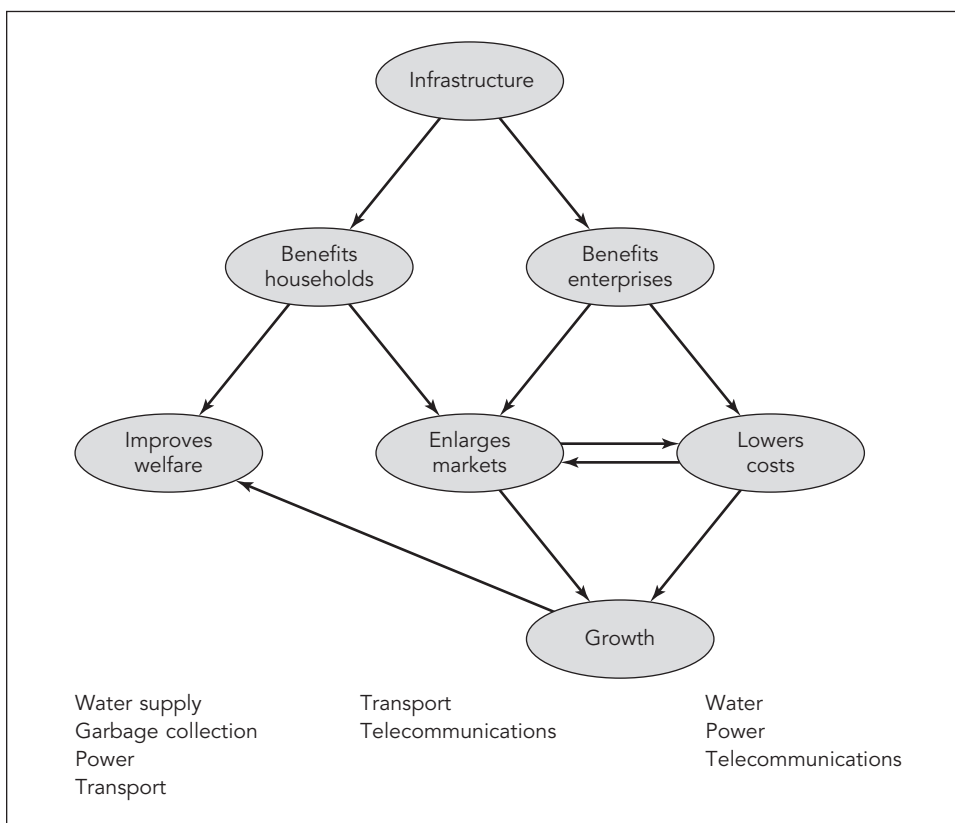
Source: Author.

The two other mechanisms, which are interrelated, have a direct impact on GDP. First, infrastructure supply lowers the cost of some of the inputs used by enterprises. In power or transportation or communication, infrastructure can lower costs by impressive amounts. In so doing, infrastructure acts exactly like technological progress. Lower input costs mean lower total costs, which in turn mean larger markets and further cost reductions.

The most interesting and perhaps most important mechanism of the infrastructure-development relationship is market enlargement. This applies to goods markets and to labor markets, and even (via telecommunication infrastructure) to capital markets. As noted by Adam Smith more than two centuries ago, transport infrastructure enlarges goods markets, by lowering transport costs, and by speeding access for perishable merchandise. The progress of telecommunication infrastructure has intensified this enlargement process. With it come intensified competition, greater specialization, and economies of scale.

Improved infrastructure functions exactly like lower tariffs. It facilitates economic exchanges, and yields the same type of economic benefits. All the analyses that assess

the economic benefits of increased trade can be used to show the contribution of improved infrastructure to economic growth.

Perhaps less known is the impact of infrastructure—in this case urban transport infrastructure—on urban labor markets. In a world where more than half the population, and a much larger share of output, is located in cities, the efficiency of cities has a macroeconomic importance. As is well known, the productivity of a city (per worker, but also per unit of capital) increases with the size of the city. Why? The most plausible reason is that larger cities have larger labor markets. The larger the labor market, the greater the probability that each individual will find exactly the kind of job that matches his or her capabilities and the greater the probability that each enterprise will find exactly the kind of workers it needs. A larger labor market ensures a better match of labor demand and supply, and this in turn ensures greater productivity. However, what matters here is not so much the potential size of the labor market (the total number of jobs), but rather its *effective* size (the number of people who can on average access jobs at a reasonable time and money cost). This is where infrastructure enters the picture. For the effective size of the labor market is (more or less mathematically) a function of three factors: the total number of jobs in the urban area, the relative location of jobs and houses, and the speed at which people move to their jobs—a function of urban transport services, which themselves depend upon transport infrastructure capital.

This simple model of urban productivity was tested and confirmed on the case of 22 urban areas in France (Prud'homme and Lee 1999). The elasticity of productivity (corrected to eliminate the influence of differences in industry-mix) with respect to effective size of labor market was about 0.18, and the elasticity of the size of the labor market with respect to transport speed was around 1.6. This means that the elasticity of productivity to transport speed was about 0.29. Increasing transport speed by 10 percent increases productivity and output by nearly 3 percent, as illustrated in figure 2.

The relationship between transport infrastructure and transport speed is obvious, although not well known. In a study of the Paris area, we estimated that road investments over the period 1983–91 had increased traffic speed (relative to what it would have been in the absence of such investments) by 5 percent. Using the elasticity mentioned earlier, this made it possible to estimate the increased productivity and output due to these road investments, and to derive a rate of return of about 60 percent. The numbers are certainly fragile, but the causal linkages are probably quite robust.

Financing Options and Economic Efficiency

The question of whether an infrastructure investment should be undertaken or not has typically been discussed independently of who will undertake and finance it. Yet, it can be shown that the institutional and financial context and constraints have a direct bearing upon the economic desirability of the project, and also on public budgets. Whether infrastructure is constructed and operated by the public sector,

FIGURE 2.
Impact of Transport Infrastructure on the Efficiency of Cities
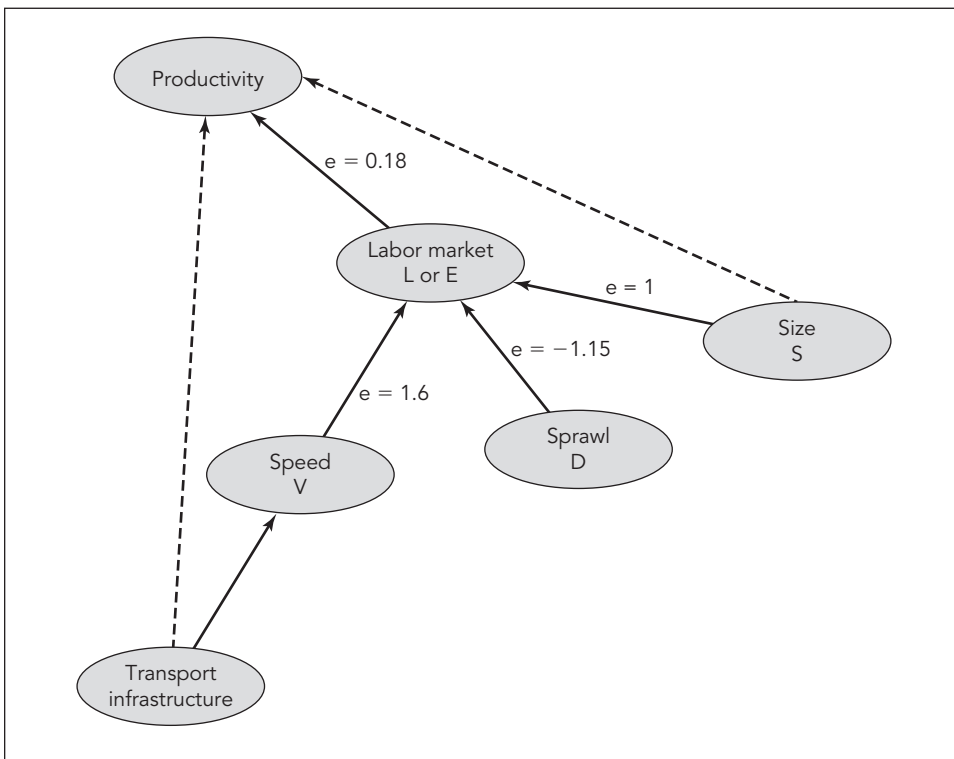
Source: Prud'homme and Lee 1999.

contracted out to private enterprises, or jointly constructed or operated and financed by both actors may make the project more or less valuable. Whether it is paid for and financed by users or by taxpayers also has a direct impact on the socioeconomic viability of the infrastructure project. In a sense, this discussion illustrates the distinction between infrastructure capital and services. In all cases, the physical infrastructure capital is the same; but the services it provides, or the cost at which they are provided, vary with the financing and institutional regime chosen.

A Menu of Options

Let us begin by describing the menu—or rather a menu—of institutional and financing options available, using the example of a bridge or a road. Seven options are considered.

- *Pure public option.* In the pure public option, the bridge is built in year one by a government entity that operates the bridge; usage is free.
- *Pure private option.* In the pure private case, the bridge is built and operated by a private enterprise, in the framework of a contract or concession or authorization

granted by a public body. Users pay a toll, the proceeds of which will compensate the private enterprise. There is a toll level that maximizes toll profits, but the effective toll level is usually negotiated with the granting entity—and must be sufficiently high to ensure the financial viability of the private enterprise investment (i.e., meet a financial rate of return constraint).

- *Public cum toll option.* The infrastructure may be built and operated by a public body that imposes a toll on users. Over the course of time, toll proceeds will accrue to government coffers, and might be assumed to substitute for ordinary taxes. The toll level may or may not be the same as in the pure private option. It is usually lower because the financial rate of return constraint for the public body is lower. (A variant of the public cum toll option, not considered here for the sake of simplicity, allows the public entity to borrow from a bank the money needed for the investment and use toll proceeds to pay interest on the loan).

- *Private cum subsidy option.* In the private cum subsidy option, the private enterprise that builds and operates the bridge at an agreed toll level argues that it needs a subsidy to meet its financial rate of return constraint, and obtains one. This subsidy can be a percentage of the initial investment and be paid up front. (In a variant not considered here, the subsidy is paid over the course of time, as a percentage of toll proceeds or as a prescribed amount).

- *Shadow-toll option.* In the shadow-toll option, the private enterprise builds and operates the bridge. There is a toll, but the toll is not paid by users, for whom bridge crossing is free. It is instead paid by the granting authority, based on the number of users. The toll level is also negotiated, and can be lower than the toll level of the pure private option because the number of users will be greater.

- *Delayed public option.* In the delayed public option that often prevails when governments find themselves cash-strapped, the bridge construction and operation is simply postponed; aside from the delay, this option is similar to the pure public option.

- *Do-nothing option.* Obviously, not constructing the bridge, and letting potential users continue to make a long detour to cross the river (or not cross it), is always an available option. It is even the baseline option, or the one with which the other options can and should be compared.

Economic Impact Mechanisms

As is well known, a cost-benefit analysis of this bridge can be synthesized with two indicators: the discounted net value (economic DNV) of the flow of costs and of benefits; and the internal rate of return (economic IRR), the social discount rate that equalizes the discounted value of costs and of benefits. This is not the place to discuss the relative merits of both indicators, which in practice usually tell very much the same story. The important point here is that the institutional and financial options for the infrastructure being considered are not identical from an economic viewpoint. For

the same bridge, the cost-benefit analysis of the various options will not produce the same results. There are three reasons for this, three important basic mechanisms.

User's exclusion. The toll charged for the use of the bridge will eliminate some users. Since the economic cost of supplying bridge service is normally not affected by usage, excluding some users implies a welfare cost. The surplus generated by the bridge is inversely related to the toll level. Since our options carry different toll levels (including zero levels), they have different effects on the benefits associated with the infrastructure. If this mechanism were the only one at work, the pure public and the shadow toll options (both with a zero toll level) would clearly dominate the other options.

Two remarks can be added. First, what has been said about the user's exclusion is only true for a non-congested bridge or road. If, or rather when, there is congestion, then (at least in principle) a congestion charge is appropriate to maximize the benefits from bridge usage. There is little chance that the prevailing toll will be exactly equal to the optimal congestion toll, but the prevailing toll will nevertheless increase rather than decrease the surplus associated with bridge usage.

Second, caution is required to extrapolate this mechanism to other types of infrastructure. Tolls, called fees, are common with many other infrastructure projects. But the exclusion of users caused by these is only a welfare cost when marginal production costs are zero (or at least lower than the fee charged), and when, as in the case of the bridge, there is no congestion problem.

Greater efficiency of private operation. There are theoretical and empirical reasons to expect private operations to be more efficient—or to consume less economic resources—than publicly managed operations.

There are at least four reasons for this greater efficiency of the private sector.[6] First, the incentive system prevailing in the private sector is more effective than the one that prevails in the public sector; for respectable reasons, the people who deliver are better rewarded (and those who do not more punished) in the private sector; there are strong built-in cost-minimizing mechanisms. Second, and also for respectable reasons, procurement, accounting, and disbursement procedures are more complicated and formal in the public sector; doing things according to the rules is more important than doing them swiftly and efficiently. Third, and somewhat paradoxically, the private often benefits more from economies of scale than the public; this is because the public may consist of relatively small local or regional governments,[7] whereas the private often consists of large companies operating in the entire country or even the entire world. Fourth, technical knowledge and innovation, the mother of productivity, is by now more common in the private sector than in the public sector. And unfortunately, these reasons are likely to have even more force in developing than in developed countries.

Assuming that maintenance and operation costs are negligible, this means that Ig, the economic cost of a publicly built and operated infrastructure, will be higher than Ie, the cost of the same infrastructure developed by a private enterprise, by a margin of α:

$$Ig = (1 + \alpha)*Ie$$

The value of α varies greatly from case to case and country to country. There might be cases when $\alpha < 0$, particularly if and when the private entity is an

uncontrolled monopoly. But in general, the distribution of α seems to be centered around a positive value. Twenty percent sounds like a reasonable order of magnitude.

This means that institutional and financial options in which the bridge construction and operation are done by a public enterprise will (all other things being equal) have economic costs higher by α—higher than what the costs would be if bridge construction and operation were managed by a private enterprise.

Tax distortions. The third mechanism to be taken into account is related to the economic cost associated with tax-financed expenditures. Taxes are generally distortive,[8] and modify the incentive system in ways that decrease output and associated welfare. This deadweight loss, or opportunity cost of tax income, equal to $\lambda*$ tax proceeds, varies with both the tax/GDP ratio and the structure of the tax system. The value of λ might be as high as 20 percent.[9] This means that when a government entity spends 100 financed by tax income, the economic cost to the economy of this expenditure is about 120. Conversely, when this government entity raises 100 in the form of tolls, thereby decreasing other distortive tax resources by 100, there is a welfare gain of about 20.

This has implications for the valuation of the costs and benefits associated with the various options. Costs financed by tax income must be increased by λ, and benefits resulting from a tax reduction must be taken into consideration.

Cost-benefit analysis of the various financing options must take into consideration these three interacting mechanisms, and the outcomes are hard to predict. General formulations, which quickly become complex, do not throw much light on such outcomes. We have preferred a simple simulation that produces different IRR and DNV for our different options, and suggest an economic ranking of these options. Before turning to these numbers, however, we must discuss another dimension of the issue, the budgetary approach.

Budgetary Approach

So far, we have examined the problem in purely economic terms. In practice, the problem has a budgetary dimension as well, often a dominant one. Ministries of finance (even when they are not separate from ministries of the economy) try, all other things equal—and at times when they are not equal—to minimize budgetary expenditures. This means spending less, and spending as late as possible.

An infrastructure investment, however, when it is successful and produces utility, also produces additional taxes and public revenues. Additional utility is not exactly additional economic output, but it is akin to it, and a large fraction of it. As a first approximation, we can say that, every year, additional tax output ΔR is a fraction γ of additional utility or welfare ΔW:

$$\Delta R = \gamma*\Delta W$$

The value of γ varies with the type of infrastructure investment, with the nature of the tax system, and with the level of government considered. It is much higher for a

central government than for a local government, because local government tax rates are much lower than national ones, and because welfare benefits usually leak out of the area where the investment is made. A plausible order of magnitude could be $\gamma = 20$ percent. This would be commensurate with a 30 percent tax-to-GDP ratio, and a two-thirds ratio of GDP to welfare.

With a value of γ one can determine, in each of the financing options discussed, the flow of government revenue generated by the infrastructure investment considered, and compare it with the associated government expenditures. This is done by calculating the DNV (Discounted Net Value) with a discount rate, but this rate need not be identical to the rate of discount utilized for the economic DNV.

Comparing Financial Options

To compare our seven institutional financial options, we can compare the economic IRR, the economic DNV, and the budgetary DNV associated with each. Let P(Q) be the demand curve for the crossing of the river, as shown in figure 3.

Before construction of the bridge the price of crossing, Po (which implies a long detour), is high, and the traffic, Qo, is modest. We are in A. After the bridge, with a toll Pp, we move to B, with a traffic Qp. If there is no toll (Pt = 0), we move to C. The yearly utility or social benefit associated with the bridge is OPoABQp, or OPoAQo+QoABQp. Let us assume that the bridge is built in one year, in year 1, at a private investment cost of Ie, with α the sur-cost of public construction and operation, and λ the opportunity cost of tax resources. γ is the ratio of additional tax to additional welfare.

FIGURE 3.
Demand for River Crossing
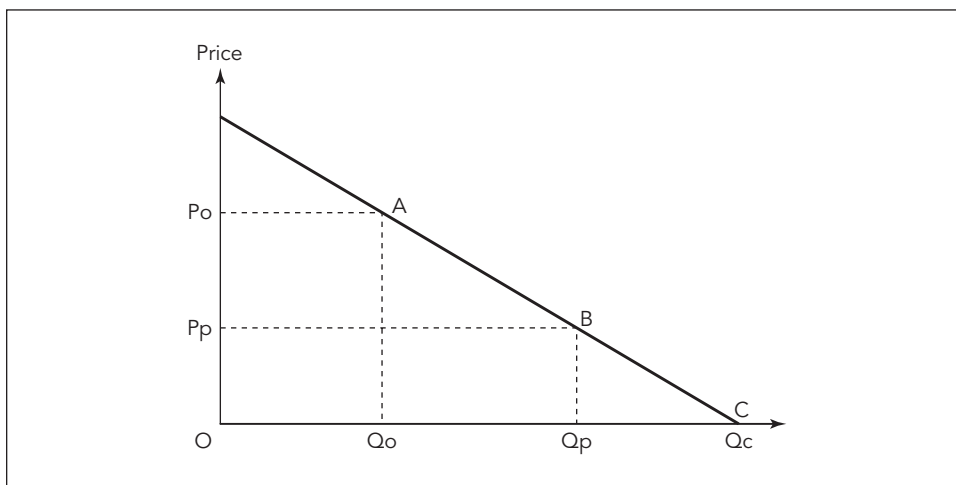
Source: Author.

For a given option, the economic IRR is the value of r for which:

$$\Sigma_t \left[Po*Qo + \int_{Qo}^{Qt} D(P)dQ \right]_t *(1 + r)^t - \alpha*\lambda*Ie = 0;$$

the economic discounted net value DNVe is:

$$DNVe = \Sigma_t \left[Po*Qo + \int_{Qo}^{Qt} D(P)dQ \right]_t *(1 + r^\circ)^t - \alpha*\lambda*Ie$$

with r° an appropriate social rate of discount; and the budgetary DNVb is:

$$DNVb = \Sigma_t(Pp*Qp)'_t *(1 + r')^t + \Sigma_t\gamma \left[Po*Qo + \int_{Qo}^{Qt} D(P)dQ \right]_t *(1 + r')^t - \alpha*Ie - S$$

$$- \Sigma_t(Pp*Qp)''_t *(1 + r')^t$$

With $(Pp*Qp)'$ the public toll proceeds (when they exist), S is the subsidy to a private enterprise (when there is one), $(Pp*Qp)''$ the toll paid to a private enterprise (when they exist), and r′ the social rate of discount for public funds.

To produce orders of magnitude, we used the following values for the parameters utilized. The demand curve for the crossing of the river is assumed to be:

$$P(Q) = 15 - 5*Q$$

$$Q(P) = 3 - 0.2*P$$

This defines a price-elasticity of demand that varies along the demand curve, but which is about −0.5 for P=5, in the lower ranges of P that matter, a realistic elasticity. We assume the initial situation to be Po = 10 and Qo = 1. The demand curve is assumed to be constant over time.

The cost of the construction of the bridge by a private enterprise Ie is 100. We assume α, the public construction sur-cost to be 20 percent. The opportunity cost of tax resources λ is also assumed to be equal to 20 percent (but the two values could be different). We also assume γ (the marginal ratio of tax income to welfare) to be equal to 20 percent (but γ need not be equal to α or λ). The social rate of discount r° used to calculate the economic DNV is taken to be 6 percent. The social rate of discount r′ used to calculate the budgetary DNV is also taken to be 6 percent (but here too, the two values could be different).[10] Both IRR and DNV calculations are based on a 30-year period.

In the pure public and shadow-toll options, there is no toll, and price P paid by users is therefore 0. Different tolls could be retained for the other options. The profit-maximizing toll (the one that equals to zero the derivative of P*Q(P)) is 7.5. This is the toll level that the private enterprise would choose if it were left to decide. But this would lead to a restricted patronage of the bridge and reduce its economic utility to a low level. We assume that the negotiated toll level in the pure private option will be 5. This is consistent with a 9.3 percent financial internal rate of return for the enterprise, which may be considered sufficient. In the public-cum-toll option, we assume a lower toll level of 4, because the public entity can function with a lower financial internal rate of return. In

the shadow-toll option, the toll is paid not by users but by the government entity to the private enterprise; and because it is paid on all users, the toll can be lower. We take it to be 3.33, the toll level that yields yearly toll proceeds equal to the toll proceeds of the pure private option. In the private-cum-subsidy option, the toll remains at 5.[11]

Table 4 presents the parameters attached to each option, and above all the value of the indicators produced by the model. Other numbers for the parameters would produce different values for the indicators and in certain cases different rankings of the options. Nevertheless, the values shown in table 4 are not unreasonable, and the rankings obtained deserve attention. They suggest several conclusions.

First, different financial options for the same infrastructure investment (here, a given bridge) lead to different economic IRR or economic DNV, and also to different budgetary DNV. Institutions and finance do matter for economics.

Then, the two economic rankings are practically identical.[12] The budgetary ranking is different but tells a story that is not much different from the economic one.

This economic story is that the pure public option does not fare well. It has the lowest economic IRR of all options. It can marginally be improved by the introduction of a toll (what is lost in terms of consumer surplus is more than compensated by what is gained through a reduction in tax-associated damage); in addition, the toll is attractive from a budgetary viewpoint. Delaying the pure public option by a few years is worst in economic (DNV) terms, and not much better in budgetary terms.

TABLE 4.
Comparisons of Various Financial Options

	Pure public	Pure private	Public +toll	Shadow toll	Private +subsidy	Public delayed
Features:						
α	0.2	—	0.2	—	—	0.2
λ	0.2	—	0.2	0.2	0.2	0.2
γ	0.2	0.2	0.2	0.2	0.2	0.2
P (toll level)	—	5	4	3.33	5	—
Q (traffic)	3	2	2.2	3	2	3
P*Q (proceeds)	—	10	8.8	10	10	—
S (subsidy)	—	—	—	—	30	—
U (utility)	20	17.5	18.4	20	17.5	20
I (invt cost)	144	100	144	100	100	122
r° (% discount rate)	6	6	6	6	6	6
r' (% id for budget)	6	6	6	6	6	6
Financial IRR (%)	n.a.	9.3	6.1	9.3	14.0	n.a.
Indicators :						
Economic IRR (%)	**13.6**	**17.4**	**15.1**	**17.9**	**16.3**	**13.4**
Economic DNV	**124**	**133**	**126**	**139**	**127**	**95**
Budgetary DNV	**−61**	**+45**	**+49**	**−78**	**+17**	**−53**

Source: Author's calculations.

Note: n.a. = not applicable, — = zero. IRR and DNV are calculated over a 30-year period; in the public delayed option, the delay is 3 years (i.e., the investment is made in year 4).

The pure private option is—in the example studied—far superior to the public options, in both economic and budgetary terms. Even the combination of a private provision and a subsidy is, in socioeconomic terms, more attractive than the pure public option, although it does not fare very well in budgetary terms. The private provision with subsidy fares better than the pure public option, but not as well as the tolled public option.

The shadow-toll system is the best system in socioeconomic terms. In budgetary terms, however, it fares badly, even worse than the public options.

Finally, in budgetary terms, the do-nothing option, which evidently has a budgetary DNV of zero, is more attractive than the pure public option. This provides a justification for doing nothing, but is a bad justification. Doing nothing is (in the example studied) the worst option in socioeconomic terms; and even in budgetary terms it is worse than either the private options or the public-cum-toll option.

Forecasting Errors, Uncertainties, and Risks

Traditional cost-benefit analysis implicitly assumes that the flows of costs and benefits generated over the course of time by an infrastructure project can be correctly forecasted. This assumption is, however, often erroneous. The comparison between *ex ante* forecasts and *ex post* events can show enormous discrepancies. Some of the methodological refinements of cost-benefit analysis that "improve" accuracy of analysis by 1 or 2 percentage points are applied to data that may be off the mark by 30 or 40 percentage points. This is a worrisome contrast. More generally, forecasting errors are a measure of the uncertainties related to the life of infrastructure projects, and of the associated risks. Some might say that there is nothing new here, and that most business decisions are taken in the face of uncertainty. But it is a matter of degree, and uncertainty in infrastructure decisions is generally much greater than in ordinary business decisions. Reducing errors, dealing with uncertainties, and allocating risks efficiently constitute major tasks of infrastructure policy decisions.

Magnitude of Forecasting Errors and Uncertainties

Errors in infrastructure projects are defined as the difference between *ex ante* and *ex post* numbers. They relate to costs and completion dates (delays are a major source of additional costs), and to benefits, which, in many cases, and certainly in transportation projects, are closely associated with patronage and traffic.

Systematic studies of such errors are scarce (Pickrell 1990; Flyvbjerg and Skamris 1997; Flyvbjerg, Hom, and Buhl 2002; Flyvberg, Bruzelius, and Rothengatter 2003; Odeck 2004), because they are difficult to conduct. Cost-benefit analysis assumes that there is a well-defined project to be analyzed, decided, and implemented. This is a fiction. In practice, the story of many infrastructure projects, particularly large ones, begins with a concept to which a few costs and benefits numbers are attached. It continues with a draft project, in which these numbers are refined. The numbers later are further modified, because additional information becomes available, and because

additional negotiations are conducted. New numbers appear—and even after a decision has been finalized, there are often further negotiations, new information, changes, improvements, additions, and so forth, producing revised forecasts. The net result is that a simple question such as: "What was the *ex ante* cost of the project?" is often very difficult to answer. In addition to these conceptual difficulties, there are practical difficulties. *Ex ante* data may have never existed, the data may have been lost, or those who have it might be unwilling to communicate it.[13]

The most comprehensive study of such forecasting errors is the one undertaken at Aalborg University under the leadership of Bent Flyvberg, on more than 200 transport projects, in 20 developed and developing countries. The findings of that study, summarized in table 5, are very much in line with the findings of other studies. In his pioneering work on 10 U.S. rail transit projects, Don Pickrell (1990) found average capital cost overruns of 61 percent (compared with 45 percent for rail projects in table 5), and average ridership overestimates of 65 percent (compared with 39 percent in table 5). Odeck (2004), looking at construction costs of 620 road projects in Norway, finds average overuns of 8 percent (compared with 20 percent in table 5). A Transport and Road Research Laboratory study of subways in developing countries produced construction cost underestimates and ridership overestimates of similar magnitude.

The picture is therefore quite clear and consistent. In transport projects errors on construction costs and on ridership are very common, very large—and are systematically on the "wrong side," with costs underestimated and patronage overestimated. Errors are significantly larger for rail projects than for road projects. There is apparently no progress in the accuracy of forecasting over the course of time. The size of projects does not seem to matter; indeed, Odeck (2004) finds greater errors in small projects than in larger ones. Errors seem to be both largely independent of the project country and equally important in developed and developing countries.

These conclusions relate to transport infrastructure projects. Studies of cost and patronage forecasts in other infrastructure sectors are less systematic (or less known to us), but the available information suggests that similar errors are common in those as well.

TABLE 5.
Forecasting Errors on Construction Costs and Traffic Forecasts in Transport Projects

	Construction costs			Traffic		
	Number	Error	sd	Number	Error	sd
		(%)			(%)	
Rail projects	58	+45	(38)	27	−39	(52)
Road projects	167	+20	(30)	183	−9	(44)
Fixed links	33	+34	(62)	n.a.	n.a.	n.a.
All projects	258	+28	(39)	210	n.a.	n.a.

Source: Flyvbjerg, Bruzelius, and Rothengatter 2003, chapters 2 and 3.

Note: sd = standard deviation, n.a. = not available.

Explaining Errors and Uncertainties

Why are such massive errors made, and what uncertainties do they reflect? It might be useful to distinguish four main causes, or four main types of errors: substantive, economic, technical, and institutional.

First, there are errors and risks related to the nature or the *substance* of the infrastructure project. The *ex post* project might not be the same as the *ex ante* project. The project may have started as a 2 × 2 lanes project and evolved into a 2 × 3 lanes project. Environmental or safety constraints may have been added to the initial project. In such cases, the drift is not one of costs but a drift in the scope of the project. Similarly, in traffic forecasts, an alternative road, which was not planned and therefore not taken into account, may have been built, changing the context and nature of the initial project. Such errors and the uncertainties they reflect are largely specific to infrastructure projects, meaning that there are *substantive risks* in infrastructure investments.

Second, there are *economic* errors and risks, that is, risks associated with the evolution of the overall economic climate. Most studies of demand and patronage are heavily dependent upon income, and therefore upon income and activity forecasts. The economic development of a country is beyond the responsibility of infrastructure planners. Overly optimistic forecasts usually result in overestimates of patronage, a risk that is often referred to as market risk. It could be argued that a similar risk exists for all goods and services; for example, for toothpaste production. The difference is that in toothpaste production forecasting errors can be much more easily corrected, because toothpaste production does not involve massive, long-lived, immobile capital.

There are also errors linked to the *technical* difficulty of forecasting costs and usage for an infrastructure project. These arise from the fact that many such projects are unique. They are not goods and services that are mass-produced in an easy-to-predict fashion. These projects are made to measure, and are often large in size. This makes them complex, and completing them may take years, which increases the probability that something could go wrong. Infrastructure projects are exposed to strikes, flooding, supplier bankruptcies, and so forth. They are often dependent upon geological unknowns—and often use new and not yet fully mastered technologies. For usage forecasts, planners are dependent upon imperfect models and insufficient information, not to mention uncertainties about the economic, social, psychological, or political environment. The resulting uncertainties, which are also specific to infrastructure projects, mean that there are *technical risks* associated with such projects.

Most of these "economic" and "technical" errors, however, could and should play in both directions. Such uncertainties should lead to both overestimates and underestimates. They should explain the standard error of errors, not the average, which should be zero. But they cannot explain fully the systematic errors that are so common.

A complementary, and probably better, explanation is behavioral and *institutional*. Infrastructure developers make errors because they have an interest in making errors.

Errors in Public Projects

Flyvbjerg and others (2002) put this more bluntly: "Underestimating Costs in Public Work Projects: Error or Lie?" In purely public projects—and most of the projects studied fall into that category—civil servants in the technical ministries involved very much want the projects to be constructed. Their prestige, carriers, power (and in extreme cases, income) are often attached to such projects. Hence the easy to understand tendency to underestimate costs and overestimate utility in order to ensure that "their" projects will be decided. If things go wrong afterwards, these civil servants are unlikely to be affected.

There is a similar asymmetry with decisionmakers (i.e., politicians), who are also quite willing to be misled. They will derive a political benefit from the decision to build, and an even greater benefit from inaugurating the project. But the potential failure of the project, in terms of costs overruns or underpatronage, will probably not be damaging for them. In many cases, politicians will no longer be in office when this failure becomes apparent, because the life cycle of an infrastructure project is usually longer than that of a political term. In addition, the benefits of an infrastructure project are often visible and concentrated, whereas the costs are hidden and diluted. On political scales, the benefits outweigh the costs.

Errors in Private Projects

One would expect privately financed infrastructure projects to be protected from such institutional biases in favor of errors. The private capitalist who underestimates costs and/or overestimates usage is likely to be penalized for his errors, often severely. The private sector employees responsible for the errors will be sacked, unlike what happens in the public sector. In any case, the banks that lend money to the private enterprise will scrutinize the project and double-check its seriousness because it is their own money that is at stake.

Yet, it appears that even private infrastructure projects are not immune from errors. The record is not as bad as in the case of public projects, but it is far from perfect. The most glaring case is perhaps that of the Channel tunnel connecting Great Britain and France. This US$8 billion infrastructure investment (1985 prices) was built with private money from banks and capital markets, without public subsidies. Nevertheless, actual costs were 80 percent higher than projected costs, actual traffic is 40 percent below projected traffic, and the private development company for the tunnel is on the verge of bankruptcy. Studies of concession contracts in Latin America point to similar errors. How can they be explained?

In certain cases cost overruns are for substantive errors. Additional constraints are added, increasing delays and costs. This was a major factor in the Channel tunnel case. In other cases, it is reported that the losses of the private infrastructure enterprise are often realized as profits of the construction companies that are the dominant stockholders of the infrastructure enterprise.

But the main explanation is that there are no purely private enterprises in infrastructure construction and operation. Some public entity is always involved. Some

public agency or ministry always intervenes to define the project, to select the private enterprise, to decide on toll levels, to choose the concession period, to grant subsidies in certain cases, and so forth. When things go wrong, the public sector rarely lets the private enterprise go bankrupt. Instead, the public sector usually bails out the private enterprise and renegotiates the contract. In certain cases (rare it seems), when things go too well, the public sector also intervenes and imposes additional taxes or constraints. The public agent is a very active back seat driver, which does not facilitate good driving.

This is why many of the weaknesses described above apply. Public technicians and politicians very much want the infrastructure project to be undertaken—as much as the private enterprises—and they are ready to distort (unconsciously perhaps) forecasts and concession contracts in order to achieve their goals.

Traffic forecasts, for instance, are typically prepared by ministries of transport, and are often included in the documentation given to prospective bidders. The estimates are as overly optimistic as if intended for direct public provision. In Colombia, note Engel and others (2003, p. 8), "traffic was 40 percent lower than predicted by Invias" (the public agency responsible for highways). In a number of cases, private enterprises need not care much about the accuracy of forecasts because they enjoy minimum traffic guarantees. If traffic is not what it was forecasted to be, the government will pay a subsidy to the enterprise.

Cost estimates do not matter much either for private enterprises, because in many cases they are *de jure* or *de facto* protected from cost overruns. Legal profit guarantees are not uncommon, and when they do not exist, contracts can often be renegotiated. Indeed, renegotiation seems to be the rule rather than the exception. In other words, private enterprises involved in infrastructure projects generally face soft budget constraints. The reasons they should have (as with hard budget constraints) to ensure that their cost and benefit forecasts are accurate are in practice dampened or eliminated.

Engel et al. (2003) go even further and suggest that the recourse to private enterprises in infrastructure projects may be part of a political strategy that implies errors. Politicians, they argue, want infrastructure projects to be undertaken now, before the next election. When purely public, these projects are taken into consideration as part of the budgetary process, where they may be fought by the opposition. But contracting out projects to private enterprises on an error-ridden basis allows the government to increase infrastructure now, at the cost of bailing out private enterprises later, without increasing apparent debt. Errors in this analysis are the counterpart of a politically convenient hidden debt.

Dealing with Uncertainties and Risks

Whether substantial, economic, technical, or institutional, forecasting errors in infrastructure projects are economically damaging. They flout cost-benefit analysis. They mean that projects which in reality have a very low economic internal rate of return (or a negative economic discounted net value[14]) and should never be undertaken are made to appear desirable and are undertaken.

Every effort should be made to reduce the uncertainties involved in infrastructure projects and the forecasting errors associated with them.

Public Projects

For purely or mostly public projects, what does this imply? To reduce substantive risks and avoid costly changes in project design, focus, objectives, and constraints, it is important to involve as many stakeholders as possible from the beginning. The time spent initially trying to achieve a consensus or at least to engage in an open debate may seem a waste of time and money. But if it can help avoid major changes at a later stage, it will actually save time and money.

Technical and economic uncertainties cannot be eliminated. They are a feature of many infrastructure projects, and there will always be uncertainties and therefore forecasting errors of that type. Several actions, however, can reduce the number of uncertainties or their adverse consequences. More publicity should be given to methods used and the hypotheses made in preparing forecasts. *Ex post* comparisons should be made (by independent analysts), so that everybody can learn from errors that are made. *Ex ante* studies should build scenarios, perform sensibility analysis, and produce estimates in the form of ranges rather than as single numbers. Forecasts should not, inasmuch as possible, be made in-house, by the ministry or the agency concerned, but be contracted out to other agencies or independent consultants, or at least submitted to outsiders for review.

It is more difficult to reduce institutional sources of uncertainty, because these involve sovereign decisionmakers and their political interests. In many cases, however, public projects involve decisionmakers in only one ministry or agency rather than in the entire government. Involving other agencies or ministries, in particular the ministry of finance, which do not have as much interest in seeing the infrastructure undertaken, may serve as an effective check. Auditing and reporting, when there are independent courts of accounts, can also play a role.

Private Projects

For privately financed projects, market mechanisms provide, in principle, an important check. In practice, as we have seen, such mechanisms are often dampened by public interference, and are nonoperative. What is the appropriate risk allocation? Risks are high, and if they are borne entirely by the private enterprise, serious and reputable enterprises might refrain from bidding on the project, or ask for exorbitantly high prices. The choice here would be not to do business with these enterprises, and forgo the potential benefit of private participation, or to do business with less reliable enterprises, which could be even worse. On the other hand, if all risks are borne by the public sector, market discipline will not work, as we have seen, and the benefits of private participation will also be forgone. Finding the appropriate balance, the point at which the marginal damage of public risktaking is equal to the marginal benefit of public risk-taking, is a delicate task.

It is often argued that privately managed risky projects will be more costly than similar publicly managed projects. The argument is that the privately managed

projects will bear an insurance premium that publicly managed projects will not bear, because the public sector is its own insurer. The implication here is that private management may be more efficient, but that in the presence of risk it is also more costly. Or that, if the project benefits from a public guarantee, it may not be more costly but—because the guarantee erodes the incentives to efficiency—it will not be more efficient. This argument is not convincing. Government is indeed its own insurer, but not paying the insurance premium is not the same as bearing no costs. Self-insurance is in the end about as costly as commercial insurance.

Discussion of infrastructure-related risks in general is probably not very fruitful. We have seen that risks are diverse, so the discussion might be helped by distinguishing between types of risks.

Technical risks, or risks that do not arise from public decisions, should be borne by private enterprises. These include the risk of extra costs due to natural disasters, or of supplier bankruptcies, and insufficient revenues due to erroneous patronage forecasts. Private enterprises should not be protected from their own mistakes on these types of uncertainties, as this is the only way to induce them to minimize such mistakes. In addition, these risks are generally insurable. Minimum revenue guarantees, cost escalation protection clauses, minimum profit provisions, or government-guaranteed loans should be systematically avoided.

Substantive risks, or risks created by post-contract public decisions, should be borne by the public sector. If the government, for reasons only it can fully appreciate, decides to increase environmental or safety constraints, it should bear the cost of this change. If the government drastically increases road charges, or creates an unplanned alternative road (thereby significantly decreasing traffic on the transport infrastructure considered), the public agency should also bear the cost of its unplanned and unpredictable—and in most cases uninsurable—change.

Pure economic risks—those associated with forecasting errors caused by errors on the evolution of activity and income—are much like substantive risks. They are beyond the control of infrastructure developers, and are very hard to insure because, unlike technical risks, they hit all infrastructure developers at the same time. Having the public sector compensate for such risks would not create perverse incentives, and would be desirable.

In practice, of course, distinctions between these types of risks are not always easy to make, opening the door to litigation and negotiation. Disentangling pure economic risks from economic forecasting errors is likely to be particularly delicate. The more explicit the concession contract, the better. The more open the litigation or renegotiation process, the better. The more independent the arbiter of potential conflicts, the better: a tribunal or a regulatory agency is highly desirable, but not necessary.

Because private involvement in infrastructure projects is potentially a great source of savings and efficiency, some people see "privatization" as a panacea—particularly so when facing a corrupt and inefficient government. But in reality, "private" provision is never pure, and always involves (and should involve) a dose of public decision and control. The efficiency of the private sector is contingent upon the form and magnitude of this public control. Unfortunately, governments that are unable to deliver

efficient public services are also unable to control private enterprises contracted to do so. And these governments are even more unable to create the independent bodies or regulation agencies needed to arbitrate disputes between public and private entities. The sad—and well-known—paradox is that the countries that would most benefit from a large dose of privatization are usually the countries that are least equipped to inject it properly. Conversely, the countries that are best able to conduct and oversee the privatization process are also those where this is needed least. Privatization of infrastructure should therefore be seen not as a panacea but as a desirable goal at the end of a long and arduous road.

Conclusions

This brief paper on a vast subject shows that "infrastructure" does not easily lend itself to generalities. These projects have a number of common features that distinguish them from ordinary private capital, and provide justification for the use of a specific concept, but infrastructure projects are very heterogeneous in type, in context, in financing schemes, pricing practices, and so forth. What is true for road construction may be wrong for power generation; what is true in the year 2000 may be wrong in 2010; what is true in a bottleneck situation may be wrong in an over-supply situation; what is true for an unpriced infrastructure may be wrong for a charged one. Market failures justify infrastructure as a category, but planning failures deprive the category of the powerful homogenizing forces of the market.

This makes it difficult, if not impossible, to design and recommend "infrastructure policies" in general. We cannot, for a given country, define an "optimal" level of infrastructure endowment and say by how much infrastructure investment efforts should be increased or decreased—although in many cases it is quite clear that more would be better. We cannot identify optimal institutional, financial, pricing, or decentralization designs—even if in many cases it seems safe to suggest that greater doses of privatization, of charges, of decentralization, or of independent regulation would be appropriate. For each country, we must proceed sector by sector, even project by project, and bring to bear all the resources of public policy analysis. In infrastructure policies, the devil is in the details.

Notes

1. As late as 1973, the editors of *Urban Studies* deleted the word "infrastructure" from a paper this writer contributed to this well-written British journal, and replaced it with "social overhead capital."
2. It has been noted that in some cases, such as harbors, reuse might be an alternative to mobility.
3. For France, input-output table data showed that, in 2001, household consumption of infrastructure-related services was exactly one-third of the total consumption of infrastructure-related services.

4. Even this author added one stone to the monument to infrastructure and growth (Fritsch and Prud'homme 1997).

5. Holtz-Eakin (1994) for example, on the United States (controlling carefully for state-specific effects), finds that public infrastructure contributes nothing to private output or productivity; note, however, that this definition of "infrastructure" is state and local government capital, which ignores federal highways and most privately owned utilities capital.

6. The dichotomy public-private oversimplifies the issue: In reality, there are also public enterprises that are less flexible and responsive than private enterprise but more flexible than a government ministry.

7. This is not always the case, however. In many developing countries a powerful ministry of public works is in charge of most infrastructure projects throughout the country, although the progress of decentralization erodes this potential benefit.

8. Lump-sum taxes and taxes on negative externalities are exceptions. Unfortunately, lump-sum taxes are a textbook curiosity. There are no tax systems with only lump-sum taxation, and taxes on externalities are very rare indeed.

9. There is of course, no reason why λ should be equal to α. For the United States, λ has been estimated to be 17 percent by Ballard and others (1985) and 47 percent by Jorgensen and Yun (1990).

10. These values are on the low side, especially for developing countries.

11. The private-cum-subsidy option corresponds to the case in which the financial IRR (9.3 percent) that prevails in the absence of subsidy is considered too low by the market. A subsidy of 30 percent is granted, which will increase the financial IRR (to 14 percent) but decrease the economic IRR—because of the economic tax cost associated with the subsidy.

12. The change from the pure public to pure public delayed options significantly deteriorates the DNV but changes the IRR very little. Calculations were made over a 30-year period, beginning with year 1, 2, 3, in which nothing happens.

13. Some people believe that refusal to communicate is, in the name of competitive secrecy, more common in the private sector, and they worry that a greater role of the private sector in infrastructure projects will translate into a greater paucity of data. Others believe that in many countries, particularly developing countries, public sector secrecy is perhaps even more formidable.

14. In the simulation exercise discussed earlier, if investment costs increase by 50 percent and traffic decrease by 40 percent (relative to forecasts used in the analysis), then for all options the IRR will be drastically reduced (to 4 and 6 percent instead of 14 and 17 percent for the pure public and pure private options), and the DNV will be negative (-48 and -5 instead of $+123$, and $+132$ for the pure public and pure private options).

References

Aschauer, D. 1989. "Is Public Expenditure Productive?" *Journal of Monetary Economics*. 23 (2): 177–200.

Ballard, Charles, J. Shovan, and John Whalley. 1985. "Equilibrium Computations of the Marginal Welfare Cost of Taxes in the U.S." *American Economic Review* 75: 128–38.

Bauer, Peter, and Basil Yamey. 1957. *The Economics of Under-Developed Countries*. Chicago: University of Chicago Press.

Easterly, William, and S. Rebello. 1993. "Fiscal Policy and Economic Growth: An Empirical Analysis." *Journal of Monetary Policy* 32 (4): 417–58.

Flyvberg, Bent, Nils Bruzelius, and Werner Rothengatter. 2003. *Megaprojects and Risk—An Anatomy of Ambition*. Cambridge, U.K.: Cambridge University Press.

Flyvbjerg, Bent, Mette Skamris Hom, and Soren L. Buhl. 2002. "Underestimating Costs in Public Works Projects: Error or Lie?" *Journal of the American Planning Association* 68 (3): 279–95.

Flyvbjerg, Bent, and M. Skamris. 1997. "Inaccuracy of Traffic Forecasts and Cost Estimates on Large Transport Projects." *Transport Policy* 4 (3): 141–46.

Fritsch, Bernard. 1999. *La contribution des infrastuctures au développement des régions françaises*. Paris: Presses de l'Ecole Nationale des Ponts et Chaussées.

Frish, B., and R. Prud'homme. 1997. "Measuring the Contribution of Road Infrastructure to Economic Development on France." In E. Quinet and R. Vickerman, eds., *Econometrics of Major Transport Infrastructures*. London: Macmillan. (Reprinted in Stough et al. 2002, pp. 363–85).

Gramlich, E. 1994. "Infrastructure Investment: A Review Essay." *Journal of Economic Literature* 32 (3): 1176–96.

Higgins, Benjamin. 1959. *Economic Development*. New York: Norton & Company.

Holtz-Eakin, Douglas. 1994. "Public Sector Capital and the Productivity Puzzle." *The Review of Economics and Statistics* 86 (1): 12–21.

Institut National de la Statistique et des Etudes Economique. 2003. *Rapport sur les comptes de la nation 2002*. Paris: INSEE.

Jorgenson, Dale, and Kun Young Yun. 1990. "The Burden of Taxation in the U.S." Harvard University, Department of Economics, Discussion Paper 1528.

Kuznets, Simon. 1959. *Six Lectures on Economic Growth*. New York: The Free Press of Glencoe.

Lewis, Arthur. 1955. *The Theory of Economic Growth*. Homewood, Ill: Richard Irwin.

Meier, Gerald, Dudley Seers, eds. 1984. *Pioneers in Development*. New York: Oxford University Press for the World Bank.

Nadiri, Ishaq, and T. Mamuneas. 1994. "The Effect of Public Infrastructure and RD Capital on the Cost Structure and Performance of U.S. Manufacturing Industries." *The Review of Economics and Statistics* 76 (1): 22–37.

Odeck, James. 2004. "Cost Overruns in Road Construction—What Are their Sizes and Determinants?" *Transport Policy* 11 (1): 43–53.

Oxford University Press. 1952. *The Concise Oxford Dictionary* (4th ed.). London: Oxford University Press.

Pickrell, Don. 1990. *Urban Rail Transit Projects: Forecasts versus Actual Ridership and Cost*. Washington, D.C.: U.S. Department of Transportation.

Prud'homme, Rémy, and Chang-Woon Lee. 1999. "Size, Sprawl, Speed, and the Efficiency of Cities." *Urban Studies* 36 (11): 1849–58.

Real Academia Espanola. 1950. *Diccionario Manual e illustrado de la Lengua Española* (2d ed.) Madrid: Espasa-Calpe.

Stough, R., R. Vickerman, K. Button, and P. Nijkamp, eds. 2002. *Transport Infrastructure* (Classics in Transport Analysis Number 4). Cheltenham, U.K.: Edward Elgar.

World Bank. 1994. *World Development Report 1994: Infrastructure for Development*. New York: Oxford University Press for the World Bank.

Comment on "Infrastructure and Development" by Rémy Prud'homme

EDUARDO ENGEL

Rémy Prud'homme provides an authoritative and friendly survey on the relation between infrastructure and economic development. This is a welcome contribution on an important topic, and I strongly recommend reading it.

Among the many things I liked, let me mention three. First, the paper tracks the concept of infrastructure through (economic) history, arguing convincingly that for a long time it was ignored by most leading development economists. Second, it provides a good review of the literature that quantifies the relation between infrastructure and growth. Given the author's important contributions on this topic, this skeptical survey should be taken seriously.

Third, the paper provides a refreshing discussion of the large forecasting errors observed for infrastructure projects. Demand is usually overestimated while costs are underestimated, in both cases reflecting strategic behavior by public and private agents involved. Yet after correcting for these systematic sources of errors, the residual component, even though unbiased, has a very large variance, reflecting the inherent uncertainty that is part of most infrastructure projects.

There are, however, some things I missed in this paper. I would have liked a stronger stance on some of the policy choices considered. Too often the author concludes that "the devil is in the details," which of course is true but not very useful for policymakers. I would have also liked more examples from developing countries. Finally, some important topics are barely mentioned. Admittedly, my complaints reflect differences in preferences and style, and should therefore not be taken too seriously.[1]

As noted in the paper, Professor Prud'homme only considers marginally the issues of privatization and regulation of infrastructure, a topic of major concern for policymakers in developing countries given the widespread disillusion with recent

TABLE 1.
Support for Privatization in Latin America
(percentages of those surveyed)

	Argentina	Bolivia	Brazil	Colombia	Costa Rica	Chile
1998	39	52	49	40	59	50
2001	17	24	49	13	31	43
2002	14	23	38	23	—	22
2003	12	19	33	24	—	28

	Ecuador	El Salvador	Guatemala	Honduras	Mexico	Nicaragua
1998	52	53	61	46	50	46
2001	33	25	22	21	28	31
2002	40	35	29	34	31	20
2003	20	15	16	25	31	20

	Panama	Paraguay	Peru	Uruguay	Venezuela	Average[a]
1998	20	46	43	43	51	*46.3*
2001	37	34	22	23	49	*29.4*
2002	31	19	32	16	38	*28.2*
2003	10	23	22	16	32	*21.6*

Source: LatinoBarometro.

Note: — = not available. Those surveyed were asked: *Do you (a) strongly agree, (b) agree, (c) disagree, (d) strongly disagree with the statement "the privatization of public utilities has been beneficial for the country"?* Possible answers included options (a)–(d) listed here, as well as *(e) do not know, (f) does not respond.* Data reported in this table reflect the percentage of individuals that chose only option (a) or (b).

a. All reported averages exclude Costa Rica, since 2002 and 2003 data were not available for this country.

privatization efforts. Indeed, as shown in table 1, support for privatization in Latin America fell substantially between 1998 and 2003, from an average across 16 countries of more than 46 percent to less than 22 percent. By 2003, more than two-thirds of the population surveyed in every country in the region disagreed with the statement that "the privatization of public utilities had been beneficial.[2]

Thus, an important question for policymakers is, "What went wrong with infrastructure privatization, and how can it be fixed?" In the remainder of this comment, I will present an overview answer to this question, for the particular case of highways.[3] A more ambitious project, considering other infrastructure sectors (such as telecommunications, electricity, and water), would make a good topic for a future Annual Bank Conference on Development Economics conference paper.[4]

Highway Privatization: Recent Experience and Policy Lessons

The "lost decade" of the 1980s led to low investment and inadequate maintenance of infrastructure, creating a major highway deficit across Latin America. This was the origin of the wave of infrastructure privatization that began in the 1990s, as this highway deficit, combined with chronic budgetary problems, led governments to

embrace a scheme whereby the private sector financed urgently needed infrastructure investments, to free up public resources for other priority areas.

Private financing of new highways throughout Latin America freed up fewer government resources than expected (Engel, Fischer, and Galetovic 2003). In several cases, public funds were diverted to bail out franchise holders in financial trouble.[5] So government guarantees for private highway franchises added to the fiscal burden, especially since these guarantees were paid out mainly during economic downturns, when government budgets were under pressure.[6]

Before proceeding, it is useful to clarify what this paper means by public and private provision of roads. Under public provision (the *traditional approach*), the government designs, finances, and operates the road. Private firms may participate in the construction stage and may be selected in competitive auctions. But once the highway is built, the government operates and maintains it. Taxpayers finance the road and, even when users pay tolls, these are usually unrelated to construction costs. By contrast, when roads are privatized, a concessionaire finances, builds, operates, and maintains the facility. The franchise owner collects tolls for a long time—usually between 15 and 30 years—and when the franchise ends the road reverts to the government. Such Build-Operate-and-Transfer (BOT) contracts can be awarded either through direct negotiation between the transit authority and an interested firm, or through a competitive auction for the franchise of a well-defined project.[7]

Highway privatization in Latin America promised not only to free up government resources, but also to deliver some of the standard advantages expected from privatization.[8] First, a firm that is responsible for construction and maintenance has the right incentives to invest in road quality (Tirole 1997). Second, private firms are better managers than state-owned highway authorities. Third, BOT contracts may be desirable on distributional grounds, since roads are paid by those who benefit. In particular, cost-based tolls are easier to justify politically when infrastructure providers are private.[9] Finally, and in contrast to public provision, under BOT only privately profitable roads will be built, thus using the market mechanism instead of central planning to screen projects. This reduces the likelihood of building white elephants, a common occurrence in Latin America (and on other continents).[10]

But the promised benefits of highway privatization in Latin America failed to materialize (Engel, Fischer, and Galetovic 2003). The main reason for the failure was the continuous process of renegotiation of franchise contracts.[11] In most countries concessionaires renegotiated their contracts without public scrutiny, which facilitated the shifting of losses to taxpayers. Such renegotiations negate the public benefits of private highways by giving an advantage to firms with political connections, thereby limiting the risk of losses—and reducing the incentives to be efficient and cautious in assessing project profitability.

Highway franchises need to be regulated. Building deadlines and quality standards must be enforced during the construction phase. Tolls, quality of service, and maintenance must be regulated while the concession is operating. And most important, a mechanism for solving contractual oversights fairly and promptly must be put in place.

Opportunistic renegotiations have been pervasive because of two design flaws that are present in all major franchising programs undertaken in Latin America (Engel,

Fischer, and Galetovic 2003). First, countries have followed a "privatize now, regulate later" approach. The lack of a clear contractual structure has often led to cost overruns and renegotiation of the conditions of the original contracts. Moreover, the government agency interested in the success of the franchise program was usually the same agency that supervised the franchise contracts. Since the success of these agencies is often measured by the number of projects they succeed in building, they tend to be lax in enforcing compliance with franchise contracts and are inclined to ease the conditions for franchise holders.

Examples of lax regulation follow (see Engel, Fischer, and Galetovic 2004). A report published by Argentina's National Comptroller in 2003 (Auditoría General la Nación 2003) concluded that the equipment needed to measure a highway's friction coefficient had been out of service since 1994, so this index had not been measured for any franchised highway for nearly a decade. The same report pointed out that highway quality immediately after construction was often considerably below specifications, and the highways deteriorated faster than stipulated in the contracts. Building delays were also recurrent, while fines the government was entitled to were rarely collected. In Colombia many concessionaires did not obtain financing, and faced no penalty for failing to do so. And in Chile the regulator did not collect data on traffic flows, relying instead on reports provided by the franchise holders to pay out minimum traffic guarantees.

The second pervasive design flaw is that most concessions have been awarded using fixed-term contracts, which shift most of the demand risk to franchise holders and create demand for subsidies and guarantees. This is troublesome, since demand risk for highways is particularly high (Engel, Fischer, and Galetovic 1997, 2001). And since the franchise holder has little ability to influence demand, there is no point in having the franchisee bear this risk. Fixed-term franchises allocated in competitive auctions make it almost certain that firms will lose money in low-demand states, which generates pressure for renegotiations and guarantees.

Optimal risk-sharing (between users, the government, and the franchise holder) is achieved through a flexible-term contract, which can be implemented with a present-value-of-revenue (PVR) auction (Engel, Fischer, and Galetovic 2001).[12] In this type of contract the regulator fixes user fees, announces a discount rate, and awards the franchise to the firm that asks the least present value of toll revenue. The franchise ends when the present value of toll revenue is equal to the winning bid.[13]

PVR franchises have four advantages over their fixed-term counterparts (Engel, Fischer, and Galetovic 1997, 2003). First, by adjusting the franchise length to demand realization, a PVR contract substantially reduces the demand risk faced by the franchise holder, and therefore reduces the demand for guarantees. Second, PVR franchises avoid lengthy negotiations on what the fair compensation should be if the franchise must be terminated early, because, for example, additional lanes need to be built. The difference between the winning bid and the present value of tolls collected at the time of termination is a good estimate of fair compensation. No such measure exists for a fixed-term franchise. Third, PVR franchises are more amenable to toll changes in response to changes in demand than are their fixed-term counterparts,

since tolls may vary substantially without affecting the franchise holder's present value of toll income.[14] For example, in the case of urban highways, a PVR contract could stipulate that tolls will be reset by an independent agency every year (in response to demand conditions), so that users internalize any congestion costs.[15] Finally, it can be argued informally that opportunistic behavior, by both the franchise holder and the government, is less likely under a PVR contract. The main drawback of PVR contracts is that they provide less incentive for maintenance than their fixed-term counterparts. But as long as quality can be easily verified by independent parties—which is the case for highways—this is not a major concern.

Summing up, the Latin American experience with highway privatization during the past decade was disappointing: Both the reduction of the fiscal burden and efficiency gains were considerably below expectations.[16] Weak regulation and pervasive opportunistic renegotiations explain these outcomes, but considerable improvement can be expected if concessions are regulated seriously, and flexible term concessions are used for future infrastructure projects.

Notes

1. With possibly one exception. I believe the exercise comparing alternatives for providing infrastructure services is conducted in a framework that is too simple to warrant many of the conclusions the author obtains. Needless to say, this does not mean that I disagree with the conclusions.

2. Even though part of this decline may be explained by cyclical factors, most analysts believe there is a marked negative trend as well.

3. The evidence I discuss is from Latin America, even though the policy implications are likely to apply in other regions as well.

4. For important insights on some of these topics see Gómez-Ibáñez (2003), Gausch (2004), and Laffont (2004).

5. For example, Mexican taxpayers spent more than US$8 billion to bail out the franchise owners and the banks that lent to them.

6. See, for example, "World Bank Warns of New Debt Dangers." *Financial Times*, May 30th, 1997.

7. Under most circumstances the latter option should be preferred to the former. See Demsetz (1968) for a forceful argument in favor and Williamson (1976) for a critique.

8. For example, an official 1999 document from ALIDE (Latin American Association of Financial Institutions for Development) states (in the following translation from Spanish by the author): *"The fiscal and financial crisis […] of the eighties led to the end of the traditional model of infrastructure financing, that considered the state as the main investment agent, and opened space for important participation by the private sector […] with the objective of not only bringing relief to the burden supported by public finances, but, more importantly, to improve the allocation of risk and improve the efficiency of management."*

9. This is important if trucks are ever to pay tolls that reflect the road deterioration they cause.

10. Where a white elephant is defined as a project whose net (of costs) social value is negative. For an extreme example of a white elephant consider the Túnel Las Raíces that spans the border between Argentina and Chile. Built in the 1940s and still the longest in Latin America, this tunnel has never been put to its intended use.

11. This is not limited to highway franchises. Contractual terms changed substantially, within three years, for more than half of the concessions awarded during the 1990s in Latin America (Guasch 2004).

12. In the 1990s, the U.K. was the first country to use a PVR-like franchise contract with a flexible franchise term, yet the franchise was not awarded in a competitive auction. Colombia auctioned a highway to the bidder demanding the least toll revenue in the mid-1990s, yet toll revenue was not discounted. The first PVR auction took place in Chile in 1998, when the US$400 million improvement and expansion of the Santiago-Valparaíso-Viña del Mar highway was auctioned. This was the first highway franchise in Chile that required no (explicit) government guarantee.

13. The discount rate should be a good estimate of the cost of funds faced by franchise holders and could be variable (such as the London Interbank Overnight Rate [LIBOR] plus some fixed-risk premium).

14. Profits are affected, since the franchise term determines maintenance and operational costs. Yet the PVR contract can be modified to incorporate maintenance costs (Engel, Fischer, and Galetovic 2003).

15. Discretion in toll-setting may be limited by fixing a lower and upper bound (in real terms) on possible tolls.

16. Of course, this does not mean that the traditional approach would have led to better results.

References

Auditoría General de la Nación. 2003. "Concesiones viales y DNV: Muestra de las principales auditorías sobre óranos de control y empresas adjudiacatarias del proceso de privatización. Síntesis y Conclusiones. 1993–2003." Buenos Aires, Argentina.

Demsetz, Harold. 1968. "Why Regulate Utilities?" *Journal of Law and Economics* 11: 55–66.

Engel, Eduardo, Ronald Fischer, and Alexander Galetovic. 1997. "Infrastructure Franchising and Government Guarantees." In Timothy Irwin, Michael Klein, Guillermo Perry, and Mateen Thobani, eds., *Dealing with Public Risks in Private Infrastructure*. Washington D.C.: The World Bank.

———. 2001. "Least Present Value of Revenue Auctions and Highway Franchising." *Journal of Political Economy* 109 (5): 993–1020.

———. 2003. "Privatizing Highways in Latin America: Fixing What Went Wrong." *Economia, The Journal of the Latin American and Caribbean Economics Association (LACEA)* 4 (1): 129–58.

———. 2005. "Soft Budgets and Highway Franchising." In R. Noll, ed., *Infrastructure and Resources in Latin America*. Chicago: University of Chicago Press.

Gómez-Ibáñez, José A. 2003. *Regulating Infrastructure*. Cambridge, Mass.: Harvard University Press.

Guasch, José L. 2004. *Granting and Renegotiating Infrastructure Concessions: Doing It Right*. Washington, D.C.: World Bank Institute.

Laffont, Jean-Jacques. 2004. *Regulation and Development*. Cambridge, Mass.: MIT Press.

Tirole, Jean. 1997. "Comment on a Proposal by Engel, Fischer, and Galetovic for Highway Auctions." In Spanish. *Estudios Públicos* 65: 201–14.

Williamson, Oliver. 1976. "Franchise Bidding for Natural Monopolies: In General and with Respect to CATV." *Bell Journal of Economics* 7 (1): 73–104.

Comment on "Infrastructure and Development" by Rémy Prud'homme

NEMAT T. SHAFIK

It is normally not the role of discussants to be advocates for the paper that they are reviewing, but I too would strongly recommend that you read Rémy Prud'homme's paper. It made me think differently about something that I think about every day, and that's unusual. So first, a strong endorsement for reading a very readable paper.

Let me start with one of the main themes of the paper: the neglect of infrastructure in the economics literature. I hadn't realized how recent the origins of the word "infrastructure" were in the English language. In Arabic, which is my original language, the word for "infrastructure" actually has a much more evocative meaning—"the basic foundation"—which I think is a better description, since infrastructure actually underpins all economic and social activity. The insight I gain from the paper, as to why infrastructure is so underrepresented in the early economic literature, is the fact that in our economic models we treat capital as undifferentiated, so the specificities of infrastructure are not captured. In that sense much of infrastructure economics doesn't lend itself to what economics is very good at, which is generalizing theory and making policy recommendations in the abstract rather than in the devilish detail that Professor Prud'homme has mentioned. And in the end, good infrastructure economics is fundamentally empirical—and in many ways is antithetical to where much of contemporary economics is going, which is deeply theoretical.

In that sense I am very excited about some new work we have launched here with the development economics vice presidency (DEC) and the research department to try and do much more serious, empirical, operational research linked to World Bank infrastructure projects, to try and better understand some of these economic relationships. And it seems only appropriate to do this work in an institution that has infrastructure at its core, since the Bank's very first loan to France was for infrastructure.

Let me turn to the question of the benefits of infrastructure, and here I think the Bank's own evidence on the benefits is in many ways consistent, but in some ways

Nemat T. Shafik served as Vice President for Infrastructure at the World Bank until October 2004, when she became Director General, Regional Programmes, in the Department for International Development of the U.K. government.

189

even more compelling than the average economic rates of return of 15 percent that Professor Prud'homme mentions. We recently looked at the ex post economic rates of return on the Bank's own history of infrastructure investments since 1964. Our Operations Evaluation Department independently assessed ex post economic rates of return and found that the rates have averaged 20 percent since 1964, and, interestingly enough, over the past three years have actually gone up to 35 percent. I think that the increase in recent years is driven by three factors: first, our own efforts to clean up the portfolio and to restructure or cancel projects that were performing poorly; second, greater attention to the quality of implementation throughout the Bank; but, third, the huge collapse in private investment in infrastructure in recent years—and the fact that in a sense we were cherrypicking. In recent years there was such a backlog of infrastructure investments that needed to be made that the economic rates of return have been quite high.

Professor Prud'homme's point on market enlargement is a very important insight that is underappreciated in popular perception. We all know that infrastructure services—water supply, energy, sanitation, and transport services—are essential to economic activity. But when you look at which infrastructure projects in the Bank's own portfolio have the highest economic rates of return, they are telecommunications and transport. Those are the sectors that enlarge markets. The lowest return is actually on water supply, despite the fact that water supply is the most popular sector because everybody thinks that water supply is essential, which it is, but water supply is more about economic welfare than about economic growth.

The benefit of urbanization is another area where popular perception may not be consistent with economic reality. Development thinking has traditionally emphasized the importance of investing in rural areas to address poverty and stem migration. Yet, the evidence is so compelling about the returns of getting people to cities, because their productivity is so enhanced.

Now let me quickly run through some points on the institutional issues of public versus private provision. I think the empirical model developed in the paper is quite interesting because it shows the kind of details that you need to look at before making general comments about the desirability of public versus private provision. And the empirical conclusion in the paper is quite consistent with the Bank's own thinking about assessing public versus private finance, a position that really is less about ideology and more about methodology and doing the hard empirical work to make those judgments.

I think the political economy issues associated with the perennial cost underestimates that we see in the infrastructure sectors and the overestimate of demand are really a familiar story. It's the conspiracy between the politicians who like to cut ribbons, and the engineers who like to build things. And that makes economists, who are driven by the principle that the world consists of scarce resources and competing ends, suspicious. The fear of white elephants is real.

There are some interesting alternatives to the cost escalation that is so common in infrastructure projects. First, the high economic rates of return in the Bank's portfolio and the relative absence of white elephants indicate that having independent

technical assessment helps. Second, the U.K.'s experience with its Private Finance Initiative and the fact that hundreds of public-private partnerships issued in this way have come in within cost—because there was no alternative, and on average the costs have been 15 percent lower than the alternative of public procurement—is an interesting option, which reduces the risks of cost escalation.

But private participation in general is not a panacea, because it is difficult to write contracts that are complete over the long life of infrastructure assets. The recent World Bank book on infrastructure by Luis Guasch (2004) demonstrated that the probability of renegotiation is high (on average 30 percent) in infrastructure sectors, but very high (55 percent) in sectors such as transport. The fact that the government is an active back-seat driver, even when infrastructure services have been privatized, further complicates matters. I recall one Brazilian infrastructure investor who said to me, "The government's attitude is if it moves, you should tax it; if it keeps moving, you need to regulate it; and if it stops moving, you should subsidize it."

We at the Bank have been trying to develop alternative models of service provision that include some of the properties of Professor Eduardo Engel's PVA approach. We call this output-based aid, or performance-based contracting where, in effect, we bid negative concessions under which providers identify the minimum subsidy they need to deliver a service. For example, we have financed connection subsidies in Nepal for telecommunications, in Cambodia with water, and in Uganda for electricity. These are mainly associated with expansion of a network rather than subsidies for ongoing operations. So in Cambodia the average price for a water connection for rural areas was $340. The government promised to pay $340 for every household that was connected, and the household then took over the costs of running the service. This was a way to control costs, increase efficiency, and bring in private provision, but have it focus on providing services to poor households, thus meeting universal service requirements in a much more efficient way.

On risk allocation, which Professor Prud'homme's paper rightly identifies as a key driver for efficiency, I think the Bank has a key role to play. The issue is both about risk allocation and mitigation, where instruments like the Bank guarantee can help. Policy and foreign exchange risks are probably the most difficult to manage. Governments don't like to admit that they might break their promises in the future. So providing guarantees is problematic, but necessary for investors. In the long run better, more credible institutions are the answer, but until then poor countries will have to pay the price of higher risk premiums. The Bank must focus on bringing these risks down—through better policy environments, and stronger regulatory institutions.

Thank you.

Reference

Guasch, Luis J. 2004. *Granting and Renegotiating Infrastructure Concessions: Doing It Right.* Washington, D.C.: World Bank.

Trade and Development

Trade Liberalization in a Globalizing World

RICCARDO FAINI

International economic integration has been on the rise since at least the mid-1980s. But the recent episode of globalization has a number of distinctive features that make it vulnerable to a turnaround in even one of its components. First and foremost, the nature of foreign direct investment (FDI), particularly toward developing and emerging markets, has changed. In the past, FDI was primarily directed to establish production facilities in foreign markets, with a view to catering to foreign consumers and circumventing restrictions to trade. More recently, however, investment decisions by international firms appear to be increasingly driven by the desire to cut production costs—by slicing the value-added chain among affiliates in different locations, as a function of relative factor prices. Hence trade costs play a very different role in this context: They no longer encourage multinational firms to invest abroad with a view toward gaining better access to foreign consumers. To the contrary, the impact is to increase the fragmentation costs of the value-added chain, thereby discouraging firms from investing abroad. In other words, trade and FDI have increasingly become complements. As a result, any step back in either dimension would reflect negatively on the other.

Similar considerations apply to the link between trade and migration. The two have typically been seen as substitutes. The creation of NAFTA was indeed hailed by then-President Salinas of Mexico as reflecting the desire of his country "to export goods not people." This relationship may still hold true, but increasingly less so. In particular, the growth in service trade, one of the most dynamic components in the expansion of international trade, very much depends on the ability to supply such services in loco through the firm's own personnel, and is therefore positively linked

to migration. Once again, restriction on one component of globalization (immigration, for example), carries negative implications for other aspects (such as trade in services) of the process.

The main findings of this paper can be summarized as follows. First, while trade and FDI policies have become steadily more liberal throughout the world, immigration policies in receiving countries have grown quite restrictive. Only for skilled workers has immigration policy in the primary receiving countries become more generous, raising fears among the sending countries of a brain drain. Second, at least during the 1990s, import liberalization fostered not only trade but also inward investment, confirming that trade and FDI toward developing countries have been, most recently, largely complements. Third, the presence of a skilled labor force is a relevant factor to attract FDI. Moreover, trade policies and the stock of FDI have a positive impact on the incentives to acquire education. This set of findings highlights the possibility of a low-equilibrium trap, where the lack of human capital discourages FDI and inadequate investment from abroad limits the domestic incentives to acquire education. Rich countries, by encouraging skilled immigration from relatively poor countries, definitely aggravate such a risk. Fourth, we find little evidence to support the contrary argument of a brain gain, where the possibility for skilled workers to migrate abroad raises the return to education and the investment in human capital at home.

International economic integration has been on the rise since at least the mid-1980s, and trade in goods and services has been a key component in the process. While world gross domestic product (GDP) rose on average by 3.8 percent between 1985 and 2000, world exports expanded at a substantially faster rate during the same period: 6.1 percent on an average annual basis. The growth in trade did not come at the expense of reduced factor mobility, as traditional trade theory would typically imply. Between 1985 and 2000, real foreign direct investment (FDI) increased at an average annual rate of 17.7 percent. The stock of inward FDI rose from 8.4 percent of world GDP in 1985 to 22.3 percent in 2002. Even migration, in many respects largely absent in the current globalization episode, played a non-negligible role in fostering international integration. In the United States, for example, the stock of foreign-born population increased from 6.2 percent of the total population in 1980 to 10.4 percent in 2000.

More recently, however, there have been signs that globalization is retreating. FDI fell markedly, by 41 percent in 2001 and again by 20 percent in 2002. International trade stagnated in 2001 and failed to rebound in 2002: The volume of merchandise trade actually fell in industrial countries during these years, so a key question is whether this is simply a cyclical phenomenon reflecting the global economic slowdown, or whether it carries more worrisome implications.

Clearly, it is too early to tell. What we can say, however, is that the current episode of globalization has a number of distinctive features that make it more vulnerable to a turnaround in even one of its components. First and foremost, the nature of foreign

direct investment, particularly toward developing and emerging markets, has changed. In the past FDI was mainly directed to establish production facilities in foreign markets, with a view toward catering to foreign consumers and circumventing restrictions to trade. Accordingly, an increase in trade barriers would have been associated with a rise in FDI. By and large, therefore, trade and FDI were substitutes. More recently, however, investment decisions by international firms seem increasingly to be driven by a different set of considerations. Improvements in communication and transportation technology allow firms to achieve substantial cost reductions by slicing the value-added chain among affiliates in different locations as a function of factor prices. Hence, trade costs play a very different role in this context; they no longer encourage multinational firms to invest abroad with a view toward gaining better access to foreign consumers. To the contrary, trade costs increase the costs attendant on the fragmentation of the value-added chain, thereby discouraging firms from investing abroad. In other words, trade and FDI have increasingly become complements. As a result, any step back in either dimension would reflect negatively on the other. In particular, increasing trade restrictions would not only depress trade but also have a negative effect on the incentives to undertake FDI.

Similar considerations apply to the link between trade and migration, which have typically been seen as substitutes. The creation of NAFTA was indeed hailed by then-President Salinas of Mexico as reflecting the desire of his country "to export goods not people." This relationship may still hold true, but increasingly less so. In particular, the growth in service trade, one of the most dynamic components in the expansion of international trade, very much depends on the ability to supply such services *in loco* through the firm's own personnel and is therefore positively linked to migration. So once again, restrictions on one component of globalization—immigration, for example—have negative implications for other components of the process (trade in services).

These considerations have substantive implications for both policy and research. At the academic level, they highlight the need to study globalization in a fully integrated way, not just as the sum of its components. Separate analyses of trade, FDI, and migration will not do the job. Even bivariate studies of the link between trade and migration, or between trade and FDI, may not tell the whole story. We know little, empirically or theoretically, about the links between FDI and migration. And we know even less about the intricate relationships that link trade, foreign direct investment, and migration.

From a policy point of view, the need for policy coherence is paramount. Restricting trade may be detrimental to a country's ability to attract FDI. Similarly, limiting migration may discourage FDI and depress trade. Unfortunately, policy coordination is weak, particularly in developing countries, with key decisions affecting trade, inward FDI, and migration policies often made by different public bodies. Policy coherence is also particularly inadequate at the international level. Coordination among international institutions is the exception rather than the rule. Also, there is no international institution in charge of migration issues. Similarly, attempts to establish multilateral rules for FDI have completely floundered and have all but been abandoned. Finally, the drive to coordinate aid, trade, and capital flows policies has so far yielded no tangible results.

The remainder of this paper is organized as follows. The following section briefly reviews the major trends in international economic integration since the early 1990s. The focus is on trade, FDI, and migration, with an eye toward policies. We find that while trade and FDI policies have steadily become more liberal throughout the world, immigration policies in receiving countries have grown quite restrictive, especially in Europe. Only for skilled workers has immigration policy in the primary receiving countries become more generous, a trend that has raised considerable concern among developing countries afraid of losing their best and most highly educated talents.

These trends have significant implications for the ability of developing and emerging markets to fully benefit from globalization. Trade liberalization across the world should foster trade, boost openness, and encourage FDI. We next test these conjectures, and as expected, find that, at least during the 1990s, import trade liberalization fostered both trade and inward investment. At the same time, we also find that the presence of a skilled labor force is a relevant factor in attracting FDI. We therefore take a close look at the determinants of investment in human capital in developing countries—in an effort to assess whether, in addition to more traditional determinants, trade policies and the stock of FDI also have a positive impact on the incentives to acquire education. We find that this is indeed the case, which suggests the possibility of a low-equilibrium trap where the lack of human capital discourages FDI, and that inadequate investment from abroad limits the domestic incentives to acquire education. Rich countries, by encouraging skilled immigration from relatively poor countries, definitely aggravate such a risk. Developing countries are therefore right to be concerned about the negative growth implications of the brain drain.

However, the case could be made that the brain drain is not necessarily a curse for sending countries.[1] It may well be that opening the borders of industrial countries to skilled migration from the developing world raises the return to education in sending countries, thereby boosting investment in human capital. Under these conditions trade liberalization, FDI, and the brain drain would no longer work at cross purposes but would be mutually reinforcing. Trade liberalization and the brain drain would both be associated with a larger flow of FDI. This is because the sheer ability to migrate abroad would boost the incentive to education, potentially raise the domestic supply of skilled workers, and in the end trigger an even larger flow of foreign direct investment. We assess the empirical plausibility of this argument and find evidence suggesting that this rosy scenario is somewhat unlikely. Overall, the results can only be seen as preliminary, but they clearly indicate the need to examine the effects of trade, FDI, and migration policies in a fully integrated manner.

Trade and Factor Mobility: Trends and Policies

The key facts of globalization have been well documented elsewhere and need only to be briefly recalled here. Between 1985 and 2000, world real GDP increased at an average annual rate of 3.8 percent. During the same period, real exports increased at
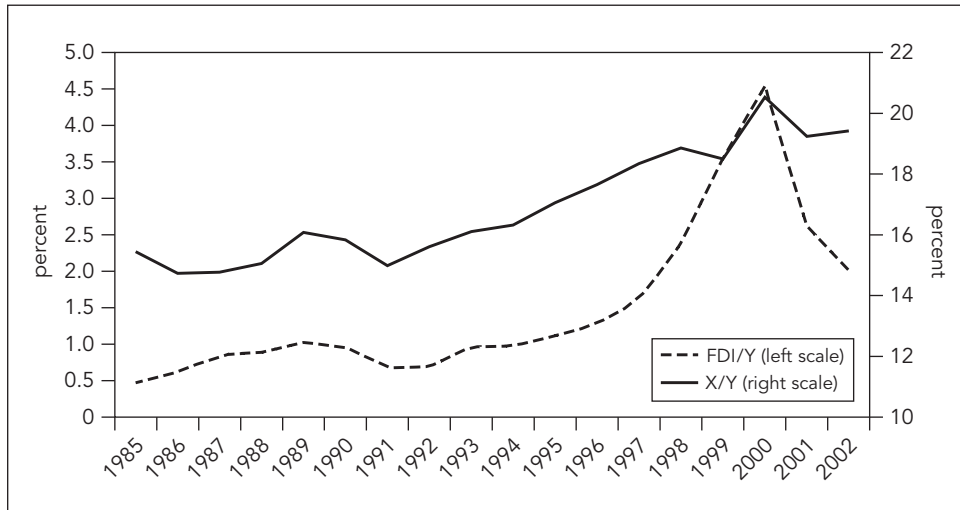
TABLE 1.
Stock of Foreign Population in OECD Countries
(percentage of total population)

Country	1990	2000
United States	7.9	10.4
Japan	0.9	1.3
France	6.3	5.6
Germany	8.4	8.9
Italy	1.4	2.4
United Kingdom	3.2	4.0
Belgium	9.1	8.4

Source: SOPEMI 2003.

Note: OECD = Organisation for Economic Co-operation and Development.

FIGURE 1.
**World Trade and World FDI as a Percentage of World GDP**

Source: Author's compilation using UNCTAD and WTO data.

Note: X = exports, Y = GDP.

an average rate of 6.1 percent and real FDI flows by 17.7 percent. As a result, the share of both exports and FDI in world GDP increased substantially (figure 1). Migration also expanded, but at a substantially slower pace. Cross-country comparisons of migration data are marred by definitional problems. Yet, available evidence shows that migrant stocks, as a percentage of population, have risen relatively slowly for the primary receiving countries and, in a number of cases, have even declined (table 1).

For purposes of this paper, the key fact is the expanding role of developing countries. Their share in world exports has increased rapidly, especially during the 1990s, largely driven by the exceptional performance of Asia (figure 2). Developing countries are also playing an increasingly relevant role as host of FDI. While foreign direct investment goes predominantly to developed countries, the share going to developing

FIGURE 2.
Shares of World Non-Oil Exports
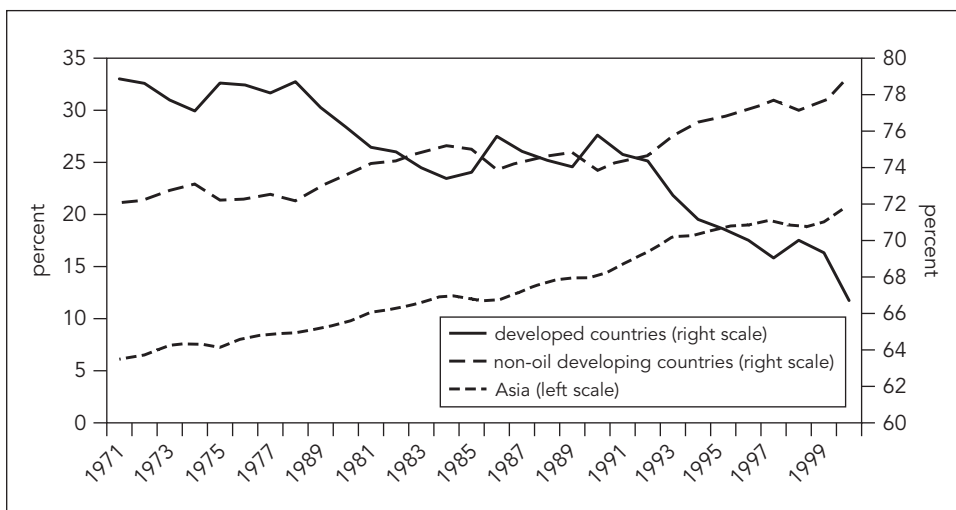
Source: Author's compilation using WTO data.

FIGURE 3.
Shares of World FDI
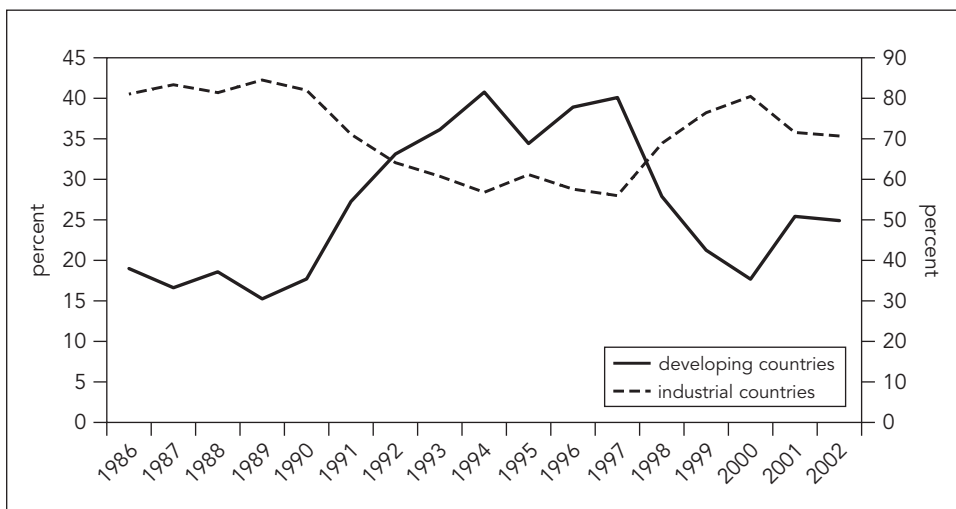
Source: Author's compilation using UNCTAD data.

countries has risen since the late 1980s. This share fell in the aftermath of the Asian crisis but has since recovered, albeit at a slow pace (figure 3). Perhaps more crucially, developing countries have witnessed a substantial rise in their exposure to the international economy. Figures 4 and 5 present the regional details for trade and FDI, respectively. Figure 4 shows how trade openness—defined as the sum of imports and exports over GDP—has risen steadily in Asia, developing Europe, and Latin America. The picture for Africa is a bit less univocal as is that of the Middle East, despite the

FIGURE 4.
Openness in Developing Countries, 1980–2000
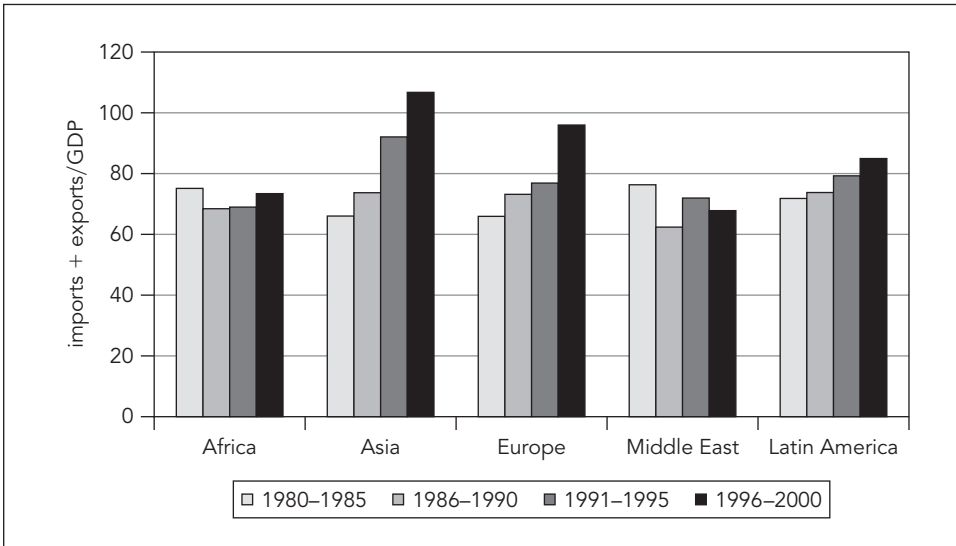
Source: Author's compilation using World Bank data.

FIGURE 5.
Inward FDI as a Percentage of GDP
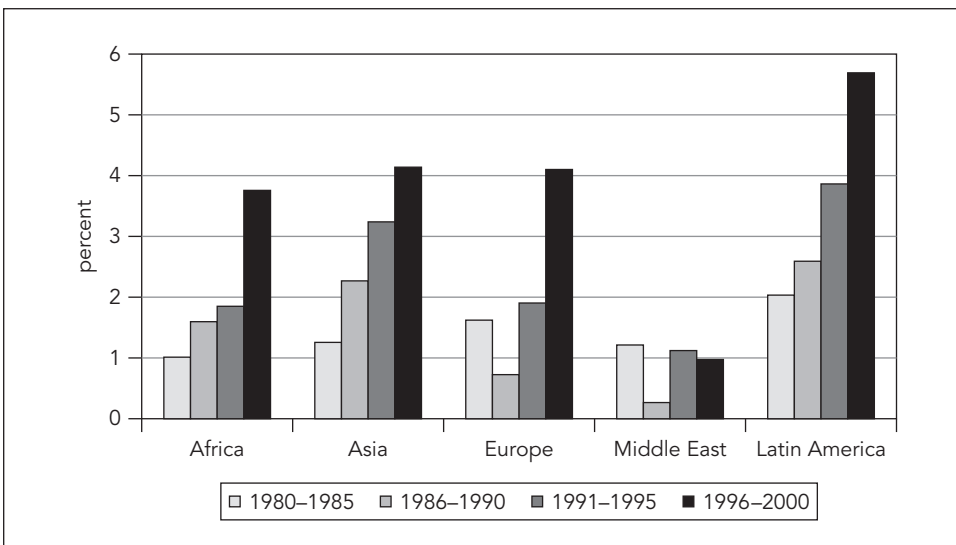
Source: Author's compilation using UNCTAD data.

fact that openness is measured at constant prices and so is not affected by gyrations in commodity prices. Foreign direct investment also shows a rising trend as a percentage of host-country GDP (figure 5). With the exception of the Middle East, inward foreign direct investment has surged in all regions and now accounts for about 4 percent of GDP in Africa, Asia, developing Europe, and Latin America. Contrary to widespread beliefs, Africa has also benefited from the rapid expansion in FDI; its

FIGURE 6.
Average Tariff Rates, 1980–2000
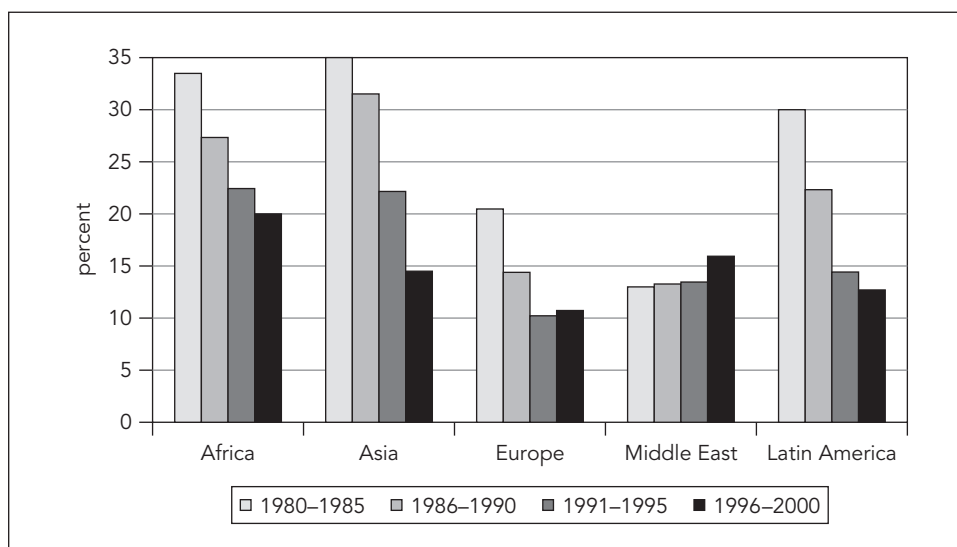
Source: Ng 2004.

low share in world FDI simply reflects its low share in world GDP. Finally, the substantial increase in FDI to Latin America is partly the mirror image of the large-scale privatization program initiated there during the 1990s.

A more liberal policy stance has been instrumental in opening up the economies of developing countries. Figure 6 illustrates how, over the past 20 years, tariff barriers have declined significantly in most developing regions, from 33 to 20 percent in Africa, from 35 to 15 percent in Asia, and from 30 to 13 percent in Latin America. Developing Europe, a relatively latecomer to the globalization process, also managed to halve its average tariff rate from 20 to 10 percent. The noticeable exception to this fairly general trend is the Middle East, where tariff barriers increased from 13 percent in the early 1980s to 16 percent in the late 1990s. Turning to nontariff barriers, these have also declined for most developing countries. Kee, Nicita, and Olarreaga (2004) have computed the tariff equivalent value of both core and non-core nontariff barriers for three selected years during the 1990s (figure 7). Perhaps the most striking finding is that, on average, nontariff barriers do not restrict trade as much as commonly believed: In all cases, their ad valorem equivalents are below 10 percent, and only for Latin America and the Middle East are they actually close to such value. For Africa, Asia, and developing Europe the rates are much lower levels, around 2 to 4 percent.

Summing up so far, two facts stand out. First, tariff barriers are still relatively high in many developing countries but have been on a downward trend since the early 1980s, with the only noticeable exception being the Middle East. Second, nontariff barriers do not restrict trade significantly, at least as measured by their tariff equivalent values.

FIGURE 7.
Ad Valorem Equivalents of Nontariff Barriers
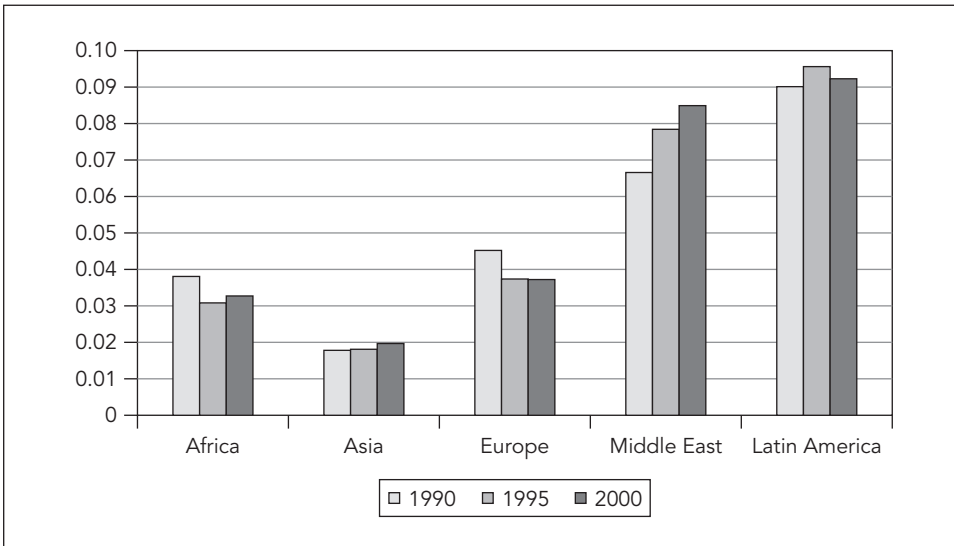
Source: Kee, Nicita, and Olarreaga 2004.

Turning to barriers with respect to FDI, the general picture of a more liberal regime still holds. Restrictive measures are more difficult to quantify in this area. Nonetheless, the United Nations Conference on Trade and Development (UNCTAD) has maintained a headcount of FDI measures and classifies them according to whether they represent a move toward a more or less liberal regime. The trend is definitely toward a more open policy regime, with liberalizing measures outranking restrictive ones by a factor of 10.

Last but not least, the stance on migration policy in most countries stands in sharp contrast with the increasingly liberal attitude toward trade and FDI. Indeed, since 1974, most industrial countries have tried to restrict new immigration and, at the same time, favor the return of earlier immigrants. While many of these policies have been only partly successful, they have nonetheless succeeded in slowing the flow of immigrants that characterized the early postwar period. More recently, in response to the growing shortage of skilled workers, most receiving countries have tried to shift the focus of their immigration policy, with a view to favoring the recruitment of highly skilled workers. This new twist in the immigration policy stance has become a source of considerable concern in countries that have traditionally sent immigrants abroad but are now afraid of losing their most skilled and entrepreneurial workers. Unfortunately, empirical evidence on the size and determinants of the brain drain is quite limited. The gap has been partly filled by the work of Carrington and Detragiache (1998). Using U. S. census data on the educational level and geographical origins of immigrants into the United States, and combining this information with the Barro-Lee data set on the level of educational achievements in sending countries, they have estimated migration rates for different educational groups. Their main

TABLE 2.
The Brain Drain: Migration Rates by Educational Attainment
(percentage of host country's educational group)

Origin country	Migration to the U.S.		Migration to the OECD	
	Secondary education	Tertiary education	Secondary education	Tertiary education
Colombia	3.6	5.6	3.8	5.6
Dominican Republic	29.7	14.2	30.5	14.7
Ghana	0.3	15.1	0.7	25.7
Guatemala	29.1	13.5	29.1	13.5
Korea	1.2	5.7	3.3	14.9
Mexico	20.9	10.3	20.9	10.3
Philippines	4.4	6.6	6.0	9.0
Uganda	0.6	15.4	0.6	15.5

Source: Carrington and Detragiache 1998.

finding is that skilled migration can represent a significant drain for at least some developing countries. Table 2 reports migration rates for secondary- and tertiary-educated migrants in a selected group of sending countries.

Trade and Factor Mobility: The Most Recent Trends

The rapid growth in trade and FDI came to a sudden stop in 2001, when the fall in FDI was a massive minus 41 percent in nominal terms. Fortunately, there was no collapse in world trade, but the volume of trade was virtually unchanged with respect to the previous year. Matters did not improve much in 2002, when FDI again fell, this time by 20 percent. The recovery in trade was modest, 3.2 percent, or basically in line with the growth in world output. The prospects for 2003 are not particularly bright, with growth in trade projected to stay, once again, below that of world output.

This slowdown in the pace of globalization can largely be attributed to a slowing of the global economy and, to a lesser extent, to the correction of the financial excesses of the late 1990s. Yet comparisons with earlier periods suggest that fluctuations of the world economy cannot fully explain what happened to trade and FDI from 2001 to 2003. For purposes of comparison, consider the 1991–93 slowdown, when world output growth stood at 3 percent (as in 2001–03). We find that world trade increased by only 3.9 percent in 1991 but then recovered rapidly, to 9.5 and 11.4 percent in the two following years. As we have just seen, there has been no such recovery of trade in the current slowdown. Similarly, FDI fell by 20 percent in 1991, but recovered rapidly in 1992—particularly so in 1993, when its rate of growth stood at around 30 percent. As argued by UNCTAD (2003), what is of concern today therefore is not only the severity of a downturn but also its duration.

These concerns should not be overblown. The recent fall in world FDI reflects the collapse in mergers and acquisitions, which is in turn linked to the decline in the

stock markets. Moreover, the data for 2003 seem to indicate a bottoming out of the FDI cycle. Yet the fact remains that trade, a component not too prone to financial excess, fell relatively more, and was relatively slower to recover compared with previous slowdowns.

The policy stance is an additional source of concern. The failure of the 2003 WTO ministerial meetings in Cancun and the collapse of the negotiations for a multilateral agreement on investment provide a hefty reminder that the trend toward more open policies should not be taken for granted. We also know from history that globalization is not an irreversible process driven by the forces of technology. Policy does matter, even more so in a context where, as noted earlier, complementarities among different facets of globalization play an increasingly relevant role. Failure to liberalize trade, or worse, a retreat from current trade policies, would not only depress the expansion of trade flows but could also possibly undermine the incentives for capital and labor mobility.

Trade Liberalization, Trade Flows, and Factor Mobility

There is considerable disagreement in the literature as to the growth effects of a more liberal trade regime. A key difficulty is whether the stance of the trade regime should be measured by an output indicator, such as the relative size of trade flows, or more directly by an index of trade policy. Output indicators are easier to compute but do not necessarily provide a good measure of the trade policy stance, as they reflect the influence of many confounding factors. Indexes are harder to determine and typically perform less well in empirical analyses. Still, until quite recently, the conventional wisdom was that trade openness, whether measured by an output or a policy indicator, was positively associated with per capita income growth. However, an influential paper by Rodrik and Rodriguez (2000) showed that standard wisdom was, as often happens, less conclusive than commonly believed. Not only were traditional measures of trade openness imperfect indicators of the actual stance of trade policy; more crucially, trade restrictions, compared with other unsound policies, were found to play a relatively minor role in determining the growth performance across countries. More recently though, Wacziarg and Welch (2003), while confirming most of the early findings of Rodrik and Rodriguez, showed that, in a time-series context, properly identified trade liberalization episodes have a positive and robust effect on growth and investment.

Trade Policy and Trade Flows

For purposes of this paper, the key finding of Wacziarg and Welch (2003) is that trade policy measures are indeed associated with greater trade openness, measured by the sum of exports and imports over GDP. This result supports the view that, typically, the effects of trade liberalization are not negated by offsetting measures or by poor implementation. We assess the robustness of the findings by Wacziarg and

Welch (2003) by adding to their regression a number of structural factors, namely, per capita GDP and total population, with a view to capturing the impact of economic development and size on openness. Trade policy is simply measured by an indicator of tariff barriers compiled by the World Bank (2003). We ran the following simple five-year regression over the 1980s and 1990s (1981–85, 1986–90, 1991–95, 1996–2000) for a large sample (92) of developing countries:

$$\frac{X + M}{Y} = \alpha + \beta \ln Y_{pc} + \gamma \ln Pop + \delta\tau \tag{1}$$

where X, M, and Y denote exports, imports, and GDP; while Y_{pc}, Pop, and τ represent per capita income, total population, and tariff barriers. Details about our estimation methods and data sources are explained in the annex at the end of this paper. Suffice it to say here that, in an effort to cope with endogeneity problems, explanatory variables are equal to their value at the beginning of each five-year period, while the dependent variable is averaged over the entire period. Also, as an indicator of τ we use both the level of import duties and its log. As expected, we find that openness is positively associated with income per capita and negatively related to population (table 3), which confirms the view that rich and small economies tend to trade relatively more. Furthermore, trade policy also affects significantly, in both statistical and quantitative terms, the level of openness. A 10 percent decline in the level of tariff is associated with a 2 percent increase in the level of openness.

Summing up, a more liberal trade stance will positively affect openness, even after controlling for the more structural determinants of trade flows.[2]

Trade Policy and Foreign Direct Investment

We can now turn to the relationship between trade policy and inward foreign direct investment. Unfortunately, theory does not provide a clear-cut answer as to the

TABLE 3.
Openness and Trade Policy

Dep. Variable: $[(X+M)/Y]_{t,t+4}$

	(1)	(2)
ln $(Y_{pc})_t$	13.1 (4.3)	9.30 (2.2)
ln $(pop)_t$	−12.4 (6.5)	−12.4 (6.8)
τ_t	−0.20 (2.01)	
ln τ_t		−12.7 (4.5)
$\chi^2(7)$	654.5	780.9
R^2	0.47	0.56
No. of observations	206	194

Source: Author's compilation.

Note: X = exports, M = imports, Y = GDP, Y_{pc}= per capita income, pop = population, τ = tariff rate.

direction of such a link. Consider first the standard model of trade. An increase in tariff barriers will typically depress both exports and imports. In a capital poor country, it will also raise the returns to capital and hence attract investment from abroad. This is the standard substitutability result between trade and factor mobility. Conversely, the lifting of trade restrictions should boost trade and discourage factor mobility. Interestingly enough, even less orthodox models come to similar conclusions. Consider, for example, the product cycle paradigm. Initially the new product will be exported by the innovating firm. However, once production becomes routine, it will be moved to a foreign location, where production costs are lower. Therefore, here too FDI is viewed as replacing trade.

The models above basically fit the description of horizontal FDI, where the shift of production to a foreign location is motivated by the desire to circumvent trade barriers, save on trade costs, and gain access to foreign consumers. Even in this setup, however, the medium-term link between trade and FDI is not unambiguous. As noticed by UNCTAD (1996), for instance, foreign affiliates typically generate a steady demand for imports of capital and intermediate goods from their parent firm. In the medium run, therefore, following the investment abroad, trade flows may grow rather than contract. However, it is still true that trade will grow less compared to the case where the firm had not established a production platform abroad. Accordingly, the prediction that trade barriers will depress trade and encourage market-seeking FDI still holds.

Market-seeking inward FDI is certainly not the rule for most developing countries, given the small size of their markets. A perhaps more relevant motivation of inward FDI has been the desire to exploit the availability of natural resources in host countries. Resource-seeking FDI is typically trade creating. Multinational firms will mostly cater to consumers in their home country or in third markets. Indeed, the main motivation of resource-seeking FDI is to produce for exports rather than for host country consumption. The key observation here is that trade barriers are unlikely to matter much. Compared to manufacturing firms, primary sector affiliates are less dependent on the imports of intermediate goods. They are also unlikely to face major barriers in export markets. At any rate, trade barriers in the host country are unlikely to be a determining factor in the location decision of resource-seeking FDI. Even imports of capital goods are typically taxed at favorable rates if not totally exempted. Accordingly, the prediction is that trade barriers should not have a substantive effect on resource-seeking FDI.

It is also worth noting that this form of FDI has been steadily losing relevance. For example, exports by U.S. primary sector affiliates as a percentage of host countries' exports of primary goods declined from 35 percent in 1977 to 11.2 percent in the early 1990s (UNCTAD 1996). This trend mostly reflects the policies of indigenization of primary sectors by host countries. Moreover, the primary sector itself has lost much ground as a source of foreign exchange for the developing world. Indeed, in the early 1990s, the primary sector accounted for more than 22 percent of the stock of inward FDI in developing countries, but during that decade only less than 9 percent of the total flow of inward FDI went to that sector.

Efficiency-seeking, or vertical, FDI represents a further motivation for the firm decision to locate in foreign country. The aim is simple, namely, to cut production costs by slicing the value-added chain and relocating abroad the production of those intermediate goods that are too costly to produce domestically. The implications for trade are immediate: Efficiency-seeking FDI is trade creating. Transactions within the firm or among firms in the home country are replaced by trade between the parent firm and its affiliates. Quite often, moreover, affiliates will start selling to firms other than their parent. Similarly, the role of trade barriers changes radically. Pervasive restrictions to trade in the host country would raise the costs of vertical disintegration between the parent firms and its affiliates abroad. Hence, trade barriers discourage vertical FDI.

To sum up, we have distinguished three types of motivations for FDI, with altogether different implications for both trade and for the impact of trade barriers. First, horizontal FDI substitutes for trade and is generally fostered by high trade barriers in host countries. Second, resource-seeking FDI augments trade but is relatively insensitive to trade barriers in host countries. Finally, vertical FDI is also trade creating but is highly responsive to trade barriers.[3]

These are all testable implications. They require, however, the ability to distinguish between these three forms of FDI. Unfortunately, this is a tall task, one reason being that most often aggregate data include all three types of investment. Moreover, even for firm-level data the different kinds of motivations can all be present at the same time. Nonetheless, there are indications of a negative relationship between trade costs and vertical FDI once data series are extended to cover the 1990s (Barba Navaretti and Venables 2004). For example, Hanson, Mataloni, and Slaughter (2001) find a negative relationship between the host country's trade costs and both the ratio of affiliate exports to affiliate sales to the local market and the ratio of affiliate imports from parents to total affiliate sales.[4]

In what follows, we take a very simple route, starting from the observation that resource-seeking FDI has been on a somewhat downward trend, at least in relative terms. We also note that vertical FDI has been gaining in importance, even with respect to market-seeking FDI. Indirect evidence in this respect comes from the fact the average export propensity of U.S.-majority-owned manufacturing affiliates in developing countries has been steadily rising, from 8.4 percent in the mid-1960s to 22 percent in the early 1980s, and to 39 percent in the mid-1990s. The biggest rises have been in Latin America and developing Asia. In Asia's newly industrializing economies, the export propensity of U.S. affiliates has actually declined, albeit from very high levels, most likely reflecting the expanding size of the host countries' domestic markets. Again, we see that a sharp distinction between horizontal and vertical FDI is hard to draw—although, together, these data point to the growing role of vertical FDI. If so, then, we would expect trade barriers to increasingly discourage FDI to developing countries, particularly in the 1990s.

We test this proposition in a simple manner, first by taking the UNCTAD measure of FDI attractiveness. This is basically a simple average of the scores achieved by different countries on a number of indicators that are deemed to attract FDI.[5] The key

observation here is that the UNCTAD FDI potential index does not include trade barriers. We therefore ran the following simple regression:

$$\frac{FDI}{Y} = \alpha + \beta PI + \gamma\tau \tag{2}$$

where Y is GDP, PI is the UNCTAD potential index of inward FDI, and τ is a measure of trade barriers. Note that a high level of PI indicates that the country is not relatively attractive as a destination of FDI. We expect therefore $\beta < 0$ (countries with high PI are unattractive to foreign investors) and $\gamma < 0$ (if vertical FDI is predominant). Here, too, details about estimation methods and data sources are relegated to the annex. Suffice it to say that the sample includes only developing countries. Industrial countries are excluded on the ground that inward FDI in these countries is still motivated by market-seeking considerations and, accordingly, is encouraged by high trade barriers (Barba Navaretti and others 2002).[6] Furthermore, we focus only on the 1990s, where efficiency-seeking FDI has played an increasingly relevant role.[7] Overall, therefore, given that we exclude industrial countries (where horizontal FDI is still very relevant) and focus only on the 1990s (when vertical FDI became more relevant) our expectation is to find that trade barriers discourage FDI (i.e., that $\gamma < 0$). As a measure of τ we take the sum of tariff barriers and the ad valorem equivalent of nontariff barriers.

The results are strikingly simple (table 4). First, as expected, the index of potential FDI is negatively associated with the GDP share of inward FDI (recall that a high level of PI means that the country is not very attractive as a host to FDI). Second, trade barriers significantly discourage inward FDI. The results are robust to the inclusion of regional fixed effects. They suggest that vertical FDI has indeed been the

TABLE 4.
FDI and Trade Policy

Dep. Var: $[FDI/Y]_{t,t+4}$

	(1)	(2)
PI_{FDI}	−1.04 (1.80)	−1.58 (2.28)
ln τ	−1.23 (2.5)	−1.03 (1.95)
Wald test $\chi^2(2)$	20.4	27.5
Hausman test $[\chi^2(4)]$		4.34
Estimation method	FE	RE
$\chi^2(4)$ for regional dummies		6.16
R^2	0.35	0.40
No. of observations	61	61
Regional dummies	no	yes

Source: Author's compilation.

Note: FDI = foreign direct investment, Y = GDP, PI_{FDI} = potential for FDI, τ = trade barriers, FE = fixed effects estimator, RE = random effect estimator.

predominant mode of investment internationalization toward developing countries during the 1990s.

We have also tried to extend our analysis to earlier periods. Unfortunately, the UNCTAD index is not available before the 1990s. We therefore had to run a more complex regression, controlling for (some of) the primary factors that are thought to affect the attractiveness of FDI. Moreover, we are less optimistic in our quest to find a significant impact of trade barriers on FDI given that for earlier periods the three motivations—resource, market, and efficiency seeking—were all even more present, therefore confounding the impact of trade restrictions. The new regression is:

$$\frac{FDI}{Y} = \alpha_0 + \alpha_1 \, K_{FDI} + \alpha_2 \, HK + \alpha_3 \, PK + \alpha_4 \, \ln(Y_{pc}) + \alpha_5 \tau \qquad (3)$$

where K_{FDI} denotes the ratio of inward FDI stock to GDP, HK and PK are two measures of the stock of human and physical capital respectively, Y_{pc} is income per capita, and τ denotes (as usual) the size of trade barriers. As a measure of the stock of human capital, we use the average number of school years per inhabitant, as recently updated by Barro and Lee (2000). As a proxy for physical capital, we take the number of telephone lines per 1,000 inhabitants. Finally, for trade barriers we only have indicators of tariff restrictions. The ad valorem equivalents of nontariff barriers that were used in table 4 are available for only the 1990s, not for earlier periods.

Three facts stand out (table 5). First, and perhaps not unexpectedly, trade barriers have no clear impact on FDI decisions when a longer period is considered,

TABLE 5.
FDI, Human Capital, and Trade Policy

Dep. Var: [FDI/Y]$_{t,t+4}$

	(1)	(2)
K_{FDI}/Y	0.05 (1.94)	0.07 (3.06)
HK	0.94 (2.26)	0.87 (2.2)
PK	0.01 (2.1)	0.01 (1.33)
ln (Y_{pc})	−1.01 (0.71)	−0.5 (0.4)
τ	0.01 (0.5)	
ln τ		0.05 (0.1)
F(5,70)	6.13	7.4
Hausman test	9.3 [χ^2(4)]	9.6 [χ^2(5)]
Estimation method	FE	FE
R^2	0.17	0.40
No. of observations	140	130
Regional dummies	No	No

Source: Author's compilation.

Note: K_{FDI} = stock of foreign direct investment, Y = GDP, τ = tariff rate, HK = average years of schooling (human capital), PK = telephone lines per 1000 inhabitants (physical capital), Y_{pc} = per capita income, FE = fixed effects, RE = random effects.

presumably reflecting the wider variety of motivations of multinational firms. Second, the availability of both physical and human capital encourages inward FDI, confirming that multinational firms are attracted by the presence of a skilled labor force and of adequate infrastructures. The role of human capital is particularly notable, as it confirms the conjecture in much of the literature (Blomstrom and Kokko 2003, Borensztein and others 1998) that it is not simply the availability of low-wage unskilled labor that attracts foreign investors.

By and large, therefore, our results so far show that FDI toward developing countries has become increasingly motivated by efficiency considerations. Both progress in communication technology and lower trade barriers have contributed to the international fragmentation of the value-added chain along efficiency lines. Furthermore, we find that the availability of a pool of educated workers is also a key consideration in the location decision of multinational firms in developing countries. This latter finding raises an intriguing possibility. Consider the case where investment in human capital is encouraged by the presence of foreign investors. We may then well have two equilibria, one with a limited presence of foreign firms and poor educational achievements, the other with a large stock of foreign capital and strong educational levels. We explore this possibility in the following section. But we first turn to the relationship between trade and migration.

Trade Policy and Migration

Trade policy can also affect the mobility of people and workers. For example, restrictive trade measures in industrial countries will discourage exports from developing countries and strengthen the push factors of migration. They will also encourage the expansion of low-skill intensive import substituting sectors in receiving countries, thereby reinforcing the pull factors of immigration. On both counts, therefore, migration pressure will increase. These effects are evident for instance in the Common Agricultural Policy of the European Union (EU). By discouraging the expansion of agricultural exports from countries in northern Africa, restrictive trade measures foster out-migration. Conversely, the disproportionate size of the agricultural sector in the EU acts as a magnet for would-be migrants.

Some further evidence in this respect comes from the sectoral allocation of migrants in receiving countries. Faini and Venturini (1993) show that by and large, immigrants are more likely to be employed in import-competing sectors.

Trade restrictions in sending countries are also likely to encourage out-migration, but there is limited evidence in this respect. Faini, Grether, and de Melo (1999) show, with the help of a simple simulation model, that in a Ricardo Viner framework trade liberalization will have fairly complex effects on out-migration. In particular, if exports respond slowly to the new trade regime, then trade liberalization will be accompanied by a relatively strong real exchange rate depreciation that will foster out-migration. Historical evidence by Collins, O'Rourke, and Williamson (1999) is also compatible with the notion that trade and migration may have been complements, at least during the 19th century.

The plausibility of a complementary relationship between migration and trade is also strengthened by the growing role of service trade—for many services must be delivered personally. Moreover, even tradable services often require skilled (or at least trained) personnel to be relocated, most likely temporarily, to the importing country.

Overall, the relationship between trade and migration is not unambiguous. While substitutability may have been the rule rather than the exception for most of the postwar period, the new forms of international exchange of goods and services increasingly imply a complementary relationship.

Foreign Investment, Investment in Human Capital, and the Brain Drain

We have seen in the previous section how foreign investment is typically attracted by an adequate supply of skilled workers. This raises an intriguing question: How are the incentives for human capital accumulation affected by the presence of foreign firms? Existing evidence shows that, by and large, foreign firms employ relatively more skilled labor than their domestic counterparts (Barba Navaretti and Venables 2004). Similarly, Feenstra and Hanson (1996a, 1996b) showed in a fairly influential set of papers, that foreign direct investment may raise the relative demand for skills in both the home and the host country.[8] We would then expect the incentive to acquire further education to be strengthened by a sufficiently large presence of foreign firms.

Does FDI Boost the Incentive to Invest in Human Capital?

For the purpose of illustration, consider a simple model where foreign capital accumulation (I_{PK}) is negatively related to its own stock (PK) but positively related to the stock of human capital (HK). The previous section presented some supporting evidence for both of these conjectures.[9] Similarly, we assume that investment in human capital (I_{HK}) is negatively affected by its own stock but is encouraged by the presence of foreign investment. Formally:

$$I_{PK} = f(PK, HK) \qquad \text{with } f_{PK} < 0 \quad \text{and} \quad f_{HK} > 0 \qquad (4)$$

and

$$I_{HK} = g(PK, HK) \qquad \text{with } g_{PK} > 0 \quad \text{and} \quad g_{HK} < 0 \qquad (5)$$

This admittedly simple model is amenable to a graphical representation (figure 8). In a steady-state equilibrium—we assume that the standard conditions for such an equilibrium to exist are fulfilled—we have that $I_{HK} = I_{PK} = 0$. Given our assumptions ($f_{HK} > 0$ and $g_{PK} > 0$), both schedules are positively sloped. For instance, an increase in PK will discourage investing in physical capital and will need to be offset by a rise in HK.

The fact that both schedules $I_{HK} = 0$ and $I_{PK} = 0$ are positively sloped raises the possibility of multiple equilibria. In figure 8 we depict three. The middle equilibrium is unstable; the other two are stable. We see therefore the emergence of a low-level

FIGURE 8.
Multiple Equilibria with Foreign and Human Capital Investment
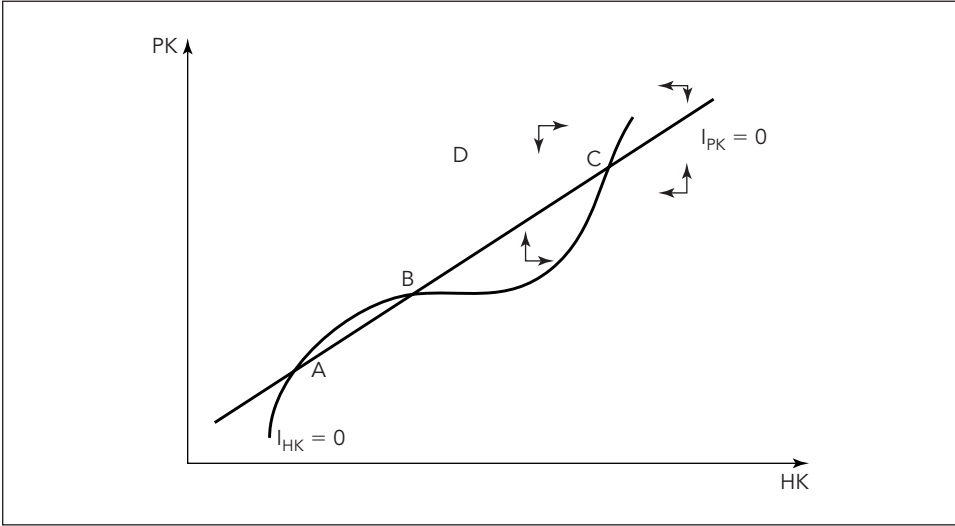
trap (point A in the figure), where foreign firms are discouraged from investing in the host country because of the inadequate supply of human capital there, while the limited presence of foreign firms reduces the demand for skills and discourages investment in education.

Is this an empirically plausible scenario? For an affirmative answer, we still must show that investment in education is encouraged by the presence of foreign firms. To ascertain this possibility, we ran a simple econometric equation with the goal of explaining the level of educational enrollment at secondary and tertiary levels. We take this variable as a reasonable proxy of investment in human capital, I_{HK}.

We then postulate that investment in human capital is a function of: a) per capita income, with a positive relationship provided that education is, as it should be, a normal good; b) trade policy, with a priori ambiguous impact on I_{HK}. In the standard model of trade with two factors (skilled and unskilled labor), trade liberalization in an unskilled abundant country should penalize skilled labor and hence discourage investing in human capital. However, the analyses of Feenstra and Hanson (1996a, 1996b) suggest that this may not always be the case, with trade liberalization instead boosting the returns to skills and the incentives to education in both skilled abundant and skilled scarce countries;[10] c) the endowment of natural resources, to allow for the possibility that the abundance of natural resources may discourage investing in human capital; and d) the presence of foreign firms, as measured by the stock of foreign capital. Formally:

$$I^i_{HK} = \alpha_0 + \alpha_1 K_{FDI} + \alpha_2 \ln(Y_{pc}) + \alpha_3 \tau + \alpha_4 D_{NR} \quad i = S,T \qquad (6)$$

where I^i_{HK} denotes investment in human capital at the secondary or tertiary level (i = secondary, tertiary), K_{FDI} is the stock of FDI over the host country GDP, Y_{pc} is

TABLE 6.
The Determinants of Secondary School Enrollment

Dep. Var: $[SSE]_t$

	Full sample	Africa	Asia	Europe	Latin America	Middle East
K_{FDI}/Y	0.34 (5.5)	0.22 (1.3)	0.75 (5.3)	0.32 (2.4)	0.3 (2.8)	−0.3 (0.7)
$\ln(Y_{pc})$	17.7 (3.2)	7.9 (1.2)	5.8 (1.0)	42.5 (4.0)	11 (1.7)	11.5 (0.7)
$\ln \tau$	−6.24 (5.4)	−4.9 (1.90)	−4.9 (1.57)	−3.4 (1.0)	−9.5 (5.0)	−1.5 (0.3)
D_{NR}	−3.0 (2.2)	−7.3 (3.1)	−5.3 (2.1)	−7.2 (1.2)	1.7 (0.7)	0.5 (0.1)
F(m,n)	41 (4,286)	5.3 (4,60)	23 (4,62)	10 (4,33)	14 (4,101)	0.7 (4,14)
Hausman test	19 $[\chi^2(4)]$					
Estim. method	FE	FE	FE	FE	FE	FE
R^2	0.51	0.72	0.14	0.29	0.31	0.22
No. of observations	378	91	80	53	127	27
Regional dummies	No	No	No	No	No	No

Source: Author's compilation.

Note: SSE = secondary school enrolment, K_{FDI} = stock of foreign direct investment, Y = GDP, τ = tariff rate, D_{NR} = dummy for natural resource abundant country, Y_{pc} = per capita income, FE = fixed effects, RE = random effects.

income per capita, τ is the level of tariffs, and D_{NR} is a dummy variable that takes a value of one if natural resource exports account for more than 40 percent of total exports.

The results for secondary school enrollment are presented in table 6. As expected, we find that income per capita is positively associated with secondary school enrollment, while a large endowment of natural resources has a negative impact on I_{HK}^S. Interestingly enough, both trade liberalization and the presence of foreign firms raise investment in human capital at the secondary school level, thereby supporting the notion that the accumulation of human capital responds favorably to a more liberal trade and foreign investment regime. This is a key result for the purpose of our analysis. It shows that foreign capital has a significant and positive impact on investment in education, as we indeed conjectured in equation 5 and in figure 8. Taken together with our earlier finding that a well-educated workforce is a crucial factor to attract foreign investors, it highlights the existence of a complementary relationship between FDI and human capital. By and large, therefore, the existence of a low-level trap, where the lack of foreign capital and of an educated workforce feed on each other, is something more than a mere theoretical possibility. In addition, our results are consistent with the findings of Feenstra and Hanson (1996a, 1996b) that FDI raises the returns to education.

We also ran the same equation separately for the five main geographical regions Africa, Asia, developing Europe, Latin America, and the Middle East. Interestingly enough, the regional estimates suggest that the previous results also hold for all the main regions, except for the Middle East. The predominance of resource-seeking FDI in such a region may explain why neither the stock of foreign direct investment nor trade liberalization has a positive effect on the incentives to accumulate human capital.

TABLE 7.
Determinants of Tertiary School Enrollment

Dep. Var: $[TSE]_t$

	Full sample	Africa	Asia	Europe	Latin America	Middle East
K_{FDI}/Y	0.12 (3.4)	0.12 (3.8)	−0.15 (1.1)	0.44 (5.3)	0.11 (3.1)	−0.7 (1.84)
$\ln(Y_{pc})$	14.6 (8.3)	5.4 (4.5)	23 (4.2)	41 (6.5)	7.1 (2.9)	6.3 (0.5)
$\ln \tau$	0.2 (0.3)	−1.4 (3.1)	2.2 (1.1)	1.5 (0.8)	−2.1 (2.8)	−1.4 (0.4)
D_{NR}	−1.6 (2.1)	−1.3 (2.9)	−0.9 (0.5)	2.3 (0.6)	1.1 (1.2)	−12 (3.5)
$F(m,n)$	29 (4,313)	17 (4,75)	9.1 (4,65)	30 (4,34)	12 (4,109)	4.5 (4,15)
Hausman test	81 $[\chi^2(5)]$					
Estim. Method	FE	FE	FE	FE	FE	FE
R^2	0.38	0.69	0.57	0.15	0.17	0.33
No. of observations	408	106	83	55	136	28
Regional dummies	No	No	No	No	No	No

Source: Author's compilation.

Note: TSE = tertiary school enrolment, K_{FDI} = stock of foreign direct investment, Y = GDP, τ = tariff rate, D_{NR} = dummy for natural resource abundant country, Y_{pc} = per capita income, FE = fixed effects, RE = random effects.

We now turn to tertiary school enrollment. The econometric results presented in table 7 are quite similar to those for secondary school enrolment, except for the fact that now trade policy is no longer a significant influence on educational investment for the pooled specification. Running the equation separately for the five main developing regions confirms the finding that the Middle East is an outlier, with the stock of FDI having a *negative* impact on tertiary enrollment. Trade liberalization raises investment in education at the tertiary level for Latin America and Africa, but not for the other areas.

Overall the results in this section suggest that trade liberalization has a twofold effect on investment in human capital. First, the lifting of trade restrictions attracts foreign firms and, accordingly, raises the demand and the returns to skills. Second, trade liberalization itself has a positive impact on the incentive to invest in human capital, at least for secondary school. As depicted in figure 8, trade liberalization would shift the $I_{HK} = 0$ schedule to the right, with two main implications. First, this would raise the steady-state levels of both human and foreign capital. Second, the rightward shift (if sufficiently large) of the $I_{HK} = 0$ schedule may eliminate the low-level equilibrium altogether. In both cases the economic and welfare effects are likely to be positive.

Perhaps more crucially, our results also show that education and FDI policies are complements. Opening up the economy to foreign capital increases the incentive to invest in education, which in turn further strengthens the attractiveness of the host economy to FDI. Conversely, any strengthening in the quality of educational facilities would have a positive impact on the steady-state levels of both human and foreign capital.

The Brain Drain: A Curse or a Blessing?

We have so far focused on trade and foreign direct investment, neglecting labor mobility. However, the previous setup lends itself quite easily to an analysis of the effects of skilled migration. To begin with, suppose that, prompted by the unexpected opening of the borders of industrial countries to highly educated immigration, the sending country suffers from a sudden loss in its skilled labor force. Let the initial equilibrium be at C, where both the human and the foreign capital stock are relatively large. Neither the $I_{HK} = 0$ nor the $I_{PK} = 0$ schedules would shift. The new (temporary) equilibrium would then be at a point like D, but the equilibrium would slowly move back to its original position. There is one case, though, where this may not happen. In particular, if the initial loss of skilled workers is large enough, then the economy may shift from the "good" equilibrium to the low-level trap in A.

Even abstracting from such a dramatic outcome, the effects of the brain drain are unlikely to be positive. Consider, for example, the case where we model the brain drain in a somewhat different way, as leading to a steady loss of skilled workers that migrate abroad. In a sense, the investment in human capital has become less productive because of higher depreciation (emigration) of the existing stock of human capital. Formally this amounts to making the $I_{HK} = 0$ steeper, with a two-fold effect. First, the steady state levels of both human and physical capital will decline. Second, the probability that the economy remains trapped in the low-level equilibrium at A will increase. Independently of how we model the impact of brain drain, its impact on the sending country's economic welfare is likely to be negative.

Our assessment of the impact of the brain drain may, however, be too gloomy. A recent stream of literature has argued that in the end the brain drain may not be as bad as previously thought for the sending country. The reason is simple. So far we have treated the brain drain as a purely exogenous phenomenon with no impact on the behavior of domestic agents. This approach may, however, be too restrictive. Suppose that the returns to skills are higher abroad, perhaps because of better technology. Consider the case where the foreign country opens up its border to skilled immigration, so that home-country educated residents now have a positive probability $(p > 0)$ to move abroad. Clearly, the expected return to education will rise, as it will now be equal to a weighted average of the higher return abroad and the initial return at home and, as a result, investment in education will also increase. If the latter effect is large enough, it may offset the loss of skills due the brain drain and allow the country to retain a larger pool of educated workers compared to the no migration case. In such a circumstance, the brain drain becomes a brain gain (Stark and others 1997, 1998; Mountford 1997).

Is this outcome empirically plausible? The evidence on the education boosting effect of the brain drain is quite limited, also because of the paucity of data on the size of skilled migration. While the gap has been partly filled by Carrington and Detragiache (1998), their work focuses almost exclusively on the United States and makes generalizations to other receiving countries tenuous. Recent work by Adams (2003) provides only a few additional observations.

Notwithstanding these limitations, there have been a number of attempts to assess whether the possibility of migrating abroad effectively boosts the investment in education. The evidence is mixed, with both Beine and others (2003) and Faini (2004) finding positive but relatively weak effects of the brain on secondary enrolment. Faini (2004) finds that the probability of tertiary migration actually depresses tertiary enrolment, a finding that is attributed to the choice by would-be migrants to pursue their graduate studies abroad.

In what follows we take a further shot at the issue, by combining the data sets of Carrington and Detragiache (1998) and Adams (2003). We then augment the equation for I^i_{HK} with a variable p^i that denotes the probability of migration for workers whose skill level is equal to i.

$$I^i_{HK} = \alpha_0 + \alpha_1 K_{FDI} + \alpha_2 \ln(Y_{pc}) + \alpha_3 \tau + \alpha_4 D_{NR} + \alpha_5 p^i \qquad (7)$$

As a measure of p^i we take the migration rate for workers with educational level equal to i. The results, presented in table 8, are far less than encouraging. Due to the limited number of observations for p^i, the size of the sample shrinks dramatically. We drop the natural resource dummy, as it is never significant. Per capita income is the only significant determinant of secondary school enrollment. The stock of foreign capital has the expected positive sign but is not statistically different from zero at standard significance levels. For tertiary school enrollment, the tariff rate again enters the equation (as it did in table 8, with a negative coefficient), suggesting that trade protection discourages investment in human capital. The noticeable fact, though, is that the probability of migrating abroad (i.e., the variable p^i) plays no role whatsoever in determining the educational choices at either the secondary or tertiary level. Moreover, the coefficient on p^i is negative in both equations, contrary to the prediction of the brain gain model. These results should be taken with a grain of salt given the limited number of degrees of freedom. But they provide no evidence to support the brain gain argument.

TABLE 8.
Educational Achievement and the Brain Drain

	Secondary school enrollment	Tertiary school enrollment
K_{FDI}/Y	0.18 (1.26)	−0.06 (0.98)
$\ln(Y_{pc})$	22.8 (7.1)	8.4 (5.97)
$\ln \tau$	−1.41 (0.4)	−3.73 (2.4)
p^i	−0.06 (0.3)	−0.07 (1.4)
Estimation method	RE	RE
Hausman test [$\chi^2(4)$]	3.56	3.46
R^2	0.52	0.54
Number of observations	47	50

Source: Author's compilation.

Note: K_{FDI} = stock of foreign direct investment, Y = GDP, τ = tariff rate, abundant country, Y_{pc} = per capita income, p^i = migration rate for educational group i, FE = fixed effects, RE = random effects.

Conclusions

The interaction between trade, capital, and labor mobility is a key factor in determining the impact of globalization on developing countries. Separate analyses of the effects of trade, migration, and FDI are increasingly at risk of missing some key feedbacks between the various components of globalization.

One of the main findings of this paper is that trade liberalization, in addition to its standard and somewhat controversial effects on growth, also increases the host country's attractiveness for foreign direct investment. This adds a new channel through which a more liberal trade regime can favor growth.[11] Moreover, trade liberalization can also boost the investment in education and, hence, allow an economy to escape from a low-equilibrium trap.

We have also shown how skilled migration can interact with FDI and the investment in human capital to generate a welfare-inferior equilibrium, where foreign firms refrain from investing in the host country because of its inadequate supply of skills, while the incentives to become educated remain weak because of the lack of foreign capital. The brain drain in this setup means that a substantial share of skilled workers will migrate abroad, thereby aggravating the disincentive for foreign investors. The possibility that the brain drain may be turned into a brain gain, by raising the returns to skills, seems in this context remote and, at any rate, is not supported by the available evidence.

The policy message is that policy coherence, at both the domestic and international levels, is key. International institutions should lead the way in this respect, by better integrating their policy advice in the fields of trade, FDI, and migration. National policymakers should make sure that various policy provisions do not work at cross purposes with each other.

Finally, and perhaps more generally, we have shown how complementarities between the different aspects of globalization have become increasingly pervasive. This is both good and bad news for the world economy. It is good news to the extent that the effects of a more liberal regime tend to be mutually reinforcing; but it is bad news when backtracking in one area feeds negatively on other areas. Concerns about the recent trends in globalization and in the policy stance may be exaggerated. But it is a far greater risk to underestimate the dangers and the costs of a policy backtracking.

Annex 1. Estimation Methodolgy

All equations were first estimated with a random-effect estimator. The Hausman test was used to assess whether the unobservable random effects were correlated with the regressors. If so, the equation was reestimated with a fixed-effect estimator.

Regional dummies were added to each equation, and their joint significance was tested. Dummies were not included in the fixed-effect estimation.

In all but one equation, the dependent variable was measured by its average value over a five-year period (1981–1985, 1986–1990, 1991–1995, or 1996–2000). For

right-hand-side variables, we used their value at the beginning of the relevant five-year period.

The school enrollment equation relies on annual data from 1983 to 2000.

Annex 2. Data Sources

FDI (flows and stocks): UNCTAD database
Tariff rates: Ng (2004)
Openness (in constant prices): Heston, Summers, and Aten (2002)
GDP per capita in PPP (purchasing power parity): Heston, Summers, and Aten (2002)
Population: Heston, Summers, and Aten (2002)
Tertiary and secondary school enrollment: World Bank (2003)
Telephone lines: World Bank (2003)
Years of schooling: Barro and Lee (2000)

Notes

1. See, for example, Stark and others (1997, 1998) and Beine and others (2001, 2003).

2. Had we not controlled for the structural determinants of openness (income per capita, and population) the impact of tariffs would have been substantially larger. See Frankel and Rose (2000) for a similar result.

3. A further source of complementarity between trade and FDI may stem from the growing role of services. Particularly for less developed countries, inward FDI in marketing and services is complementary to their ability to export. This further strengthens the argument that during the 1990s FDI and trade in developing countries became strong complements.

4. However, Markusen and Maskus (2001), using aggregate U.S. data, found a positive relationship between the host country's trade costs and the affiliate's exports back to the home country. They interpret this finding as evidence of export platform FDI, where the firm's locational choice is dictated by the desire to cater to third markets.

5. The indicators are real GDP growth, GDP per capita, total exports as a percentage of GDP, telephone lines and mobile phones per 1000 inhabitants, per capita commercial energy use, research and development spending as a percentage of GDP, tertiary enrollment, country risk, exports of natural resources as a percentage of world total, exports in services as a percentage of world total, and inward FDI stock.

6. Inappropriate pooling of industrial and developing countries can strongly bias the results. See Blongen and Wang (2004).

7. The index PI is available only for the 1990s.

8. This is because the goods whose production is relocated to the south are skill-intensive for the south but unskilled-intensive for the north.

9. The derivative of FDI with respect to the stock of foreign capital is likely to be negative for reasonable values of the depreciation parameter.

10. Schiff and Wang (2004) found a similar result for Latin America.

11. The growth impact of FDI is also controversial in the literature. The findings by Borensztein and others (1998) have been challenged by Carkovic and Levine (2002).

References

The word *processed* describes informally reproduced works that may not be commonly available through library systems.

Adams, R. 2003. "International Migration, Remittances, and the Brain Drain." Policy Research Working Paper 3069. World Bank, Washington, D.C.

Barba Navaretti, G., J. Haaland, and A. Venables. 2002. "Multinational Corporations and Global Production Networks: The Implications for Trade Policy." Report prepared for the European Commission. Centre for Economic Policy Research (CEPR), London.

Barba Navaretti, G., and A. Venables. 2004. *Multinational Firms in the World Economy*. Princeton, N.J.: Princeton University Press.

Barro, R., and J. W. Lee. 2000. "International Data on Educational Attainments: Updates and Implications." Center for International Development Working Paper 42. Harvard University, Cambridge, Mass.

Beine, M., F. Docquier, and H. Rapoport. 2001. "Brain Drain and Economic Growth: Theory and Evidence." *Journal of Development Economics* 64: 275–89.

———. 2003. "Brain Drain and LDCs' Growth: Winners and Losers." IZA DP n. 819. Institute for the Study of Labor (IZA), Bonn, Germany.

Blomstrom, M., and A. Kokko. 2003. "Human Capital and Inward FDI." CEPR DP 3762. Centre for Economic Policy Research (CEPR), London.

Blongen, B., and M. Wang. 2004. "Inappropriate Pooling of Wealthy and Poor Countries in Empirical FDI Studies." NBER Working Paper 10378. National Bureau of Economic Research, Cambridge, Mass.

Borensztein, E., J. De Gregorio, and J-W. Lee. 1998. "How Does Foreign Direct Investment Affect Economic Growth?" *Journal of International Economics* 45: 115–35.

Carkovic, M., and R. Levine. 2002. "Does Foreign Direct Investment Accelerate Economic Growth?" University of Minnesota. Processed.

Carrington, W., and E. Detragiache. 1998. "How Big Is the Brain Drain?" IMF Working Paper 98/102. International Monetary Fund, Washington, D.C.

Collins, W., K. O'Rourke, and J. Williamson. 1999. "Were Trade and Factor Mobility Substitutes in History?" In R. Faini, J. de Melo, and K. Zimmermann, eds., *Migration: The Controversies and the Evidence*. Cambridge, U.K.: Cambridge University Press.

Faini, R. 2004. "Does the Brain Drain Boost Growth?" University of Rome Tor Vergata. Processed.

Faini, R., J.-M. Grether, and J. de Melo. 1999. "Globalization and Migratory Pressures from Developing Countries: A Simulation Analysis." In R. Faini, J. de Melo, and K. Zimmermann, eds., op. cit.

Faini, R., and A. Venturini. 1993. "Trade, Aid, and Migration: Some Basic Policy Issues." *European Economic Review* 37: 435–42.

Feenstra, R., and G. Hanson. 1996a. "Foreign Investment, Outsourcing, and Relative Wages." In R. Feenstra et al., eds., *The Political Economy of Trade Policy. Papers in Honour of J. Bhagwati*." Cambridge, Mass.: MIT Press.

———. 1996b. "Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras." *Journal of International Economics* 42: 371–93.

Fontagne, L. 1999. "Foreign Direct Investment and Trade: Complements or Substitutes?" STI Working Paper 1999/3. Organisation for Economic Co-operation and Development (OECD), Paris.

Frankel, J., and A. Rose. 2000. "Estimating the Effect of Currency Union on Trade and Output." NBER Working Paper 7857. National Bureau of Economic Research, Cambridge, Mass.

Goldberg, L., and M. Klein, 1999. "International Trade and Factor Mobility: An Empirical Investigation." NBER Working Paper 7196. National Bureau of Economic Research, Cambridge, Mass.

Hanson, G., R. Mataloni, and M. Slaughter. 2001. "Expansion Strategies of U.S. Multinational Firms. In D. Rodrik and S. Collins, eds., *Brookings Trade Forum 2001*. Washington, D.C.: Brookings Institution.

Heston, A., R. Summers, and B. Aten. 2002. Penn World Table 6.1, Center for International Comparisons at the University of Pennsylvania, Philadelphia.

Kee, H. L., A. Nicita, and M. Olarreaga. 2004. "Ad Valorem Equivalents of Nontariff Barriers." International Economics Department, World Bank, Washington, D.C. Processed.

Markusen, J. 1983. "Factor Movements and Commodity Trade as Complements." *Journal of International Economics* 13: 341–56.

———. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, Mass.: MIT Press.

Markusen, J., and K. Maskus. 2001. "Multinational Firms: Reconciling Theory and Evidence." In M. Blomstrom and L. Goldberg, eds., *Topics in International Economics: A Festschrift in Honor of Robert Lipsey*. Chicago: University of Chicago Press.

Mountford, A. 1997. "Can a Brain Be Good for Growth in the Source Economy?" *Journal of Development Economics* 53: 287–303.

Ng, F. 2004. Tariff Data on Trade and Import Barriers. World Bank Trade Web site. Washington, D.C.

Rodrik, D., and F. Rodriguez. 2000. "Trade Policy and Economic Growth: A Sceptic's Guide to the Cross-National Evidence." In B. Bernanke and K. Rogoff, eds., *NBER Macroeconomics Annual*. Cambridge, Mass.: MIT Press.

Schiff, M., and Y. Wang. 2004. "Education, Governance, and Trade-Related Technology Diffusion in Latin America." IZA DP 1028. Institute for the Study of Labor (IZA), Bonn, Germany.

Shatz, H., and A. Venables. 2000. "The Geography of International Investment." Policy Research Working Paper 2338. World Bank, Washington, D.C.

SOPEMI. 2003. *OECD SOPEMI Trends in International Migration*.

Stark, O. 2004. "The Economics of the Brain Drain Turned on Its Head." *ABCDE—Europe 2004 (Oslo Proceedings)*. Washington, D.C.: World Bank.

Stark, O., C. Helmenstein, and A. Prskawetz. 1997. "A Brain Gain with a Brain Drain." *Economic Letters* 55: 227–34.

———. 1998. "Human Capital Formation, Human Capital Depletion, and Migration: A Blessing or a Curse?" *Economic Letters* 60: 363–67.

UNCTAD. 1996. *World Investment Report: Investment, Trade, and International Policy Arrangements*. Geneva: United Nations Conference on Trade and Development.

———. 2003. *World Investment Report: FDI Policies for Development: National and International Perspective*s. Geneva: United Nations Conference on Trade and Development.

Wacziarg, R. and K. Horn Welch. 2003. "Trade Liberalization and Growth: New Evidence." NBER Working Paper 10152. National Bureau of Economic Research, Cambridge, Mass.

Walmsey, T. and A. Winters. 2002. "Relaxing the Restrictions on the Temporary Movement of Natural Persons: A Simulation Analysis." University of Sussex, U.K. Processed.

World Bank. 2003. *World Development Indicators*. Washington, D.C.: World Bank.

Comment on "Trade Liberalization in a Globalizing World" by Riccardo Faini

GERRY K. HELLEINER

One can only applaud any paper that attempts to offer an integrated analysis of the impact upon developing countries of the globalization phenomenon, incorporating factor movements as well as flows of goods and services. Its plea for more coherent development policies, at both national and international levels, is one that will resonate profoundly with all who have attempted to address these issues on a partial basis—and that, I am afraid, must include all of us. Indeed, as will be seen, the author might well have tried to incorporate still more factor flows and other economic influences within a more fully integrated analysis.

Professor Faini's paper properly calls attention to the fact that whereas trade and foreign direct investment (FDI) policies have generally been liberalized, contributing to a disproportionately rapid increase in international trade and FDI flows in recent years (at least until 2001–03), immigration policies in receiving countries have become more restrictive and potentially damaging. It also highlights the growing relative importance of and increased openness to trade and FDI in developing countries. These are key contextual elements for any current and prospective analysis of globalization-development interactions.

The principal object of Faini's paper is to explore the interactions, potential and actual, between trade, FDI, and labor flows; and policies relating to each. Curiously, it does not address the surely no less important trends in capital flows and capital account regimes more generally. Most accounts of the globalization phenomenon, after all, rate capital market integration at or near the top of their lists of its primary features. Moreover, FDI is seen in the dominant international business (Caves/ Dunning) paradigm as a matter primarily of industrial organization—relating to the exploitation of firm-specific assets, location, and internalization decisions—rather than of either capital flow (which often, when locally financed, fails to accompany FDI), or real new bricks and mortar (which is absent in the case of mergers and

Gerry K. Helleiner is Professor Emeritus in the Department of Economics and Distinguished Research Fellow at the Munk Centre for International Studies at the University of Toronto, in Ontario, Canada.

acquisitions). The paper also does not directly address the trends in technology flows and technology policies, which possess lives of their own quite apart from FDI or trade, and therefore constitutes a promising beginning rather than an attempt at a fully integrated analysis of the economic aspects of globalization.

I believe the paper has serious problems, even within the confines of its prescribed subject matter. The author posits that continued trade liberalization by developing countries will be good for their growth, not only for the "usual" orthodox reasons (about which he concedes there may be some doubt) but also because it will increase both FDI (the effects of which, he notes, may also be in some doubt) and some kinds of human capital. Professor Faini's concessions to the doubters have clearly not altered his "priors" favoring the positive effects of both trade liberalization and FDI, and he believes he has found further reasons for developing countries to liberalize their trade (and encourage FDI). He consequently expresses concern that what he sees as a current slowdown in trade and (FDI policy) liberalization may be especially "bad news" in that it may be doubly harmful.

The author constructs his case for trade liberalization (and concern about its slow-down) on the basis of a little theory, some very aggregative data, and regression equations that he himself describes as "very simple." I do not myself believe he has made a persuasive case.

It is certainly true, as Professor Faini notes, that the literature is now inconclusive as to the growth effects of trade liberalization or, for that matter, on the relative importance of trade (as against other) policies. It is equally inconclusive as to the poverty effects of liberalization (Winters et al. 2004), a matter of prime current policy concern not discussed in this paper.

The author begins his empirical analysis by building upon a particular section of a recent paper by Wacziarg and Welch (2003), which found, not surprisingly, that trade liberalization increases trade openness (a result that, Faini fails to note, Wacziarg and Welch did not obtain for the 1990s). Other parts of that paper were rather more interesting for the purposes of Faini's analysis. For instance, Wacziarg and Welch confirmed the Rodriguez-Rodrik (2000) results as to the absence of any cross-country growth impact of trade restrictions. They also found that the much-cited Sachs-Warner results (1995), purporting to show negative growth effects from trade barriers (and criticized by Rodriguez-Rodrik), break down completely in the 1990s.

Perhaps most interesting of all from the standpoint of the paper now under discussion, Wacziarg and Welch found that, within countries, trade liberalization "episodes" were on average associated with subsequent increases not only in openness but also in investment and growth—and that these results were stronger in the 1990s than before. Professor Faini describes this finding as an indicator that trade liberalization has "a positive and robust effect on growth and investment." The authors themselves are more cautious. They call attention to the "vast amount of heterogeneity across countries in the extent to which growth rose after trade reforms." "While the average effect of change in the large sample is positive," they continue, "roughly half of the countries experienced zero or even negative changes in growth post-liberalization" (p. 3). Indeed Wacziarg and Welch go to some length to explore the reasons for this

heterogeneity—institutions, politics, the nature of the reforms, concurrent macroeconomic policies, and so on—but find it difficult to draw generalized conclusions. It is worth noting in this context that another recent econometric analysis, using standard cross-national regression methodology and a wide sample, found that trade barriers, while negatively associated with openness, are positively associated with growth, particularly in developing countries (Yanikkaya 2003). This would all seem to argue for much more research on the reasons for such diverse growth (and poverty alleviation) experience with trade liberalization, and indeed with all manner of other trade and trade-related policies.

But Professor Faini instead questions, again, whether trade liberalization does, after all, increase trade openness. He reverts to cross-national regression analysis, adds a couple of variables (population and per capita income), uses a crude measure of trade barriers (import duties over import value), and derives the expected result: it does. Yanikkaya, with more careful analysis, also got this result. To my knowledge, such an association has never been in doubt. But, more intriguing, Yanikkaya also found, as already noted, that trade barriers were positively associated with growth. This suggests that some kinds of "activist policies" (proxied by our inadequate measures of trade barriers) may both reduce openness and increase growth. That is, to put it the other way round, trade liberalization can both increase openness and, at the same time, reduce growth. And there is supportive empirical evidence that it does. This possibility does not exist in the present author's (or most other mainstream) analysis, but would seem to cry out for more detailed country-level exploration.

In any case, it is important to get beyond the crude measures of trade policy upon which such regression analyses rely. (The paper under discussion relies upon import duties divided by import value, sometimes supplemented by the average tariff equivalent of nontariff barriers.) This is necessary for at least four reasons: 1) The structure of trade barriers—their dispersion and the nature of their selectivity—is likely to be at least as important to growth (and poverty alleviation) as their average size. 2) Direct incentives for exports have historically been as important as import barriers, and, for large numbers of developing countries, they remain WTO-compatible (while some other countries are finding ways around the legalities). 3) The behavior of the real exchange rate is of major significance to the overall incentive structure, and over the medium term it usually swamps the effects of trade barriers. 4) Non-trade policy incentives are usually also of major significance (particularly those in the financial sector), and they need to be integrated into any overall analysis of incentive structures.

But let me return to Faini's paper. What about the relationship between trade barriers and FDI? The author posits that the nature of FDI in developing countries has changed in recent years. No longer primarily seeking natural resources or jumping trade barriers to gain access to markets, FDI is now primarily "efficiency-seeking." It seeks to cut production costs by "slicing the value-added chain" and relocating intermediate production of goods and services. (Others would call this "outsourcing," which may or may not involve FDI.) "Vertical" FDI, he argues, now dominates "horizontal" FDI. From this it follows that trade barriers will now, on balance, discourage (total) FDI. Restricting the data to the 1990s and to developing countries,

and again using fairly crude indicators, this is indeed what he finds. Total inward FDI (normalized by GDP) is higher when trade barriers are lower. Replicating the test for earlier periods and adjusting his methodology because of data problems, Faini finds that trade barriers previously had no effect on FDI. Trade liberalization, he concludes, will now also encourage expanded FDI; whereas this was not the case in the past.

Most of us, and certainly all policymakers, would want to be more careful to specify the particulars of industries and countries before offering such bold assessments of trade policy-FDI interactions. Nothing has changed in the relationship between trade barriers and either vertical or horizontal FDI. Whether FDI is vertical or horizontal still depends upon country characteristics (notably the factor endowments) of home and host countries; and the specifics of trade-FDI relationships still change over time with changing costs of trade and other factors. (For a recent analysis see Amiti and Wakelin 2003.) Trade liberalization will still, most assuredly, reduce market-seeking inward FDI. Ask any government that has reduced its import barriers. Is it really helpful to try to analyze FDI experience, still less to offer policy recommendations, at such an aggregated level? In any case, if one wants to know whether the importance of "efficiency-seeking" FDI and other outsourcing is increasing, would it not make more sense to seek its measurement directly?

It may be worth observing, in this context, that multinational firms have always favored trade liberalization, particularly for the products in which they trade. From their standpoint, trade and FDI have always been complementary. But from the standpoint of developing country policymakers, the questions are more complex. They seek guidance on which types of FDI are most valuable in their particular circumstances, and want to know what policies are necessary to acquire and benefit from it. Trade policies may matter in this respect, as for instance in export-processing zones or in spillover-generating import substituting activities, but they do not always dominate. Generalizations are therefore of dubious value.

Faini also finds that inward FDI is positively related to an element of the host countries' infrastructure and human capital. He notes the two-way causality between skilled labor and FDI: To be productive, FDI requires a certain absorptive capacity (including skills); at the same time, FDI may also create local demand for such skills. Fully aware that there may still be some "leakages" through brain drain, most economists would probably recommend breaking through any possible resulting low-level equilibrium with direct subsidies to education, which will increase the skill base and thus attract more FDI (e.g. Hoffman 2003). But the paper under discussion is on trade policy. Heroically, the author therefore seeks an empirical link between investment in human capital (secondary school and tertiary enrollment ratios), trade barriers, and the total stock of FDI. He makes no attempt to distinguish differential effects of different kinds of FDI. Secondary school enrollment, he finds, is associated with lower trade barriers (and higher FDI stock). Lower trade barriers are sometimes also associated with tertiary enrollment in Africa and Latin America.

From all of these regression results Faini concludes that trade liberalization will both attract foreign firms and increase investment in human capital. This seems to me to be quite a stretch.

There is so much we still need to know about the real-world relationships between trade, investment, government policies, and growth, not to speak of poverty alleviation. I seriously question whether there are still significant returns to further fiddling with cross-national regressions based on such crude measures of complex realities.

Allow me one further word. Are the world economy and the fate of the developing countries really at risk from the purported pause in liberalization or instead, as the author says, from the collapse of the multilateral agreement on investment (MAI) and the deadlock in the World Trade Organization (WTO)? I hardly think so. The MAI was severely flawed in many dimensions, and its provisions would never have been acceptable to most developing countries. It is more important for the WTO and other rules systems to be broadly fair and acceptable, however long it may take to get them right, than to rush to further liberalization (but never for labor, of course) as interpreted by the major economic powers. There has already been significant worldwide liberalization, and there has probably been a high return from it. But it is not obvious, on the basis of existing evidence, that what development and poverty alleviation most now require is an immediate further (and probably seriously biased) burst of liberalization. The completion of a Doha Round with a serious and careful emphasis upon development effects could be very helpful. But, so far, the development emphasis has only been rhetorical. If the current round of WTO negotiations fails it will not necessarily be, as some suggest, a disaster for development. What we should have learned from the past, it seems to me, is that if the current so-called "Development Round" fails, we shall have to try again.

References

Amiti, Mary, and Katharine Wakelin. 2003. "Investment, Liberalization, and International Trade." *Journal of International Economics* 61 (1): 101–26.

Hoffman, Anders N. 2003. "Education, Trade, and Investment Liberalization." *Journal of International Economics* 60 (2): 433–53.

Rodriguez, Francisco, and Dani Rodrik. 2000. "Trade Policy and Economic Growth: A Skeptic's Guide to Cross-National Evidence." In B. Bernanke and K. Rogoff, eds., *NBER Macroeconomics Annual 2000*. Cambridge Mass.: MIT Press.

Sachs, Jeffrey, and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity 1995*:1: 1–118.

Wacziarg, Romain, and Karen Horn Welch. 2003. "Trade Liberalization and Growth: New Evidence." NBER Working Paper 10152. National Bureau of Economic Research, Cambridge, Mass.

Winters, L. Alan, Neil McCulloch, and Andrew McKay. 2004. "Trade Liberalization and Poverty: The Evidence So Far." *Journal of Economic Literature* 42 (1): 72–115.

Yanikkaya, Halit. 2003. "Trade Openness and Economic Growth: A Cross-Country Empirical Investigation." *Journal of Development Economics* 72 (1): 57–89.

Comment on "Trade Liberalization in a Globalizing World" by Riccardo Faini

JEFFREY G. WILLIAMSON

Motivation and Agenda

The links between trade and factor mobility have always been central to traditional trade theory. Indeed, Eli Heckscher and Bertil Ohlin viewed them as substitutes. After all, if, as they thought, factor endowments dictate trade specialization, then anything that reduces differences in factor endowments should reduce the gains from trade and the amount of trade. Thus, labor migrations from labor-abundant to labor-scarce countries should diminish trade, and capital flows in the opposite direction should do the same. More to the point of Riccardo Faini's typically thoughtful paper, consider the obvious corollary: If trade barriers go up, there is more pressure on labor and capital to migrate to countries where they are scarcest and best rewarded; and if trade barriers go down, there is less pressure for labor and capital to migrate. The corollary has important policy implications, as illustrated recently by the political expectation that NAFTA would reduce Mexican migration to the United States.

It has been a long time since Heckscher and Ohlin wrote down their theorem, and since then economists have come to recognize that these predictions are very much the result of their assumptions. Modern theory simply will not tell us whether trade and factor flows are complements or substitutes (Faini et al. 1999). However, even an ambiguous theory can offer powerful morals. Here are two: First, it is unwise to analyze either trade, labor migration, or capital flows independent of the other two; and second, only good empirical analysis will overcome the theoretical ambiguities. Professor Faini acts on both of these morals and in so doing advances our knowledge into areas where the World Bank definitely needs to go. However, a discussant is obliged to have some complaints.

Jeffrey G. Williamson is the Laird Bell Professor of Economics at Harvard University in Cambridge, Massachusetts.

Measuring What Matters: Trade Policy

Professor Faini deals with issues of trade openness and trade policy throughout the paper he presented at this conference, but I am not completely happy with where this discussion leaves us.

For example, the standard approach in the empirical literature on trade and development is to use the trade share as a measure of openness. We know this is an endogenous variable, driven by trade policy *and* a ton of domestic factors. Since I am an economic historian, you will not be surprised by my favorite example of how very bad $(X + M) / Y$ is as a measure of openness. Over the half-century from the 1870s to the Great War, Latin American trade grew faster than in any other region, and trade shares rose to higher levels there than in any other region (Bulmer-Thomas 1994). And over the same half-century, Latin American tariffs were also higher than anywhere else, five times as high as those of the allegedly "protectionist" European continent (Coatsworth and Williamson 2004). Latin America had high and rising trade shares during its *belle époque* in spite of restrictive trade policy, not because of liberal trade policy.

The regression in table 3 of the paper is supposed to persuade me that the history just cited is too old, and that modern economies are somehow different. But I am not yet persuaded. First, liberal policies come in packages (Lindert and Williamson 2003). Thus, there is a tendency for declining tariffs to be correlated with the deregulation of domestic capital, labor, and commodity markets. The regression in table 3 makes no effort to control for this reality. The correlation between high trade shares and low tariff rates may be spurious, both of them correlated with domestic liberalization. Second, while Professor Faini reports and discusses the coefficients in table 3, he doesn't tell us whether changing tariff rates explain much of the observed changes in trade shares. Big or small? Third, the data underlying table 3 are a panel, four periods (1981–85, 1986–90, 1991–95, 1996–2000) times 92 developing countries. Anyone in this room who has used such panel data knows that the vast majority of the variance is between countries, not over time. I don't know about you, but I like to see policy assessed by doing comparative history, not by exploring the difference between countries that also differ in culture, geography, demography, institutions, religion, and policy, and where the analysis fails to control explicitly for such influences.

Measuring What Matters: FDI

I found Professor Faini's separation of foreign direct investment (FDI) into three parts very useful. First, there is *horizontal FDI,* where the motivation is to circumvent trade barriers, save on trade costs, and gain access to foreign markets. However, since most developing countries have small GDP and small markets, Professor Faini correctly observes that this motivation is unlikely to play a big role in most developing countries. The important point, however, is that horizontal FDI is negatively correlated with low trade barriers and is trade destructive. Second, there is

resource-seeking FDI, where the motivation is to exploit local resources and to foster the developing country's primary-product-producing export sector. This kind of FDI is clearly trade creating. Third, there is *vertical FDI,* where the motivation is to cut costs by outsourcing intermediate stages of production. This kind of FDI is also trade creating, since what was previously produced in the advanced economy is now partly produced in the developing economy, and traded.

These useful descriptive labels imply an equally useful empirical agenda, but I think Professor Faini could have pushed his agenda much farther. The best way to test these propositions is to control for country attributes (e.g., endowments and technologies) that distinguish the three types of FDI. To control for *resource-seeking FDI,* why not add the standard variable that Jeffrey Sachs and Andrew Warner (1995) made popular, the share of exports that are resource-intensive? Alternatively, since African exports are more resource-intensive than are those of the rest of the developing world, why not explore regional variance in the FDI-openness correlation? Similarly, to uncover the role of *horizontal FDI,* why not control for the size of developing country GDP? Alternatively, is the FDI-openness correlation different for Brazil, China, India, and Indonesia than for the rest? And to uncover the role of *vertical FDI,* why not control for the education and skill attainment of the developing country's labor force (using Barro-Lee)? The econometric style should be to *interact* such variables, not simply to add them on.

Turning to table 4, I recommend some further additions. While Professor Faini doesn't tell us how the United Nations Conference on Trade and Development (UNCTAD) constructs its "potential index of inward foreign direct investment" (PI), I don't see how it gives us more insight to discover that FDI is highly correlated with an index that has been constructed to measure FDI attractiveness. What I would like to see instead is a regression that identifies the fundamentals driving PI, or a detailed discussion of the UNCTAD weights and their justification, or both. I would also like to see how UNCTAD's potential index matches up against the World Bank's "investment climate" index.

Measuring What Matters: World Migration

Professor Faini poses two great questions in the last section of his paper, where he confronts the connections between FDI, emigration, brain drain, and human capital accumulation. Does the presence of foreign firms raise the incentive for human capital accumulation, thus fostering growth? Does emigration of the skilled raise the incentive for human capital accumulation at home, thus converting the brain drain into a brain gain? Alas, the answers are not yet quite as great as the questions.

I like the motivation behind table 5, or at least the part which implies that FDI is attracted by good schooling endowment. I like it especially since this correlation is as old as global capital markets. During the first global century before World War I, the Lucas Paradox (Lucas 1990) was just as apparent as it has been during the second global century after World War II (Clemens and Williamson 2004). Namely, foreign

capital tends to flow to rich countries, not to poor countries. So what are the missing complements to capital whose absence lowers the rate of return to capital and thus deflects foreign capital from poor to rich countries? One of the most important missing complements during the first global century was schooling. It is comforting to find this result again in table 5, this time for recent FDI. Once again, the past is confirmed by the present.

Table 5 has already shown us that FDI chases after good schooling; so I am happy to see that positive coefficient on FDI in table 6, which explains secondary school enrollment rates. So far, so good. But Professor Faini doesn't tell me what theory predicts that a Third World pro-trade policy, and thus presumably a trade boom, will foster a heightened demand for secondary schooling. Don't poor countries have a comparative advantage in manufacturing activities that require lower skills and schooling? Higher tariffs are correlated with lower secondary school enrollment rates, but it is not at all clear what the channel of impact is. What would be far more persuasive would be an analysis broken down by sectors. And while Professor Faini is plunging ahead on this very productive agenda, I also urge him to pay more careful attention to causality.

Finally, the brain drain analysis needs to be sharpened. At the start of this section of the paper, we are asked to consider the following thought experiment, to wit: "suppose … that the brain drain leads to a sudden loss in the sending country's skilled labor force." This thought experiment invites us to treat the brain drain as exogenous. Of course, it is not: It is the low returns on educational investment at home that sends skilled and schooled people abroad. And these low returns are generated by the absence of the complementary inputs that raise those returns—sound property rights, stable governments, the absence of violence, technological advance, excellent social overhead, and so forth. In the absence of those complements, it is only the employment options abroad that keep the return on education from falling still further. Table 8 does not control for these complements. Thus, the negative (but insignificant) coefficient on the emigration rate of the schooled regressed against enrollment rates may tell us nothing about whether a brain drain stimulates schooling at home. Instead, it may only tell us that both are driven by third factors that are missing from the analysis.

In spite of these complaints, I applaud Professor Faini's start and urge him to continue this agenda—and hope he can persuade the World Bank to join him. It is exciting, and it is important.

References

Bulmer-Thomas, Victor. 1994. *The Economic History of Latin America since Independence.* Cambridge, U.K.: Cambridge University Press.

Clemens, Michael, and Jeffrey G. Williamson. 2004. "Wealth Bias in the First Global Capital Market Boom 1870–1913." *Economic Journal* 114 (495): 304–37.

Coatsworth, John, and Jeffrey G. Williamson. 2004. "The Roots of Latin American Protectionism: Looking before the Great Depression." In Antonie Estevadeordal, Dani

Rodrik, Alan M. Taylor, and Andres Velasco, eds., *Integrating the Americas: FTAA and Beyond*. Cambridge, Mass.: Harvard University Press.

Faini, Riccardo, Jaime de Melo, and Klaus Zimmermann, eds. 1999. *Migration: The Controversies and the Evidence*. Cambridge, U.K.: Cambridge University Press.

Lindert, Peter H., and Jeffrey G. Williamson. 2003. "Does Globalization Make the World More Unequal?" In Michael Bordo, Alan M. Taylor, and Jeffrey G. Williamson, eds., *Globalization in Historical Perspective*. Chicago: University of Chicago Press.

Lucas, Robert. 1990. "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review* 80 (2): 92–96.

Sachs, Jeffrey D., and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1: 1–118.

Themes and Participants for the

17TH ANNUAL WORLD BANK CONFERENCE ON DEVELOPMENT ECONOMICS (ABCDE)

Dakar, Senegal

"GROWTH AND INTEGRATION"

JANUARY 27, 2005

Growth and Integration

Financial Reforms

Economic Development

Trade and Development

Abdoulaye Wade • Abdoulaye Diop • François Bourguignon • Augustin Fosu • Stephen O'Connell • Gaye Daffe • Jean-Paul Azam • Abdoulaye Diagne • Lemma W. Senbet • Jean-Claude Berthelemy • Mthuli Ncube • Tche Tche Nguessan • Larry Hinkle • Richard Newfarmer • Alan Gelb • Vijaya Ramachandran • Ben Eifert • Leonce Ndikumana • Ademola Oyejide

The Annual World Bank Conference on Development Economics (ABCDE) brings together the world's leading scholars and development practitioners for a lively debate on state-of-the-art thinking in development policy and the implications for the global economy.

The 16th conference was held in Washington, DC, on May 3–4, 2004. The theme of the conference was lessons of development experience, which was divided into four topics: behavioral economics, lessons of experience, infrastructure and development, and trade and development.

IN THIS VOLUME

Introduction by François Bourguignon and Boris Pleskovic; opening address by James D. Wolfensohn; keynote addresses by François Bourguignon and Vernon L. Smith; papers by Sendhil Mullainathan, Richard N. Cooper, Gustav Ranis, Rémy Prud'homme, and Riccardo Faini; and comments by Colin F. Camerer, Karla Hoff, Abhijit V. Banerjee, Pedro S. Malan, Eduardo Engel, Nemat T. Shafik, Gerry K. Helleiner, and Jeffrey G. Williamson.

OXFORD
UNIVERSITY PRESS

THE WORLD BANK