Annu

World Bank

Confer

on Develop

Econom

1996

Edited by Michael Bruno and Boris

Annual World Bank Conference on Development Economics 1996

Edited by **Michael Bruno and Boris Pleskovic**

The World Bank Washington, D.C.

First printing January 1997

Michael Bruno is senior vice president, Development Economics, and chief economist at the World Bank. *Boris Pleskovic* is deputy administrator, Research Advisory Staff, at the World Bank.

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 subject to an internal and external peer review process. Summaries are included of floor discussions, which attempt to convey the sense and substance of what was discussed, interventions by participants from the floor, and responses by panelists. These summaries have not been reviewed by the presenters, the discussants, or other participants. Participants' affiliations identified in this volume are as of the time of the conference, April 25–26, 1996.

Meta de Coquereaumont, Paul Holtz, and Barbara Karni edited this volume. Pat McNees worked on the floor discussion. Mark Bock was responsible for composition and layout, and Wendy Guyette helped with coordination. All are with American Writing Corporation.

Annual World Bank Conference on Development Economics is indexed in *Human Resources Abstracts,* the *Index of Economic Articles,* the *Index of Social Science and Humanities Proceedings,* the *Index to International Statistics,* the *Public Affairs Information Service, Sage Public Administration Abstracts,* the *Standard Periodical Directory,* and *Ulrich's International Periodicals Directory* and online by ABI/INFORM and DIALOG.

ISSN 1020-4407
ISBN 0-8213-3786-6

Contents

Introduction

Michael Bruno and Boris Pleskovic

The articles in this volume were initially presented as papers at the Eighth Annual World Bank Conference on Development Economics, held April 25–26, 1996, in Washington, D.C. The conference series seeks to expand the flow of ideas among development policy researchers and practitioners around the world. It is a forum for World Bank and outside experts to exchange ideas, challenge and build on one another's findings, and expand knowledge of the theories and practice of development. Each year the topics selected for the conference represent new areas of concern or areas that will benefit from a review of what is known and the identification of what still needs to be explored. To further these ends, participants at the eighth annual conference were drawn from a broad range of research, academic, and policymaking institutions in both developing and industrial countries.

The eighth conference addressed four themes: banking failures as crises or opportunities for reform, reducing poverty through targeted programs and rural finance, legal systems and economic development, and labor and environmental standards in international trade. Despite these varied topics the articles in this volume share two overarching themes: the appropriate role of government in economic development is changing, and there is a central role for government policy in each area addressed that is critical to development. The articles also focus on the role of institutions, the importance of political and economic competition, and the role of law and political organization.

Several articles stress the vital role of institutions and related incentives in finance: Frederic S. Mishkin and Gerard Caprio Jr. and Daniela Klingebiel on banking and Timothy Besley and Ernest Aryeetey on rural finance and poverty. These articles and Joseph E. Stiglitz's keynote address discuss the role of government as an institution and the need to strike a balance between introducing competition within government and devolving government activities to the market, and between deregulation and correction of market failures and the creation of markets.

Michael Bruno is senior vice president, Development Economics, and chief economist at the World Bank. Boris Pleskovic is deputy administrator, Research Advisory Staff, at the World Bank.

Annual World Bank Conference on Development Economics 1996

In their discussions of the importance of political as well as economic competition Stiglitz and others note that political competition in bureaucracies may be an effective surrogate for market competition. Also clear from the discussions, especially the article by Alan Krueger, is the growing recognition that as income rises so does demand for labor and environmental standards. Kym Anderson shows that linking environmental standards to trade will not achieve the goals sought by the standards and could disrupt trade. Multilateral organizations may have an important advocacy role in ensuring fair play in administering standards.

The discussion of the role and rule of law emphasizes the links between demand for law and the legal and political systems. Legal systems must be country or context specific, argue both Robert Cooter and Avner Greif. There is no unique, optimal combination of contract designs and enforcement mechanisms because culture plays an important role in the interaction of social, political, and economic forces.

Roles of Multilateral Organizations and Government

In his opening remarks James D. Wolfensohn, president of the World Bank, notes that these conferences, which began in 1989, have become a premier forum for the presentation and dissemination of cutting-edge thinking on some of the most important issues in development. If our goal is to improve results on behalf of developing countries, he argues, we need to expand and intensify our thinking on development issues. In today's rapidly changing global economy the World Bank is an important connecting rod between research and practice. Wolfensohn stresses that our role in generating ideas requires complete openness and interchange between ourselves and our colleagues outside the organization. He sees this annual conference as an important part of our effort to fulfill our global role and as a laboratory for the ideas that will make a difference to the quality of life of the world's poor. He notes that successful development should be based on strong macroeconomic policy fundamentals and that the topics of the conference are all priority issues for developing countries.

In the keynote address Joseph E. Stiglitz, chairman of the U.S. Council of Economic Advisers, focuses on the appropriate role of government in economic development, the topic of the 1997 *World Development Report*. Stiglitz advocates a balanced perspective in assessing the role of government in today's economy. Instead of repeating old arguments about the superiority of private over public enterprise, he said, researchers and policymakers should identify areas where the activities of governments and private markets can most effectively complement each other. Stiglitz argues that the role of government is dynamic. Thus changes in the economic environment fundamentally alter what government can and should do. Stiglitz notes that earlier U.S. experience and more recent experiences in East Asia suggest that government policy can play a useful economic role in such key areas as education, technology, financial regulation, infrastructure, environmental protection, and social welfare.

Banking Failures—Bane or Boon?

Frederic S. Mishkin develops a tool for analyzing banking and financial crises in developing countries based on an asymmetric information framework that provides insights into causes and treatments of financial crises. Some of the causes are the same in industrial and developing countries—rising interest rates, increased uncertainty, deteriorating balance sheets, and bank panics. But some factors differ in ways that have important policy implications for developing countries. In industrial countries the monetary authorities can promote recovery through expansionary monetary policies, but in developing countries the institutional features of the financial system (high and variable inflation, liabilities denominated in foreign currencies) make promoting recovery much more difficult. Expansionary monetary policies would likely cause expected inflation to rise dramatically and the domestic currency to depreciate sharply, worsening rather than improving the balance sheets of households, firms, and banks. Mishkin argues that in developing countries monetary policy ought to be restricted to promoting low inflation and restoring confidence in the economy, a much more complicated exercise. Slowly phasing in financial sector deregulation and liberalization until supervisory systems are in place is crucial, since risk-taking behavior will otherwise not be adequately constrained, with potentially disastrous economic and financial consequences.

Also on the theme of banking, Gerard Caprio Jr. and Daniela Klingebiel note that in the past fifteen years bank insolvencies have occurred around the globe, affecting developing and industrial countries alike. Where these failures are systemic, they can drain a country's financial, institutional, and policy resources—resulting in large losses, misallocated resources, and slower growth. The authors explore the reasons for these failures and consider what can be done about them. Using a new database covering some eighty cases of insolvency, Caprio and Klingebiel analyze the microeconomic and macroeconomic factors contributing to these crises and assess how governments and bankers have responded—unfortunately, not well. Too often, banks conceal the magnitude of their financial woes. Some continue to make loans in an effort to spur recovery; others fail to account for nonperforming loans in their balance sheets. And because many countries have lax regulatory standards for banks—in many countries banks are essentially self-regulating yet have little at risk—these practices are able to continue until it is too late to prevent a financial crisis. To better manage insolvencies, Caprio and Klingebiel conclude, policymakers must develop a regulatory framework that allows banks to respond more robustly to various shocks. That bankers have not regularly planned for shocks suggests that they have not had the incentives to do so. Promoting greater diversification, especially in small developing economies, will also be essential to improving the stability of banking systems.

Reaching the Poor

Exploring the topic of poverty, Timothy Besley states that understanding how to alleviate poverty is the central concern of development economics. Although long-term

economic growth is known to significantly reduce poverty, over the shorter term governments and aid agencies must develop policies to ensure that all people have the means of achieving basic subsistence. Besley identifies two somewhat opposing perspectives on these issues: the technocratic and the institutional. Economists tend to rely on the technocratic approach, which emphasizes designing policies that target the poor as effectively as possible. Noneconomists and nongovernmental organizations are more comfortable with the institutional approach, which emphasizes the development of institutions and the improvement of government performance as means of alleviating poverty. Because those who finance transfer programs want them to be designed in a way that minimizes their financial burden, efforts are often made to carefully identify those who deserve resources under a given program. But such research is costly. Thus better targeting and cost minimization are not always parallel objectives—targeting is less about economic efficiency and more about redistributive tradeoffs. Optimal policy depends on country-specific factors—institutions, history, and especially political structure. Ultimately, Besley concludes, a two-pronged assault on poverty is required: better-targeted interventions and ways of delivering resources to the poor that defy or change the governing political mechanism.

In a second article on poverty Ernest Aryeetey notes that liberalizing financial markets will not of itself improve credit delivery in Africa. The obstacles to financial market development stem not only from the wrong policies but also from deep structural and institutional roots that cannot be addressed with macroeconomic policy reform alone. There is almost no interaction between formal institutions, such as banks, and informal institutions, such as savings and loan associations, *susu* collectors, and moneylenders. Each type of institution lends to its own distinct clientele, largely because of weak contract enforcement and therefore fear of losses if loans are extended outside known groups. This fragmentation squeezes one group of would-be borrowers the hardest: small rural farms and businesses that need medium-term loans. Banks will not lend to them because they do not meet their eligibility criteria, while informal lenders do not have the money to lend or cannot offer them the terms they need. Women and the poor also may fail to get loans for the same reasons. Aryeetey recommends a three-tiered system in which banks lend to semiformal institutions, such as savings and loan companies or credit unions, within a strengthened framework of regulation and supervision. These institutions would then lend to susu collectors, cooperatives, savings associations, and other informal lenders. This sort of integration would permit institutions with surplus financial resources to link up with existing local institutions that lack resources but that have a comparative advantage in reaching the poor and other small borrowers cost-effectively.

Law and Development

Writing on the topic of legal systems and economic development, Robert D. Cooter notes that law supports the modern economy in several ways—for example, ensuring that firms will advance resources to others by lowering the risk of doing so, help-

ing firms create incentives for workers, and helping to ensure competition. Cooter argues that the most effective law in support of the economy is law that responds to the realities of business practice. When state laws respond to social norms, most people perceive the laws as just and tend to obey them out of respect, which creates a "rule-of-law state." But when state laws do not respond to social norms, people perceive the laws as unjust or irrelevant and are likely to disobey them or to obey them only out of fear of punishment. This is the "rule of state law." Like citizens, businesses are more likely to comply with the law when it corresponds to their norms. When law is responsive, both citizens and businesses are also more likely to help officials enforce the law by giving them information. And they are more likely to know what the law requires of them without needing costly legal advice. In an environment of open competition business norms tend to evolve toward efficiency. Law can be modernized by following what business does, rather than telling business what to do. Cooter argues that technical assistance to help countries develop legal infrastructure should avoid supporting top-down centralized planning. Instead, it should help to develop a community of judges and scholars who can shape law to the realities of business practice.

In another article on legal systems Avner Greif notes that neoclassical economists and social scientists in other disciplines make conflicting claims about the importance of formal legal systems to various forms of exchange. For neoclassical economists these systems are essential to anonymous exchange and economic growth. For economic sociologists, however, informal contract enforcement based on personal trust and reputation is more common and more effective in sustaining exchange. According to Greif, neither view is entirely accurate. Using game theory reasoning and empirical observations, he argues that there is no uniquely optimal combination of contract enforcement mechanisms. Exchange in anonymous markets governed by a legal system is not always more efficient than exchange based on reputation and trust. But that does not imply that exchange is always best organized within social structures and that these structures cannot be subjected to analysis. Instead, it is important to recognize the historical processes in which social, cultural, and economic forces interact to generate distinct contract enforcement mechanisms, each with possibly distinct economic efficiency. Thus a set of institutions may be efficient in one economic situation but not in another. And for a given economic situation the existing set of contract enforcement institutions may not be efficient. Greif concludes that policies aimed at enhancing contract enforcement must be based on the analysis of the interrelations between the economic, social, cultural, and political factors particular to that economy.

Standards, Trade, and Development

On the topic of labor standards and international trade, Alan B. Krueger notes that no one favors child labor but that disagreement mounts quickly when it comes to what to do about it and other nationally divergent labor standards. One side believes that a set of minimal standards is necessary to promote fair competition and to allow

labor markets to work efficiently. The other side considers international labor standards as either disguised protectionism or misplaced compassion. To determine whether international labor standards are simply protectionism in compassionate disguise, Krueger examined the factors influencing support for U.S. legislation that would ban imports of goods made with child labor. If support for the bill represented a concealed desire to protect constituents from foreign competition, support ought to be strongest among legislators representing districts with relatively many unskilled workers (and so more likely to compete with child labor). Krueger found no evidence that that is the case. That finding alone, however, says little about whether such a standard is economically efficient or desirable. To know that, we need to know whether standards are complied with and what their economic effects are. Data on compliance with compulsory schooling laws—a policy related to minimum-age labor laws—show that they are widely disregarded and have little effect on educational attainment in developing countries. Economic growth, by contrast, leads to a sharp reduction in child labor. If countries lack the capacity or will to enforce their labor standards, there is little value in the international community pressing for more stringent standards.

Analyzing environmental standards and international trade, Kym Anderson notes that over the past decade trade policy discussions have increasingly touched on concerns about natural resource use and the environment on the one hand and about the trade effects of environmental policies on the other. These issues are becoming more prominent as developing and industrial countries become more closely integrated in the context of the multilateral trade framework, especially the new World Trade Organization. Because industrial countries usually have higher environmental standards than do developing countries, many developing countries perceive the entwining of environmental and trade issues as a threat to their sovereignty and their economies. But links go both ways. The pressure to harmonize environmental standards is driven not just by the desire to reduce administrative and conformance costs but also by the perceived unfair competitive advantage of firms in countries with low standards over firms in countries with high standards. Moreover, many groups in industrial countries consider it ecologically unsound to trade with and invest in countries whose environmental standards are much lower than theirs. Anderson concludes that the direct effect on developing countries of a closer link between trade and environmental issues is likely to be small and may even be positive for some through improved terms of trade or compensatory transfer mechanisms. Still, industrial countries must take care to avoid inappropriate uses of trade measures to promote environmental objectives. Otherwise there is considerable risk of harm to developing and other economies through the erosion of the rules-based multilateral trading system.

* * *

As in previous years, the planning and organization of the 1996 conference was a joint effort. Particular thanks for their support are due to Gregory Ingram, the

administrator of the Research Advisory Staff; Paulo Vieira Da Cunha, economic adviser to the senior vice president, Development Economics, and chief economist; and Mark Baird, director of development policy. We would also like to thank other staff members, in particular several anonymous reviewers; Anupa Bhaumik and Clara Else for their assistance; and the conference coordinators, Evelyn Alfaro, Mantejwinder Jandu, and Jean Gray Ponchamni, whose excellent organizational skills kept the conference on track. Finally, we thank the editorial staff, especially Meta de Coquereaumount, Paul Holtz, and Patricia McNees.

Opening Remarks

The World Bank As a Global Information Clearinghouse

James D. Wolfensohn

Good morning and welcome to the eighth Annual World Bank Conference on Development Economics. It is a particular pleasure to open this conference and to find myself among some of the best minds in the business. These conferences, which began in 1989 with the purpose of increasing exchange between researchers and Bank staff, have become something more. They present and disseminate the cutting edge of thinking on some of the most important issues in development.

In the ten months since becoming president of the World Bank, I have talked a lot about improving results on the ground, of making the Bank quicker, more effective, more agile, and more responsive to its clients. This may sound like an agenda that devalues research and thinking, but I assure you that quite the reverse is true. If our goal is to improve results on behalf of developing countries, we need to expand and intensify thinking about development issues. In today' s rapidly changing global economy, the Bank plays an important role as a connecting rod between research and practice. We are a global clearinghouse that identifies and seeks out the best ideas, develops their practical applications, and gets them to end users in time to apply them to the real problems people face.

The Bank can hardly perform this role alone. That is why your contribution is so important. Our role in generating ideas requires complete openness and interchange between ourselves and our colleagues outside the organization. This annual conference is an important part of our effort to fulfill our global role. I hope that you will not consider it as a stand-alone event, but rather as one element of an effort that involves you and other top researchers more systematically in delivering the best thinking to the places it can do the most good. It is for this very reason, for instance, that we are expanding the role of the Economic Development Institute (EDI), whose mission includes identifying the best ideas, as well as the best practices, on development issues from around the world and making them immediately available to end users in developing countries.

James D. Wolfensohn is president of the World Bank.

Annual World Bank Conference on Development Economics 1996

The topics you will be considering over the next couple of days are all priority issues in developing countries:

Banking failures. As a banker myself in my former life, I very much appreciate the importance of research into ways to anticipate and fend off banking failures before they happen and to resolve them when they do. Financial system crises plague a great many of the Bank's clients, and helping them build the capacity to avoid them and to solve them is a top priority.

Reducing poverty. Poverty reduction is, of course, the World Bank's foremost objective. What our experience has shown, however, is that there is no simple formula for applying resources in a way that brings sustainable improvement to the lives of the 1.3 billion people in the world today who live in absolute poverty. We constantly need to question the assumptions and intellectual models that drive our thinking about how best to reach the poor, and to hear from researchers in the field about how we are doing.

Legal systems and economic development. The question of how legal systems relate to economic development is at the heart of another central issue in development, namely how local institutions either promote or detract from people's willingness to interact in market-based economies. Working with our clients to improve the legal and regulatory framework for market-led growth is also among our top priorities.

Labor and environmental standards in international trade. These are two issues that we tend to think involve complicated tradeoffs for industrial and developing countries alike. What research can do is make the tradeoffs less stark. Indeed, it can elucidate ways to trade successfully while developing adequate standards for labor and the environment.

As you deliberate these important issues over the next two days, I ask only that you bear in mind that there is a concrete end use for the product of your work and your discussions: making life better for the world's poor. I consider this event to be a laboratory for the ideas that will make a difference to the conditions in which the poor live. The World Bank's responsibility is to learn from you and to ensure the quickest translation possible from thought to action.

Keynote Address

The Role of Government in Economic Development

Joseph E. Stiglitz

It has been almost three decades since I began working on economic development issues. During that time the world has changed markedly—and so has the intellectual framework we use to approach development. Thirty years ago we felt both hope and concern. We believed that developing countries could close the substantial gap separating them from industrial countries, though we worried about why so few countries had actually done so. Standard textbooks discussed the leaps the Soviet Union had made between the mid-1920s and the onset of World War II. That supposed success—which now looks more apparent than real—influenced developing countries, many of which set up planning commissions to guide their economies. In many instances the state went well beyond guidance and into actual production and ownership of enterprises.

What a difference thirty years makes. The Republic of Korea—whose per capita income in 1960 was roughly the same as India's (less than $500 in 1995 dollars)—is now a member of the Organization for Economic Cooperation and Development (OECD). The successes of Korea and other East Asian economies demonstrate the effectiveness of a more market-based development strategy. In most instances East Asian governments abandoned the rigid planning model early on. But they did not err by going to the other extreme. Their governments helped to guide and create markets rather than completely supplanting or surrendering to them.

Meanwhile, the economies that stuck with the planning model experienced slow growth, stagnation, or worse; the collapse of the socialist economies was but the final nail in the planning coffin. By the 1980s countries around the world were actively engaged in privatizing state enterprises.

The dramatic failure of the grand socialist experiment had an unanticipated consequence: it lent support to extremists of the opposite ideological persuasion, according to whom government should play almost no economic role. But the rejection of one extreme is not the affirmation of the other. The real issue that both

Joseph E. Stiglitz is chairman of the U.S. Council of Economic Advisers. The interpretations expressed here are entirely those of the author and do not necessarily represent the views of any organization with which he is or has been affiliated.

Annual World Bank Conference on Development Economics 1996

the success of East Asia and the failure of the socialist experiment raise is, *what is the appropriate role of government in economic development?* There is a third way—or, I should say, many third ways—between the extremes of total government control of the economy and complete laissez-faire. At different stages of development or in different situations countries will and should choose different points along this spectrum.

I cannot possibly touch here on every aspect of the appropriate role of government. Instead I want to draw selectively on advances in economic theory, interpretations of the East Asian miracle, and my experiences within the U.S. government to highlight aspects that have not received sufficient attention in recent discussions. Certain topics (such as the role and design of industrial policies and the role of government in the financial sector) are omitted or treated only briefly because they have already been discussed widely elsewhere.

Before developing this framework, I want to make two preliminary remarks. The first relates to developments in economic theory. The perspective that I argue for here places markets at the center of the economy. The theoretical foundations for this market-oriented perspective rest on Adam Smith's notion of the invisible hand and especially its modern rendition, the fundamental theorems of welfare economics. To be sure, economists have long recognized the need for selective interventions in the marketplace to remedy well-identified problems such as externalities.[1] But developments over the past fifteen years have shown that well-designed government actions *can* improve living standards whenever there are imperfections of information or competition or incomplete markets—problems that arise in all economies, but especially in developing ones. The use of the word *can* here is crucial. Not every market "problem" calls for government action. In order to raise living standards, government actions must meet two criteria: they must address some serious imperfection in the marketplace, and they must be designed efficiently enough that their benefits outweigh their costs.

My second prefatory comment is that the key arguments in favor of a tightly circumscribed role for government have generally been shown to have only limited validity. Critics have asserted that government is unnecessary, ineffective, and to the extent that it actually affects anything, counterproductive. They argue that anything government can do the private sector can do better; that anything the government does will be offset by actions of the private sector; and that rather than improving resource allocations, government interventions actually make matters worse, especially because of rent seeking. The first proposition is simply false, while the second is true only under highly restrictive conditions. As for the third, the historical evidence shows that government actions can and have made a difference for the better; and both theory and evidence show that concerns about rent seeking, while real, are typically exaggerated. For rent seeking to completely dissipate the profits generated assumes perfect competition in rent seeking. The major thrust of these criticisms, however, is one that I have already noted and with which I fully agree: the fact that markets are not constrained Pareto-efficient does not imply that any arbitrary intervention will necessarily improve matters. The full consequences of any proposed action must be carefully assessed.

What Should Government Do?

Theory can provide valuable guidance on the appropriate role of government. In this context I believe that the East Asian experience and the experiences of the industrial countries are instructive. To be sure, there is the ever-present problem of the counterfactual: would these economies have grown even more quickly if government had not taken the actions it did? Although we may never know for sure, a wealth of evidence suggests the contrary. And I am convinced that while the United States also relied primarily on markets, its success was in part attributable to selective government actions. In both the United States and the rapidly growing economies of East Asia government has played six important roles that have spurred development.

Six Important Roles

In a sense much of the role of government can be viewed as establishing infrastructure in its broadest sense—the educational, technological, financial, physical, environmental, and social infrastructure of the economy. Since markets cannot operate in a vacuum, this infrastructure is necessary if markets are to fulfill their central role in increasing wealth and living standards. Because constructing the broad infrastructure is beyond the capacity or interest of any single firm, it must be primarily the responsibility of government.

The first role of government that the United States and the East Asian economies have in common is in *promoting education.* Even before the adoption of the U.S. Constitution the federal government of the United States, in the Northwest Ordinances of 1785 and 1787, recognized its responsibility for promoting public education by setting aside land in the newly formed states for that purpose. Later, in 1863, the federal government helped establish the public university system.

The East Asian economies also emphasized the role of government in providing universal education, which was a necessary part of their transformation from agrarian to rapidly industrializing economies. Universal education also created a more egalitarian society in East Asia, facilitating the political stability that is a precondition for successful long-term economic development. In pursuing such egalitarian policies, the economies of East Asia laid to rest the trickle-down theories of development. Simon Kuznets had argued that economic growth was associated with an increase in inequality; Arthur Lewis had suggested that such inequality was necessary because capital accumulation lay at the heart of growth. Since rich individuals were assumed to save more at the margin than poor ones, higher levels of inequality would increase savings and hence growth. The East Asian economies showed that high levels of savings could be attained in an egalitarian setting and that human capital accumulation was every bit as important as—if not more important than—increases in physical capital.

The second role of government is in *promoting technology.* In 1789 the U.S. Constitution recognized the importance of science and technology by giving

Congress the right to grant patents to promote the progress of science. Even in the early part of the nineteenth century support for research went well beyond the establishment of a system of intellectual property. Just as the modern telecommunications system—including the Internet—was fostered by government, so too were earlier advances. In 1842, for example, the federal government financed the world's first telegraph line, between Baltimore and Washington. Over the more than 150 years during which it has supported research, the U.S. government has had an impressive record of successes. In the nineteenth century agriculture was the mainstay of the economy, accounting for more than 35 percent of GDP in the 1870s. The remarkable productivity growth in the agriculture sector is largely attributable to the federal government's support for research and dissemination of its results. East Asian governments have also played a central role in the promotion and transfer of technology.

The third role of government is in *supporting the financial sector.* Sometimes depicted as the "brain" of the economy, the financial sector is responsible for deploying scarce capital resources in the most efficient way. It is concerned with gathering, processing, and disseminating information—precisely the areas in which market failures are often most marked. In 1863, in the midst of the Civil War, the United States recognized the need to create a national financial system and passed the National Banking Act, establishing the world's first supervisory banking agency. Although we now know that far more is needed to create financial stability, this legislation did much to reduce the financial instability that had characterized the economy up to that time. In later years the government created the Federal Reserve system as well as a series of financial intermediaries to spur markets that had been thin or nonexistent. Similarly, East Asian governments took an active role in ensuring the safety and soundness of financial institutions and in creating new institutions and markets to fill gaps in the private sector.

The fourth role of government is in *investing in infrastructure,* including institutions as well as roads and communications systems. In both the United States and the successful East Asian economies, governments created institutional infrastructures within which competitive markets could thrive. Only recently, as the formerly socialist economies have struggled to establish market economies, have we become fully aware of the importance of this institutional infrastructure, which includes property rights, contract and bankruptcy laws, and policies to promote competition where it is viable and to regulate markets where it is not.

The fifth role of government is in *preventing environmental degradation.* Although economists have discussed the need for government action to correct market failures at least since Edgeworth, the concept has become widely accepted only during the past quarter century. Good environmental policies should not be viewed as luxuries, to be enjoyed only by the well-off. We should not confuse increases in GDP with increases in standards of living, or increases in measured GDP today with increases in long-term wealth. Recent attempts at building "green" GDP accounts recognize these points. They highlight the enormous challenges faced by countries that have not prevented environmental degradation: it will take generations to undo

the environmental damage that has been wrought in many developing and transition economies.[2]

The sixth role of government is in *creating and maintaining a social safety net,* including access to basic health services. In some cases these activities can be justified in utilitarian terms: they increase the productivity of the labor force and foster political stability by reducing opposition to change. But they may also be justified in terms of basic values. As I noted above, standards of living embrace more than the variables captured in GDP statistics. There is a fundamental sense in which, for instance, improved health conditions represent an improvement in living standards, even if such an improvement is not reflected in GDP.

Virtually all societies have provided social safety nets, albeit not always through the government. For at least two reasons governments today may have to assume a larger role in providing safety nets than either the U.S. or East Asian governments did at comparable stages in their countries' development. First, the pressures of urbanization call for a stronger government role. In 1975 just over a third of the world's population lived in cities; the United Nations and the World Bank estimate that by 2025 that proportion will double. Urbanization—and the migration from traditional communities with which it is associated—is likely to result in less effective community-based social safety nets. Second, in the transition economies large firms traditionally provided much of the social safety net (such as pensions and health care). The transformation of these economies is being accompanied by the shedding of these social responsibilities by corporations facing new competitive pressures. The government is the only backstop.

Special Role of Government in Developing Economies

Government needs to function in the six capacities I have delineated in all types of economies. But I want to take note of the special problems facing developing and transition economies, in which more markets are lacking, the markets that do exist may function less effectively, and information problems are more severe than in industrial countries simply because of the rapid change in the economic environment.[3]

While market failures loom larger in developing and transition economies, the capacity of the government to correct these market failures is often weaker.[4] Assessing the appropriate role of government requires recognition of both the need for and the limitations of government action. Successful governments have helped create markets (such as bond and stock markets and long-term credit institutions). They have established and enforced laws and regulations that have made financial markets more stable and increased competition in all sectors. In many cases governments have acted as surrogate entrepreneurs, encouraging the establishment of firms to enter certain markets. Especially in export markets, governments have provided firms with strong incentives. (Some econometric evidence suggests that many of these interventions were quite effective. For example, an analysis of the mild financial restraint evidenced in most East Asian economies suggests that it did lead to more rapid economic growth.)

The Conservative Reply

Most economists today accept the proposition that markets alone may not succeed in ensuring economic efficiency and may fail to protect some segments of society from abject poverty. While most economists also agree that such shortcomings might provide a rationale for government action in principle, some argue that government interventions all too often have been counterproductive in practice. Any balanced account of the role of government must acknowledge that this has often been the case, a topic I touch on below. But that in itself proves nothing: the question is, can responsible democratic governments put in place policies that raise living standards? Based on the experiences of East Asia and the United States, I believe the answer is a resounding yes.

Some critics of the role of government argue for a different perspective on the East Asian experience. They contend that all—or almost all—of the growth of the East Asian economies can be accounted for by factor accumulation. Thus, they argue, there is no miracle but simply the inexorable working out of standard fundamentals: increased inputs lead to increased outputs. Total factor productivity growth has been negligible.

There are several technical problems associated with the studies reporting these results. (Does anyone who has studied wage setting in Singapore, for example, really believe that wages are set in a competitive process, so that the real wage equals the marginal product of labor, as most of the studies assume?) But even if we take at face value the findings of low total factor productivity growth, these studies do not really address the question of whether government policies made a difference. They neither ask nor answer questions such as these:

- Why were savings rates in East Asia so high? Elsewhere comparable savings rates had been attained only under the compulsion of strong government force, as in the communist countries. Although econometric studies suggest that East Asia's savings rates may be partly explained by traditional economic factors, government actions also played a constructive role.
- Why were the East Asian economies able to invest efficiently at such a rapid pace? Other countries (the former Soviet Union is the classic example) have invested heavily but ended up with high incremental capital-output ratios rather than rapid growth. Government efforts to create effective financial institutions combined with the practice of providing funds to firms that proved their mettle in the competitive export markets surely contributed to the efficiency with which East Asia's capital was allocated.
- How were the East Asian economies able to reduce the technological gap between themselves and the most economically advanced countries so quickly? The East Asian economies demonstrated an enormous capacity to absorb both capital and technology. The speed with which they closed that gap entailed more than just buying technology. Governments played a major role in investing in human capital, allowing foreign investment (with some exceptions), and creating an economic atmosphere conducive to foreign investment.

- How did the East Asian economies ensure that the benefits of rapid growth were shared broadly across the population? As already noted, the increases in inequality that earlier experiences had suggested inevitably accompany development simply did not occur in East Asia. To the contrary, there are reasons to believe that greater egalitarianism—a result of deliberate government policies—actually contributed to the remarkable growth in these economies.

Improving Government Performance and Responding to Change

I have argued above that people who advocate a tightly circumscribed role for the government—providing only for national defense, for example—go too far. But I want to stress again that government is not infallible. Even in the successful East Asian economies governments have made mistakes. (The Japanese government, for example, initially attempted to prevent Honda from entering the automobile industry.) Government cannot fix every problem. Government definitely has a place, but it must know its place.

The pragmatic framework that I set forth earlier for assessing the role of government entails a balancing of the strengths and limitations of markets and government, and the determination of how they can best complement each other. It does not begin with two columns labeled "activities to be carried out by government" and "activities to be carried out by the private sector." This careful balancing puts greater emphasis on how the government does what it does and how it interacts with the private sector. To that effect, I want to outline a few general principles, motivated by both theoretical analyses and historical experience.

Incentives and Change

In assessing the proper role of government, we must take into account two fundamental points. The first is the importance of incentives. The second is the dynamic nature of government's role; as the economy changes, so must the government.

The government is a large organization, but unlike large market organizations it is not subject to the pressures of market competition. In democracies political competition exercises some discipline: incompetence is punished and performance is rewarded. To be sure, political competition is a far cry from the textbook ideal of perfect competition. But so too is actual market competition. It is sometimes suggested that bureaucrats lack incentives, but incentives can also be distorted in large corporations. It has also been suggested that bureaucrats are not responsive to the wishes of voters. But both theory and evidence suggest that managers of many large corporations are not always responsive to the wishes of the shareholders to whom they are in principle accountable.

In short, the distinctions between the public and private sectors are often overblown. But we must be careful not to go to the other extreme: incentives do play a somewhat more important role in the private than in the public sector. Provided adequate competition policies are put in place, market competition is more

effective in providing incentives than is ersatz public competition. The question is whether, and how, the public sector can put in place an effective set of incentives.

We must also recognize that the role of government is not static. Changes in the economic environment fundamentally alter what the government can and should do. In a world with limited international trade, for example, it might have made sense for countries to worry about material balances, and there might have been some rationale for the kinds of planning exercises that once dominated development thinking.

But with the expansion in international trade and the fall in transportation costs, countries can now specialize in that part of the production process in which they have a comparative advantage; they are not limited to domestic markets on either the demand or the supply side. Consider automobile production. Assembly is only one part of a car's cost, representing only about a fourth of value added at the factory. Different parts can be constructed in different countries and then shipped to the assembly point. Modern telecommunications systems ensure that parts orders can be transmitted quickly from the assembly plant to the parts plant, wherever it is located.

In the past ten years this pattern has spread from large multinationals to much smaller companies. As a result of improved transportation and telecommunications a small or medium-size firm in the United States or Europe can develop relationships with suppliers in East or South Asia, sending them precise product specifications. While the long-run implications are not yet clear, these developments have been a boon for developing economies. The globalization of entrepreneurship has loosened the constraint on growth imposed by one of the scarcest of factors in the developing world.

Globalization is just one example of a change in the economic structure that necessitates a change in government policies. Below I discuss other examples, including how changes in technology have enhanced the scope for competition in areas that were once considered natural monopolies (telecommunications and electricity.)

Recognizing the importance of incentives and the continually changing role of government, we can now consider ways in which government performance can be improved: increasing consumer orientation, monitoring and rewarding performance, extending the scope for competition, corporatizing and privatizing, and improving regulatory policy.

Increasing Consumer Orientation

One of the problems arising from a lack of competition is a lack of choice. Consumers do not get to choose the providers that issue their driver's license or passport. When there is a choice, as in buying airline tickets, some consumers will choose to go to airlines with shorter lines, even if they have to pay a slightly higher price; the market reflects the diversity of consumer preferences. One way to address the problem of choice within the public sector is to create more competition; short

of that, government can create a culture of customer orientation. Performance measures may be effective in drawing organizational attention to relevant variables and perhaps in motivating the behavior of individual bureaucrats. Changes in attitude toward users of government services—thinking and referring to service users as customers, for example—may also help.

In some instances performance can be measured and monitored. The length of time it takes a service representative to respond to a customer is an example. Standard survey techniques can evaluate customer satisfaction with, for example, telephone interactions. Indeed, at the individual level the motivational and monitoring problems facing, say, the Social Security Administration are little different from those facing a private insurance company. In the United States, where government agencies have worked hard over the past three years to improve customer orientation, we have shown that government can in fact succeed in improving service: ratings of the Social Security Administration's services are highly competitive with the best comparable service providers in the private sector.

Monitoring and Rewarding Performance

Private firms have simple bottom-line performance measures—profits and market value. Although government as a whole has no comparable summary statistic for capturing performance, performance at particular activities (typing letters, issuing visas, processing driver's licenses) can be identified and measured. It is important that output, not process or input, measures be used: too often rewards are based on how well a worker complies with standard operating procedures.

In many cases activities performed by the public sector are similar enough to activities performed by the private sector that private sector performance can provide a yardstick. For instance, while every firm has slightly different travel needs, it is possible to obtain a range of estimates of the administrative costs associated with travel. These costs can then be used to determine how government agencies compare and to use performance relative to private sector norms as a basis for rewards.

But many public sector activities are different from activities conducted within the private sector. Public sector activities are disproportionately administrative in nature, making measurement of individual performance particularly difficult. We do not know how to measure the quality of many administrative decisions made collectively, let alone the contributions of individuals. In many other activities there is no single metric of performance. Consider education. Rewarding performance only in terms of basic skills (which can be measured more easily than other skills) will divert resources away from developing higher-order and cognitive skills. It may, however, be possible to redesign the production process to mitigate these effects (by assigning different teachers different tasks, for example). Whether this is desirable depends on the magnitude of the incentive distortions relative to the economies of scope. If economies of scope are weak (for example, between teaching basic skills and teaching higher-order cognitive skills), then it may be desirable to redesign the production process so that individual performance can more easily be monitored.

Extending the Scope for Competition

One way to provide more effective incentives, including enhanced consumer orientation, is to extend the scope for competition. Creating effective competition among vendors, for example, is an essential step in ensuring that the government procures goods and services at the lowest possible price. But the task of competitive procurement is more difficult than is often realized. It used to be thought that competitive bidding was the simplest way to ensure that government does not pay too much for a good or service. Competitive bidding, however, typically requires the government to draft precise specifications for the item being purchased—describing a simple t-shirt could take thirty small-print pages. Since most firms do not normally produce to those precise specifications, they may find it unattractive to bid even if their products have similar performance characteristics. Thus the number of bidders is often relatively small. As a result the government may have to pay higher prices than the public at large.

In a sense the problem arises because of the difficulty of developing and clearly articulating performance measures. Government uses competitive bidding to ensure that taxpayers do not overpay and to mute criticism that the government has wasted taxpayers' money. Yet these cumbersome procurement policies—a form of micromanagement of contractor production—have often contributed to higher average costs. Where a competitive marketplace exists, the discipline of competition in the marketplace may suffice. In the United States, for example, procurement reforms based on this principle enacted under the Clinton administration will save U.S. taxpayers $12.3 billion over the next five years.

Corporatizing and Privatizing

Even when competition is not a viable option, it may be desirable to incorporate features of a private firm. This objective goes beyond introducing performance pay, extending to broader issues of personnel, procurement, and budgeting. When competition is not viable, however, the danger of abuse of monopoly power is real. To constrain the abuse of power, policymakers need to ask three key questions:

- Is there a dedicated source of revenue related to the benefits conferred?
- Is there a governance structure that can ensure efficiency and a regulatory structure that can protect against abuses of monopoly power?
- Can production issues be separated from other public policy issues, including those related to externalities and safety, for example?

Privatization represents only one point along a spectrum of organizational forms that includes a variety of corporatization structures within the public sector. Sappington and Stiglitz's (1987) Fundamental Theorem on Privatization established that the conditions under which privatization could fully achieve the public objectives of equity and efficiency were extremely restrictive—and similar to the conditions under which competitive markets attain Pareto-efficient outcomes. Because of differences in risk aversion and time discounting, the state may receive less—

possibly far less—than the expected present discounted value of the profits of the enterprise. Moreover, even with a complicated set of Pigouvian taxes the state may not be able to induce the privatized industry to act in the way it would like, especially when there are complicated social objectives.[5]

The theorem's main thrust is that privatization has to be justified on a case-by-case basis: the increase in economic efficiency must be sufficient to outweigh the disadvantages of privatization. In many cases (as in telecommunications) that case clearly has been established.

When privatization has been determined to be desirable, it must be implemented correctly, with appropriate built-in protections—including protections against abuse of any monopoly power. Appropriately designed competitive auctions typically are the most effective way of ensuring that the public obtains full value for publicly owned resources. The carefully structured spectrum auctions in the United States illustrate how to raise public revenues while promoting competitive markets and innovation.

Corporatization, privatization, and the other reforms discussed earlier help to focus attention on performance—on outputs rather than on inputs and process. This focus is necessary if the efficiency of the public sector is to be enhanced. In areas where privatization may be inappropriate (such as the granting of patents), government functions can be organized to focus on performance, as the United States has done in establishing performance-based organizations that establish organizational and individual incentives to enhance performance.

Improving Regulatory Policy

A focus on performance is also critical to ensure that regulations achieve their objectives at minimum cost. In many countries the environmental regulations of the past two decades have brought about enormous improvements in air and water quality. In some cases, however, the objectives could have been obtained at lower cost. Rather than focusing on performance criteria, policymakers imposed design standards. In some cases such standards were imposed because there was no effective way to monitor performance. But appropriately designed regulations could have provided incentives for the development of monitoring technology.

Nowhere is the changing role of government and the increasing reliance on market-based regulatory policies more evident than in the telecommunications and electricity industries. We used to think of these industries as natural monopolies, where governments faced a choice between nationalization and regulation, with most governments choosing nationalization. But as the inefficiencies of state-owned enterprises became apparent, more and more countries privatized their telecommunications systems, creating a monopoly that was often subject only to weak regulation.

Few governments took the next step of asking how competition could be ensured. They did not do so because economists told them that competition was not a viable option, that these industries were natural monopolies. But on closer observation, we have realized that competition is indeed viable in many, if not most, parts of these

industries. We are therefore left with a subtle question: how do we ensure that monopolies in those segments in which competition is not a viable option do not destroy competition in other segments (through discriminatory access or pricing, for example)?

In the case of the U.S. telecommunications industry, it became clear that regulation alone could not effectively prevent discrimination; structural separation of the "last mile" (a natural monopoly) and other parts of the telecommunications system was thus required. With the appropriate institutional infrastructure, competition could be made viable in large segments of this vast market, allowing government regulators to focus on a much narrower set of issues.

The same process is now taking place within the electricity industry, another sector traditionally regarded as a natural monopoly. We now recognize that this industry has at least three major components: generation, transmission, and marketing. Already, changes in technology have made possible a competitive market in generation. And in the United States, where an effective procompetition regulatory structure is being put in place, a competitive market in electricity generation is rapidly emerging. Telecommunications and electricity represent two areas for which the role of government in most countries has changed dramatically, from protection or detailed regulation of the entire industry to regulation only to ensure that the parts of industry where competition is not viable do not abuse their market power.

Conclusion

It has become almost a cliché to refer to the vast changes in our world and the need to adapt to those changes. Yet the fact remains that extraordinary changes have taken place and that societies that adapt better to those changes will do better, in terms of raising living standards, than those that do not. Government can help societies embrace change.

Both the constants and the changes in development practice and theory are remarkable. So too is the similarity of arenas of activity between countries that developed successfully in the nineteenth century and the East Asian economies that developed largely in the second half of the twentieth century. (One difference is that the earlier development experience lacked the benefit of insights from modern economics!) Among the constants are putting competitive markets at the center of an economy, with governments acting to assist, use, and supplement those markets; providing public investment in education and technology; and constructing appropriate institutional infrastructures, including those that support dynamic and competitive telecommunications and financial sectors. Governments must also provide a basic safety net and protect the environment. Among the changes is the recognition that government can make use of many of the mechanisms that have helped make markets work so effectively and that the scope for competition is broader than had previously been thought.

Making government perform better is an important concern everywhere. Good policies on education, health, and the environment are not luxuries to be postponed

to a later date. Making government focus more on customer orientation, performance, and competition is also essential. Indeed, the scarcity of resources and the tightness of fiscal constraints facing developing countries today make it imperative that resources be spent efficiently.

Too often discussions of what the government should do present false dichotomies. Good environmental and educational policies can actually enhance economic growth. Yet it is also true that only if developing countries grow more rapidly will they be able to provide their citizens with a decent standard of living. Development and improved living standards have many dimensions, but both ultimately depend on increased production of goods and services. It is right that we redress an imbalance that saw this increased production as an end in itself. But having refocused our attention on the right set of objectives, we should not lose sight of the means by which those objectives must be attained.

The theories and historical experience to which I have alluded here may guide us in shaping the role of government. Leadership can help articulate visions of that role. But in the end it is the desires—real and perceived—of the people whom government is supposed to serve that will determine both the scope of government and its ability to be a positive and creative force.

Notes

1. Although some economists have suggested that even those difficulties could be handled privately through Coasian bargaining, most economists believe that market interventions, in the form of Pigouvian taxes, are the appropriate remedy.
2. In many cases even the short-run costs of unsound environmental policies (such as health care costs) can be high.
3. Recent literature has emphasized the importance of reputation mechanisms and implicit contracts in governing economic relations. The effectiveness of these mechanisms depends on the long-term nature of the relationship. In developing countries rapid transition threatens the long-term viability of many such relationships, the weakening of social ties reduces the role of social sanctions as an enforcement mechanism, and high interest rates encourage short-run self-interested behavior at the expense of long-run cooperative behavior.
4. I noted earlier the importance of establishing the appropriate institutional infrastructure for a market economy in the transition economies. But it may take a strong government to establish the institutional infrastructure that facilitates the viability of a strong market.
5. A Pigouvian tax, levied on producers of externalities, raises the producer's perceived private costs to the level of the social costs of the activity.

Reference

Sappington, David E. M., and Joseph E. Stiglitz. 1987. "Privatization, Information, and Incentives." *Journal of Policy Analysis and Management* 6 (4): 567–82.

Floor Discussion of "The Role of Government in Economic Development," by Joseph E. Stiglitz

Michael Bruno (chair), noting Joseph Stiglitz's comment that his view of the government's role in the economy had not changed since he had joined the government, asked whether Stiglitz would have said the same thing had he started working for a government in a developing country. On a spectrum ranging from the government doing everything to the private sector doing everything, Bruno wondered whether the quality of the people staffing the public and private sectors would alter Stiglitz's views (assuming that the United States represents the far edge in terms of quality).

Stiglitz acknowledged that market failures—absent or ineffective markets—are more prevalent in developing countries than in industrial countries, adding that the public sector is also less effective in most developing countries. Thus a balance must be struck. East Asian economies provide a good example of how governments can play a positive role and make a difference. Thirty years ago, for example, per capita incomes in the Republic of Korea were comparable to those in India. Although it is important not to generalize, many East Asian economies engaged in practices that increased the likelihood of effective government and decreased the likelihood of corruption and other abuses. Competitive procurement practices are one way a government can use market mechanisms to competitively allocate resources and improve performance in the public sector.

A participant from the World Bank's International Trade Division asked about competition between the public and private sectors and about how the two should compete with and complement one another. Stiglitz had suggested that replacing a state marketing board with a private marketing monopoly might not be an improvement if the private operation exploited farmers even more than the state operation had. Why not let the private sector compete with the state marketing board, asked the participant. Over time, even if the state board is efficient, the private sector could replace it. The participant also wondered about the relationship between government action and market failures. There were stories, for example, of pan-territorial

This session was chaired by Michael Bruno, senior vice president and chief economist, Development Economics, at the World Bank.

Annual World Bank Conference on Development Economics 1996

pricing in some Sub-Saharan African countries. No private sector transport or storage capabilities developed because, with prices fixed, there was no incentive to develop such facilities. Given the government's limited capacity, might it not be better to focus on creating the institutions that allow the private sector to flourish?

A good question, said Stiglitz. There is now scope for competition where previously none was thought to exist. But it is not so much a question of whether such competition is viable. There is no natural monopoly in marketing grain, for example, so there is no compelling argument for a government monopoly. If the concern is that private firms will exploit farmers, the appropriate mechanism is policies to ensure that competition thrives. That a market does not exist does not mean that it cannot develop. But governments must provide the institutional infrastructure that makes competitive markets work. In many areas the government has demonstrated a role for markets and markets have followed, although in some cases the government continued to play a dominant role. In the United States, for example, the government-established Fannie Mae (Federal National Mortgage Association) has played a national role in providing a market for securitized mortgages, which has lowered the cost of mortgages for millions of homeowners. Although markets could have done this, they did not, so the government created Fannie Mae, which is now a viable enterprise. Interestingly, the government had tried to create more than one agency to engage in this process, but Fannie Mae remained the dominant player in the market.

A participant from India asked Stiglitz if he thought that the private sector must take over the duties of the public sector in developing countries or if there is scope for making public sectors more efficient in terms of professionalism, continuity in management, technology, and so on. Stiglitz said he would have rephrased the question to ask which is more important: making the public sector more efficient or redefining its role? Both are important. They are not alternatives, and countries should not have to choose one or the other. In fact, the appropriate scope of government depends on how effectively it increases efficiency in the public sector. There are a number of ways that that can be done, and many of them involve emulating practices in the private sector. In fact, in many essentially public activities, there is a continuum that runs from fully private to fully public. In developing new strategies it is important to avoid pigeonholing activities in either of the two extremes. Many countries have chosen to privatize air traffic control, for example, but some have chosen to corporatize it. The patent system is another example. The United States and the United Kingdom are both redefining their patent systems using performance-based organizations.

A small business owner who had recently struggled with a maze of government bidding procedures asked Stiglitz if he had any recommendations for how governments could improve their functions, considering the shortage of bidders from the private sector. Stiglitz agreed that governments tend to produce too many bureaucratic forms. The United States recently has made efforts to simplify procedures. People are increasingly aware that bureaucratic procedures disproportionately harm small businesses, which cannot afford a full-time employee to handle paperwork the way large corporations can. The U.S. Small Business Administration, for example,

used to require so much paperwork that it was difficult for small businesses to apply for loans. But things have changed in the past few years: the agency's new, much shorter application form asks only for essential information. Pension procedures for small U.S. businesses have similarly been simplified. Now small businesses with pension plans have to file only a brief form, and it costs less than $10 a year per pension recipient to operate the system.

Another participant from the World Bank asked if the state could play any role in addressing the social and economic problems associated with ethnicity, which countries in Africa, Eastern Europe, and Central Asia have been struggling with. Stiglitz said he believes that the government has an important role to play in ensuring that all groups have access to opportunity, and a particular responsibility to ensure it for groups that have been denied such access in the past. Some countries, he noted, have done well in this role.

Banking Failures: Crisis or Opportunity for Reform?

Understanding Financial Crises: A Developing Country Perspective

Frederic S. Mishkin

How does a developing economy shift dramatically from a path of reasonable growth before a financial crisis, as was the case in Mexico in 1994, to a sharp decline in economic activity after a crisis? This article explains this puzzle by outlining an asymmetric information framework for analyzing banking and financial crises. The framework shows why the banking sector is so important to the economy, particularly in developing countries, and provides a rationale for bank regulation and supervision. The framework is also used to show why banking and financial crises occur and why they can have such a devastating effect on the economy. An important policy implication is that an appropriate institutional structure is critical for preventing banking and financial crises in developing countries and for reducing their undesirable effects should they occur.

Financial crises and their subset, banking crises, have become worldwide phenomena in recent years. Although financial and banking crises have been costly to the economies of industrial countries such as Japan, the Nordic countries, and the United States, the damage that these crises impose on developing countries seems to be far greater. An important puzzle for the study of development economics is how a developing economy can shift dramatically from a path of reasonable growth before a financial crisis, as was the case in Mexico in 1994, to a sharp decline in economic activity after a crisis occurs that is very damaging to both the economy and social fabric of the country.

Using an asymmetric information framework for analyzing banking and financial crises, I show why the banking sector is so important to the economy—especially in

Frederic S. Mishkin is executive vice president and director of research at the Federal Reserve Bank of New York, A. Barton Hepburn Professor of Economics at the Graduate School of Business at Columbia University, and research associate at the National Bureau of Economic Research. The author thanks Agustin Carstens and the staff at the Bank of Mexico, Terry Checki, Marilyn Skiles, and participants in seminars at the Federal Reserve Bank of New York, the International Monetary Fund, the Inter-American Development Bank, and the World Bank for helpful comments and Martina Heyd for research assistance. The interpretations expressed here are entirely those of the author and do not necessarily represent the views of any of the organizations with which he is affiliated.

Annual World Bank Conference on Development Economics 1996

developing countries—and derive a rationale for bank regulation and supervision. An important theme of this analysis is that an appropriate institutional structure is critical for preventing banking and financial crises in developing countries and for reducing their undesirable effects should they occur.

Asymmetric Information and the Financial System

The financial system enables funds to move from economic agents who lack productive investment opportunities to those who have such opportunities. Unless the financial system does this job effectively, the economy will not function efficiently and economic growth will be severely hampered.

A crucial impediment to the efficient functioning of the financial system is asymmetric information, a situation in which one party to a financial contract has much less accurate information than the other party. For example, a borrower who takes out a loan usually has much better information about the potential returns and risk associated with the investment projects the loan will finance than the lender does. Asymmetric information leads to two basic problems in the financial system: adverse selection and moral hazard.

Adverse Selection

Adverse selection is an asymmetric information problem that occurs before the transaction. This problem exists when the parties who are the most likely to produce an undesirable (*adverse*) outcome are the most likely to be *selected* for a loan. Borrowers who want to take on big risks are likely to be the most eager to take out a loan because they know that they are unlikely to pay it back. Since adverse selection makes it more likely that loans might be made to bad credit risks, lenders may decide not to make any loans, even though there are good credit risks in the marketplace.

This outcome is a feature of the classic "lemons problem" analysis first described by Akerlof (1970). As pointed out by Myers and Majluf (1984) and Greenwald, Stiglitz, and Weiss (1984), a lemons problem occurs in debt and equity markets when lenders have trouble determining whether a borrower is a good risk (has good investment opportunities with low risk) or a bad risk (has poor investment opportunities with high risk). In this situation a lender will only be willing to pay a price for securities that reflects the average quality of the firms issuing the securities—a price below fair market value (the net present value of the expected income streams) for high-quality firms but above fair market value for low-quality firms. Owners or managers of high-quality firms who know their quality also know that their securities are undervalued and will not want to sell them in the market. Low-quality firms, however, are willing sellers because they know that the price of their securities is greater than their value. Since asymmetric information prevents investors from determining the quality of firms, high-quality firms will issue few securities and credit markets will not work as well since many projects with a positive net present value will not be undertaken.

Moral Hazard

Moral hazard is an asymmetric information problem that occurs after the transaction. The lender is subjected to the *hazard* that the borrower has incentives to engage in activities that are undesirable (*immoral*) from the lender's point of view because they make it less likely that the loan will be paid back. Moral hazard occurs because the borrower has incentives to invest in high-risk projects in which the borrower does well if the project succeeds but the lender bears most of the loss if the project fails. The borrower also has incentives to misallocate funds for personal use, to shirk rather than work hard, or to undertake investment in unprofitable projects that increase the borrower's power or stature. The conflict of interest between borrower and lender stemming from moral hazard (the agency problem) implies that many lenders will decide that they would rather not make loans, so that lending and investment will be at suboptimal levels. (Asymmetric information is not the only source of moral hazard. Moral hazard can also occur if high enforcement costs make it too costly for the lender to prevent moral hazard even when the lender is fully informed about the borrower's activities.)

Why Are Banks So Important?

Asymmetric information theory explains why banks are such important players in the financial system.[1] Securities markets play a much smaller role, largely because of the adverse selection (lemons) problem—low-quality firms are more eager to issue securities than high-quality firms. Although the private production and sale of information could reduce the level of asymmetric information that creates the lemons problem, there is generally too little private production and sale of information because of the free-rider problem—people who do not pay for information can take advantage of the information that other people have paid for.

To understand the free-rider problem recognize that if some investors acquire information that tells them which securities are undervalued and therefore buy those securities, other investors who have not paid for this information may be able to buy the securities right along with the well-informed investors. If enough free-riding investors can do this, the increased demand for the undervalued securities will bid up their low price to reflect the securities' full net present value given this information. As a result of all these free riders, investors who have acquired information will no longer be able to earn the entire increase in the value of the security arising from this additional information. The weakened ability of private firms to profit from producing information will mean that less information is produced in securities markets, so that the adverse selection problem continues to be an impediment to a well-functioning securities market.

More important, the free-rider problem makes it less likely that securities markets will act to reduce incentives to engage in moral hazard. Monitoring and enforcement of restrictive covenants (provisions in debt contracts that restrict and specify certain activities of the borrower) are necessary to reduce moral hazard. By

monitoring borrowers' activities to ensure that borrowers are complying with the restrictive covenants, lenders can prevent borrowers from taking on risk at the lenders' expense. But monitoring and enforcement of restrictive covenants are costly activities. The free-rider problem discourages this kind of activity in securities markets because if some investors know that other securities holders are monitoring and enforcing the restrictive covenants, they can free ride on the other securities holders' monitoring and enforcement. Once these other securities holders realize that they can do the same, they also may stop their monitoring and enforcement activities. The result is that not enough resources are devoted to monitoring and enforcement, and moral hazard continues to be a severe problem for marketable securities.

This analysis explains why, as Mayer (1990) points out, securities markets are frequently a relatively unimportant source of external finance to nonfinancial businesses in industrial countries and an even less important source in developing countries. Clearly, the better is the quality of information about firms, the more likely it is that they can issue securities to raise funds. This implication of the theory explains why only the largest and best-known firms in industrial countries issue securities. Information about private firms is even harder to collect in developing countries and, not surprisingly, securities markets play a much smaller role.

Banks play the most important role in financial systems throughout the world because they are so well suited to reducing adverse selection and moral hazard problems in financial markets. They are not as subject to the free-rider problem and can profit from the information they produce by making private loans that are not traded. Other investors have more difficulty free riding off the financial intermediary and bidding up the loan's price.[2] Similarly, it is hard to free ride off banks' monitoring activities when they make private loans. Banks thus receive the benefits of monitoring and so are better equipped to prevent moral hazard on the part of borrowers.

Banks have particular advantages over other financial intermediaries in solving asymmetric information problems because of their natural advantages in collecting information and reducing moral hazard. For example, banks' advantages in collecting information are enhanced by their ability to enter into long-term relationships with customers and issue loans using lines of credit. In addition, their ability to scrutinize the checking account balances of their borrowers provides them with an additional advantage in monitoring behavior. Banks also have advantages in reducing moral hazard because, as Diamond (1984) demonstrates, they can engage in lower-cost monitoring than individuals and because, as Stiglitz and Weiss (1983) point out, they have advantages in preventing risk taking by borrowers since they can use the threat of cutting off future lending. The greater difficulty of acquiring information on private firms in developing countries makes banks even more important in the financial systems of these countries (a fact documented by Rojas-Suarez and Weisbrod 1994).[3]

Asymmetric Information and Bank Regulation

The analysis of how asymmetric information, through adverse selection and moral hazard problems, affects the structure of the financial system is also useful in under-

standing the common forms that bank regulation and supervision take in most countries. These include a safety net for depositors, restrictions on bank asset holdings, capital requirements, chartering and bank examinations, disclosure requirements, and prompt corrective action.

Government Safety Net

Although banks are well suited to solving adverse selection and moral hazard problems because they make private loans, this solution to the free-rider problem creates another asymmetric information problem. Depositors lack information about the quality of these private loans, a situation that can lead to bank panics (Calomiris and Gorton 1991).

To understand the problem, consider a situation in which there is no safety net for depositors. Suppose a shock to the economy causes such large loan losses in 5 percent of a country's banks that they become insolvent. Because of asymmetric information depositors are unable to tell whether their bank is a good bank or one of the 5 percent of insolvent banks. Depositors at bad and good banks alike recognize that they may not get back 100 cents on the dollar for their deposits and will want to withdraw them. Indeed, because banks operate on a first-come, first-served basis (the so-called sequential service constraint), depositors have a strong incentive to show up at the bank early, before it runs out of funds. General uncertainty about the health of the banking system can lead to runs on both good and bad banks, and the failure of one bank can hasten the failure of others. If nothing is done to restore the public's confidence, a bank panic can ensue, driving both solvent and insolvent banks out of business and leaving depositors with large losses.

A government safety net for depositors can short-circuit bank runs and bank panics. Deposit insurance is one such safety net. If deposits are fully insured, depositors do not need to run to the bank to make withdrawals—even if they are worried about its health—because their deposits will be worth 100 cents on the dollar no matter what. Insurance need not cover 100 percent of deposits to decrease the incentive for depositors to run to withdraw deposits when they are unsure about the bank's health.

Even in the absence of explicit deposit insurance governments often stand ready to provide support to domestic banks when they face runs. This support is sometimes provided by central bank lending to the troubled institutions and is often referred to as the lender-of-last-resort role of the central bank. In other cases governments provide funds directly to troubled institutions or take over the institutions and guarantee that depositors will receive their money in full.

Although a government safety net can protect depositors and prevent bank panics, it is a mixed blessing. Its most serious drawback stems from moral hazard, an important feature of insurance arrangements in general, which increase incentives for taking risks that might result in an insurance payoff. Moral hazard is a prominent risk in government arrangements to provide a safety net because depositors expect that they will not suffer losses if a bank fails. Thus they may not impose the discipline of the marketplace on banks by withdrawing deposits when they suspect

that the bank is taking on too much risk. As a result banks that are provided with a safety net can (and do) take on greater risks than they otherwise would.

A further problem with safety nets arises because of adverse selection. The people who are most likely to produce the adverse outcome insured against (bank failure) are those who most want to take advantage of the insurance. Because depositors have little reason to impose discipline on the bank if they know it is subject to a safety net, risk-loving entrepreneurs find the banking industry a particularly attractive one to enter. Even worse, because depositors have so little reason to monitor the bank's activities if it has a safety net, outright crooks may also find banking attractive because it is easy for them to get away with fraud and embezzlement.

The moral hazard created by the desire to prevent bank failures presents bank regulators with a quandary. Because the failure of a very large bank makes it more likely that a major systemic financial disruption will occur, bank regulators are naturally reluctant to allow a big bank to fail and create losses for its depositors. Thus they are likely to adhere to the so-called too big to fail policy, in which no depositor suffers a loss. (The phrase *too big to fail* is somewhat misleading because when a bank is closed or merged with another bank, the managers may be fired and stockholders in the bank lose their investment.) The main problem with this policy is that it increases the incentives for moral hazard, especially by large banks. Because depositors at large banks know that they will not be subject to any losses, they have no incentive to monitor the bank and discourage it from risk taking by pulling out their deposits when it takes on too much risk. As a result big banks take on even greater risks, thus making bank failures more likely. For example, Boyd and Gertler (1993) found that during the 1980s large banks in the United States took on riskier loans than smaller banks, a practice that led to higher loan losses for large U.S. banks.

Different safety nets may lead to different incentives for moral hazard. A deposit insurance scheme and a less formal scheme in which the government clearly stands ready to support any troubled bank might produce the same degree of moral hazard because in both schemes depositors know that they will be fully protected by the government. The close equivalence of the two schemes is more likely in a banking system with a small number of large banks because the too-big-to-fail policy is operational for most of the banking system.

In many situations, however, a formal deposit insurance scheme may lead to greater risk-taking moral hazard by the banking system than would occur under a less formal government safety net. Deposit insurance protects all depositors, whether the shock to the banking system is systemic (systemwide) or idiosyncratic (affecting a single troubled bank). A less formal guarantee means that the government is less likely to respond to an idiosyncratic shock than to a systemic shock that threatens the health of the banking system. The result is that banks know that depositors may withdraw funds if they think that their bank could be brought down by an idiosyncratic shock, with the result that banks will want to reduce their risk of an idiosyncratic shock. With deposit insurance, however, the bank does not need to protect itself against idiosyncratic shocks and so has greater incentives to take on more risk. This phenomenon led to the U.S. savings and loan (S&L) crisis. Many

small S&L institutions, knowing that deposit insurance protected their depositors under all conditions, pursued rapid growth that exposed them to idiosyncratic risks that eventually brought them down and imposed large losses on taxpayers.

Restrictions on Asset Holdings and Bank Capital Requirements

Bank regulations that restrict asset holdings and impose capital requirements are directed at preventing this risk-taking moral hazard, which can cost taxpayers dearly. Even in the absence of a government safety net, banks still have the moral hazard incentive to take on too much risk. Risky assets may provide the bank with higher earnings if they pay off. But if they do not pay off and the bank fails, depositors are left holding the bag. If depositors were able to easily monitor the bank by acquiring information on its risk-taking activities, they would immediately withdraw their deposits if the bank was taking on too much risk. To prevent such a loss of deposits, banks would be more likely to reduce their risk-taking activities. But acquiring information on a bank's balance sheet and off–balance sheet activities is a difficult task, and so most depositors are incapable of imposing discipline that might prevent banks from engaging in risky activities. Thus even in the absence of a government safety net there is a strong rationale for government regulation to reduce risk taking on the part of banks.

Regulations that restrict banks from holding risky assets such as common stock are a direct means of making banks avoid too much risk. Bank regulations also promote diversification, which reduces risk by limiting the volume of loans in particular categories or to individual borrowers. Requirements that banks have sufficient capital are another way to change the bank's incentives to take on less risk. A bank that is forced to have a large amount of equity capital has more to lose if it fails and thus is less likely to engage in high-risk activities. In addition, capital requirements can be tied to the amount of risk the bank is pursuing, as with the Basle agreements, in effect charging the bank a higher insurance premium when it takes on more risk, thereby discouraging risk taking.

Chartering and Examination

Because banks can be used by risk-taking individuals or even by crooks to engage in highly speculative activities, such undesirable people are likely to want to run a bank. Chartering banks is one method for preventing this adverse selection problem: proposals for opening new banks are screened to prevent undesirable people from controlling them.

Regular bank examinations, which allow regulators to monitor whether a bank is complying with capital requirements and restrictions on asset holdings, also limit moral hazard. In addition, bank examiners can assess whether a bank has the proper management controls in place to prevent fraud or excessive risk taking. With this information about a bank's activities, bank examiners can enforce capital requirements and force a bank to revise its management practices if they are jeopardizing

the bank's soundness. Actions taken by examiners to prevent banks from taking on too much risk further help to reduce the adverse selection problem because, with less of an opportunity for risk taking, risk-loving entrepreneurs will be less likely to be attracted to the banking industry.

Disclosure Requirements

The free-rider problem described earlier indicates that individual depositors and other bank creditors will not have enough incentive to produce private information about the quality of a bank's assets. To ensure that there is better information for depositors and the marketplace, regulators can require that banks adhere to certain standard accounting principles and disclose a wide range of information that helps the market assess the quality of the bank's portfolio and the amount of the bank's exposure to risk. Publicly available information about the risks incurred by banks and about the quality of their portfolios can better enable stockholders, creditors, and depositors to evaluate and monitor banks, and so to act as a deterrent to excessive risk taking.

This view is consistent with a recent discussion paper issued by the Eurocurrency Standing Committee of the Central Banks of the Group of 10 Countries (1994), which recommends that estimates of financial risk generated by firms' own internal risk management systems be adapted for public disclosure purposes (see also FRBNY 1994). Such information would supplement disclosures based on traditional accounting conventions by providing information about risk exposure and risk management that is not normally included in conventional balance sheet and income statement reports.

Prompt Corrective Action

Bank regulation can reduce moral hazard and adverse selection problems in the banking system only if regulators pursue prompt corrective action when banks fail to comply with regulatory requirements. This means that bank supervisors must enforce regulations consistently and must avoid regulatory forbearance, that is, allowing banks to keep on operating as usual despite noncompliance with regulations in the hope that the bank's problem will go away with time. Allowing the bank to keep operating as usual creates an additional moral hazard problem for the banking system. Other banks, seeing the regulatory forbearance, will assume that they too might not be punished for taking greater risks that might lead to noncompliance. Thus even if regulatory forbearance seems to be an appropriate strategy at the time, it creates incentives that in the future may lead to undesirable behavior for the banking system.

It is particularly important to avoid regulatory forbearance for noncompliance with bank capital requirements. If a bank has too little capital, its incentives to engage in moral hazard and to take big risks increase dramatically. The most extreme case is when the institution is economically insolvent and its owners have

almost nothing to lose by taking on great risk and "betting the bank": if the bank gets lucky and its risky investments pay off, the bank gets out of its insolvency. But if, as is likely, the risky investments do not pay off, the bank's losses will mount and the government will be left holding the bag.

Why the Regulatory Process Might Not Work As Intended

To act in the public interest and to lower costs to the deposit insurance agency, regulators must set tight restrictions on holding assets that are too risky, impose adequate capital requirements, and avoid regulatory forbearance, particularly that which allows insolvent institutions to continue to operate. However, this is not always what occurs in practice.

There are two reasons the regulatory process might not work as intended, one having to do with resources and one with the principal-agent problem. The first is that regulators and bank managers may not have sufficient resources or knowledge to do their job properly. An example of this occurred in the United States and helped lead to a regulatory breakdown in the S&L industry in the 1980s. Deregulation of the S&L industry in the early 1980s opened up many lines of business to these institutions. Once restricted almost entirely to making loans for home mortgages, these thrift institutions were now allowed to have up to 40 percent of their assets in commercial real estate loans, up to 30 percent in consumer lending, and up to 10 percent in commercial loans and leases. Regulators allowed up to 10 percent of assets to be in junk bonds or in direct investments (common stocks, real estate, service corporations, and operating subsidiaries).

Three problems arose from these expanded powers for S&Ls. First, many S&L managers did not have the expertise to manage risk appropriately in these new lines of business. Second, the new powers meant that there was rapid growth in new lending, particularly to the real estate sector. Rapid credit growth may outstrip a bank's information resources, resulting in excessive risk taking. Third, these new powers of the S&Ls and the lending boom meant that their activities were expanding in scope and becoming more complicated, requiring an expansion of regulatory resources in order to monitor them. The regulators at the Federal Savings and Loan Insurance Corporation (FSLIC) did not have the expertise or the additional resources needed to monitor these new activities properly. Given the lack of expertise in the S&L industry and the weakening of the regulatory apparatus, it is no surprise that S&Ls took on excessive risks, leading to massive losses for U.S. taxpayers.

Similar events have occurred in other industrial countries. Finland, Norway, and Sweden also deregulated their financial markets in the early 1980s, with similar results (Drees and Pazarbasioglu 1995). The lack of expertise both in the banking industry and among its regulators in keeping risk taking in check, particularly when bank credit was growing rapidly as a result of a lending boom, resulted in massive losses to loan portfolios when real estate prices collapsed in the late 1980s. The resultant government bailouts of the banking industry in those countries were simi-

lar in scale (relative to GDP) to those in the United States. The Japanese banking crisis that has been unfolding in recent years shares similar features. Deregulation of the financial system in the 1980s was followed by increased bank lending, especially to the real estate sector, which resulted in huge loan losses when the real estate sector collapsed.

The second reason regulators may not do their job properly is explained by the relationship between voters-taxpayers and regulators and politicians, which creates a particular type of moral hazard problem, the principal-agent problem. The problem occurs when agents have different incentives from the people they work for (the principals) and so act in their own interest rather than in the interest of their employers. Regulators and politicians are ultimately agents for voters-taxpayers (principals) because taxpayers ultimately bear the cost of any losses when the safety net is invoked. The principal-agent problem occurs because the agent (a politician or regulator) may not have the same incentives to minimize costs to the economy as the principal (the taxpayer). The principal-agent problem stems from asymmetric information because the principal does not have sufficient information about what the agent is doing to make sure that the agent is operating in the principal's interest.

A classic example of this principal-agent problem occurred in the United States during the S&L debacle of the 1980s. By the late 1970s many S&Ls in the United States were insolvent because most of their assets were tied up in fixed-rate long-term mortgages, whose rates were fixed at a time when interest rates were low. When interest rates rose in the 1970s and early 1980s, the cost of funds for these institutions rose dramatically but their fixed-rate mortgages did not generate higher income. As a result the economic net worth of the S&Ls plunged dramatically. In addition, the 1981–82 recession and the price collapse in energy and farm products hit some parts of the country (such as Texas) very hard, causing defaults on many S&Ls' loans. Losses mounted to $10 billion in 1981–82, and by some estimates more than half the S&Ls in the United States had a negative net worth and were thus insolvent by the end of 1982.

As the analysis earlier indicates, the incentives to engage in risk-taking moral hazard increased dramatically for these insolvent institutions because they now had little to lose and a lot to gain by taking on excessive risks. Clearly, it was essential that prompt corrective action be taken and that these institutions be closed down. Instead, S&L regulators loosened capital requirements and restrictions on risky asset holdings and pursued regulatory forbearance. One important incentive for regulators that explains this phenomenon was their desire to escape blame for poor performance. By loosening capital requirements and pursuing regulatory forbearance, regulators hid the problem of insolvent banks and hoped that the situation would improve. Edward Kane (1989) refers to such behavior on the part of regulators as "bureaucratic gambling."

Another important incentive for regulators is their desire to protect their careers by acceding to pressures from the people who most influence their careers. These people are not the taxpayers but the politicians who try to keep regulators from

imposing tough regulations on institutions that are major campaign contributors. Members of Congress have often lobbied regulators to ease up on a particular S&L that contributed large sums to their campaigns.

In addition, both the U.S. Congress and the executive branch promoted banking legislation in 1980 and 1982 that made it easier for S&Ls to engage in risk-taking activities. With passage of this legislation the need for monitoring of the S&L industry increased because of the expansion of permissible activities. Regulatory agencies needed more resources to carry out their monitoring activities properly, but Congress (successfully lobbied by the S&L industry) was unwilling to allocate the necessary funds. As a result the S&L regulatory agencies became so short of personnel that they had to cut back on their on-site examinations just when they were most needed. Between January 1984 and July 1986, for example, several hundred S&Ls were not even examined once. Even worse, in response to intense lobbying by the S&L industry, Congress passed legislation in 1987 (the Competitive Banking Equality Act) providing a woefully inadequate $15 billion to close down the insolvent S&Ls. Only in 1991, when the crisis had reached massive proportions requiring a bailout of more than $100 billion in present value terms, did Congress pass the Federal Deposit Insurance Corporation Improvement Act (FDICIA), which provided the necessary funds to clean up the S&L mess and tightened up the bank regulatory process.

A Theory of Banking and Financial Crises: A Developing Country Perspective

In recent years the analysis based on asymmetric information, applied above to elucidate the structure of the financial system and the rationale for bank regulation, has also been used to develop a theory of banking and financial crises (Bernanke 1983; Calomiris and Gorton 1991; and Mishkin 1991, 1994). The theory has been used mainly to explain banking and financial crises in industrial countries, particularly the United States. But the institutional framework in the United States is quite different from that in many developing countries, and thus the theory requires some modification for use in understanding banking and financial crises in developing countries.

Before beginning this analysis, we need to define financial crisis in the context of asymmetric information theory: a financial crisis is a nonlinear disruption to financial markets in which adverse selection and moral hazard problems become much worse, so that financial markets are unable to efficiently channel funds to economic agents who have the most productive investment opportunities. A financial crisis thus prevents financial markets from functioning efficiently, which leads to a sharp contraction in economic activity.

Understanding how banking and financial crises lead to contractions in economic activity requires an understanding of the factors that lead to banking and financial crises. Four categories of factors promote financial crises: increases in interest rates, increases in uncertainty, asset market effects on balance sheets, and bank panics.

Increases in Interest Rates

As demonstrated by Stiglitz and Weiss (1981), asymmetric information and the resulting adverse selection problem can lead to credit rationing in which some borrowers are denied loans even when they are willing to pay a higher interest rate. Individuals and firms with the riskiest investment projects are willing to pay the highest interest rates since they will be the main beneficiaries if the high-risk investment succeeds. Thus a higher interest rate leads to even greater adverse selection by increasing the likelihood that the lender is lending to a bad credit risk. If the lender cannot discriminate among borrowers and identify those with the riskier investment projects, it may reduce the number of loans it makes, which causes the supply of loans to decrease rather than increase as interest rates rise. Thus even if there is excess demand for loans, a higher interest rate will not be able to equilibrate the market because additional increases in the interest rate will further reduce the supply of loans and increase the excess demand for loans.

The theory behind credit rationing can be used to show that increases in interest rates can be one factor that precipitates a financial crisis. If market interest rates are driven up sufficiently, there is a higher probability that lenders will lend to bad credit risks because good credit risks are less likely to want to borrow at the higher rates while bad credit risks are still willing to borrow. Because of the resulting increase in adverse selection, lenders will want to make fewer loans, possibly leading to a steep decline in lending that leads to a substantial decline in investment and aggregate economic activity. Indeed, as Mankiw (1986) has demonstrated, a small rise in the riskless interest rate can sometimes lead to a large decrease in lending and even a possible collapse in the loan market.

Increases in Uncertainty

A dramatic increase in uncertainty in financial markets, due perhaps to the failure of a prominent financial or nonfinancial institution, a recession, political instability, or a stock market crash, makes it harder for lenders to separate good from bad credit risks. The increase in uncertainty therefore makes information in the financial markets even more asymmetric and worsens the adverse selection problem. The inability of lenders to solve the adverse selection problem renders them less willing to lend, leading to a decline in lending, investment, and aggregate economic activity.

Asset Market Effects on Balance Sheets

The state of the balance sheet of both nonfinancial firms and banks has important implications for the severity of asymmetric information problems in the financial system. Deterioration of balance sheets worsens both adverse selection and moral hazard problems in financial markets, and if the deterioration is dramatic enough it is a major factor leading to banking and financial crises.

One important way that financial markets can solve asymmetric information problems is with the use of collateral. Collateral reduces the consequences of adverse selection or moral hazard because it reduces the lender's losses in the case of a default. If a borrower defaults on a loan, the lender can take title to the collateral and sell it to make up for the losses on the loan. If the collateral is of good enough quality, the fact that there is asymmetric information between borrower and lender is no longer as important since the loss incurred by the lender if the loan defaults is substantially reduced.

Net worth performs a similar role to that of collateral. If a firm has high net worth, the lender can take title to the net worth in the case of a loan default, sell it off, and use the proceeds to recoup some of the losses from the loan. (Note that in a multiperiod context, as Gertler 1988a shows, the concept of a borrower's net worth can be broadened to include the discounted value of future profits, as reflected in the market value of the borrowing firm.) In addition, the more net worth a firm has in the first place, the less likely it is to default because it has a cushion of assets that it can use to pay off its loans. Like collateral, high net worth directly decreases the incentives for borrowers to commit moral hazard because they have more at stake, and thus more to lose, if they default on their loans. Thus when firms seeking credit have high net worth, the consequences of adverse selection and moral hazard are less important and lenders will be more willing to make loans.

Stock market crashes can precipitate banking and financial crises through their net worth effects on the adverse selection and moral hazard problems described above. As emphasized by Greenwald and Stiglitz (1988), Bernanke and Gertler (1989), and Calomiris and Hubbard (1990), a stock market crash can increase adverse selection and moral hazard problems in financial markets because it leads to a large decline in the market value of firms' net worth. (Note that this decline in asset values could occur either because of expectations of lower future income streams from these assets or because of a rise in market interest rates that lowers the present discounted value of future income streams.) The decline in net worth makes lenders less willing to lend because the lower net worth provides them with less protection, so that losses from loans are likely to be more severe. In addition, the decline in corporate net worth increases moral hazard incentives for borrowing firms to make risky investments because these firms now have less to lose if their investments go sour. Because borrowers have greater incentives to engage in moral hazard and because lenders are now less protected against the consequences of adverse selection, the stock market decline leads to decreased lending and a decline in economic activity.

In addition to a direct effect on increasing adverse selection problems, increases in interest rates can indirectly promote a financial crisis through their effect on firms' and households' balance sheets. As Bernanke and Gertler (1995) point out, a rise in interest rates, and therefore in households' and firms' interest payments, decreases firms' cash flow, which causes a deterioration in their balance sheets (see also Hubbard 1995; Cecchetti 1995; and Mishkin 1996). As a result adverse selec-

tion and moral hazard problems become more severe for potential lenders to these firms and households, leading to a decline in lending and economic activity. There is thus an additional reason why sharp increases in interest rates can be an important factor leading to financial crises.

In economies in which inflation has been moderate, as it has been in most industrial countries, many debt contracts are of fairly long duration. In this institutional environment an unanticipated decline in inflation leads to a decrease in the net worth of firms. Long-duration debt contracts have interest payments fixed in nominal terms for a substantial period of time, with the fixed interest rate allowing for expected inflation. When inflation turns out to be less than anticipated, which can occur either because of an unanticipated disinflation (as occurred in the United States in the early 1980s) or because of outright deflation (as frequently occurred in the United States before World War II), the value of firms' liabilities rises in real terms, increasing the burden of debt, but there is no corresponding rise in the real value of firms' assets. The result is that net worth declines in real terms. Thus a sharp, unanticipated disinflation or deflation causes a substantial decline in real net worth and an increase in adverse selection and moral hazard problems (of the type discussed earlier in assessing the effect of net worth declines). Again, these problems cause a decline in investment and economic activity.

By contrast, many developing countries have experienced very high and variable inflation rates, so that debt contracts are typically of very short duration. For example, almost all bank lending in Mexico is with variable rate contracts, usually adjusted on a monthly basis. In this institutional framework an unanticipated decline in inflation does not have the unfavorable direct effect on firms' balance sheets that it has in industrial countries because the terms of the debt contract are continually changed to reflect expectations of inflation. Thus one mechanism that has promoted financial crises in low-inflation countries has no role in developing countries that have experienced high and variable inflation.[4]

Another factor affecting balance sheets that can precipitate a financial crisis in developing countries but that is not operational in most industrial countries is unanticipated exchange rate depreciation or devaluation. Because of uncertainty about the future value of the domestic currency, the government and many banks and nonfinancial firms in developing countries find it easier to issue debt if it is denominated in foreign currencies (a few industrial countries share this characteristic). This was a prominent feature of the institutional structure in Chilean financial markets before the financial crisis in 1982 and in Mexico in 1994. When debt contracts are denominated in foreign currency, an unanticipated depreciation or devaluation of the domestic currency increases the debt burden of domestic firms. Since assets are typically denominated in domestic currency, there is no corresponding rise in the value of firms' assets. The result is a deterioration in firms' balance sheets. The decline in net worth then increases adverse selection and moral hazard problems along the lines described above, leading to a decline in investment and economic activity.

Bank Panics

Because banks are well suited to gathering information that facilitates productive investment, the simultaneous failure of many banks reduces the amount of financial intermediation undertaken and leads to a decline in investment and aggregate economic activity. Indeed, even if banks do not fail but just suffer a substantial contraction in capital, bank lending will decline, thereby leading to a contraction in economic activity. Research in the United States suggests that this mechanism was operational during the early 1990s and that the capital crunch led to the "headwinds" mentioned by Alan Greenspan that were hindering U.S. economic growth (*Reuters Library Report* 1991; Bernanke and Lown 1991; Berger and Udell 1994; Hancock, Laing, and Wilcox 1995; Peek and Rosengren 1995; and FRBNY 1993).

As explained earlier, asymmetric information is the source of bank panics, with depositors rushing to make withdrawals from solvent as well as insolvent banks because they cannot distinguish between them. Bank runs and panics are more likely to occur when banks' balance sheets are in a weakened state, making it more likely that the bank is insolvent. Weak bank balance sheets can occur because the supervisory and regulatory structure has not worked well enough to restrain excessive risk taking on the part of banks. In addition, banks can be buffeted by shocks that cause a rapid deterioration in their balance sheets.

Increases in interest rates, stock market crashes, an unanticipated decline in inflation (in industrial countries), or an unanticipated depreciation or devaluation (when debt is denominated in foreign currencies) can cause a deterioration in nonfinancial firms' balance sheets that makes it less likely that they can pay back their loans. These factors can precipitate sharp increases in loan losses, which increase the probability of bank insolvency.

Banks in developing countries face other potential shocks that can make a banking crisis more likely. As primary goods producers many developing countries are often subject to large terms of trade shocks that can devastate banks' balance sheets since banks' assets are composed primarily of loans to domestic firms. Also, banks in many developing countries raise funds with liabilities denominated in foreign currencies. A depreciation or devaluation of the domestic currency can thus lead to increased indebtedness, while the banks' assets do not rise in value. The resulting deterioration in banks' equity capital then increases the possibility of bank failures and panics. Even if the depreciation does not lead directly to bank failures, it can lead to substantial declines in bank lending. As their equity capital deteriorates, banks may fail to meet capital standards, such as the Basle requirements, and so they must shrink their lending until they can raise new capital to meet the capital standards.

Applying the Theory of Financial Crises to Past Episodes

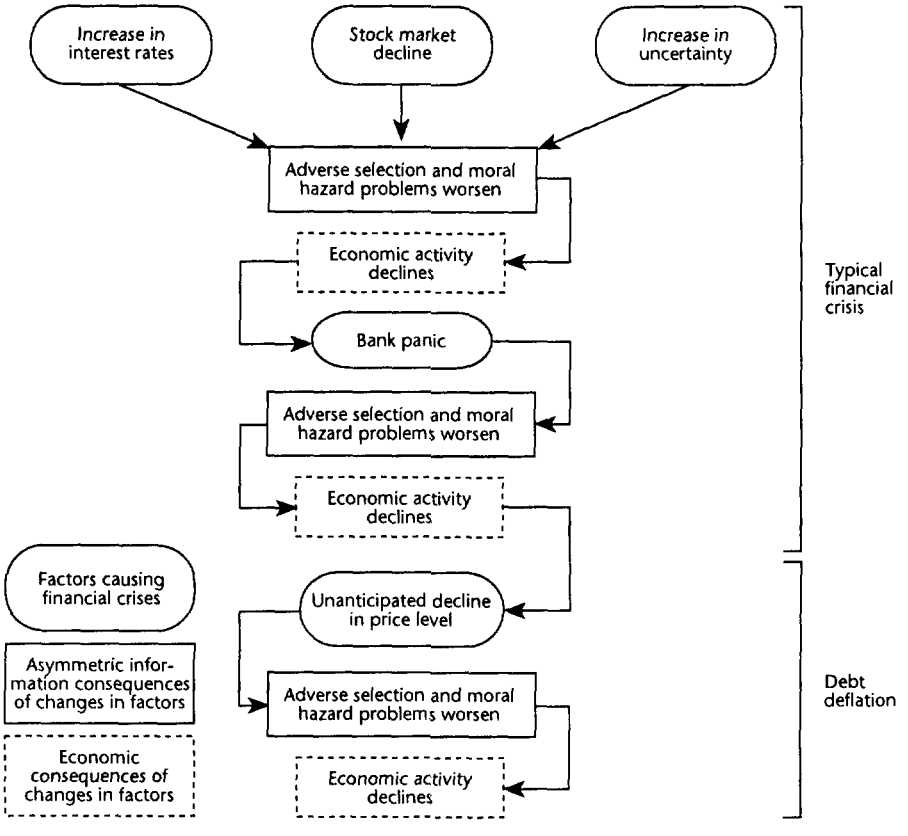
The theory of banking and financial crises laid out above can be verified by seeing how well it explains the sequence of events that have occurred during past banking and financial crises. Theory leads us to expect a somewhat different sequence of

events for industrial and developing countries because of the differences in their institutional framework. I look first at how well the theory explains the facts for an industrial country—the United States in the nineteenth and early twentieth centuries—and then at how well it does so for a developing country—Mexico in 1994 and 1995.

U.S. Experience in the Nineteenth and Early Twentieth Centuries

In the nineteenth and early twentieth centuries the United States experienced banking and financial crises every twenty years or so. As documented in Mishkin (1991) and shown in figure 1, most financial crises in the United States began with a sharp rise in interest rates (frequently transmitted from abroad, particularly from the London market), a stock market crash, and an increase in uncertainty because of recession and the failure of major financial or nonfinancial firms (Ohio Life Insurance & Trust Co. in 1857, Northern Pacific Railroad and Jay Cooke & Co. in 1873, Grant & Ward in 1884, National Cordage Co. in 1893, Knickerbocker Trust Company in 1907, and the Bank of the United States in 1930.) During these finan-

Figure 1. Sequence of Events in U.S. Financial Crises of the Nineteenth and Early Twentieth Centuries

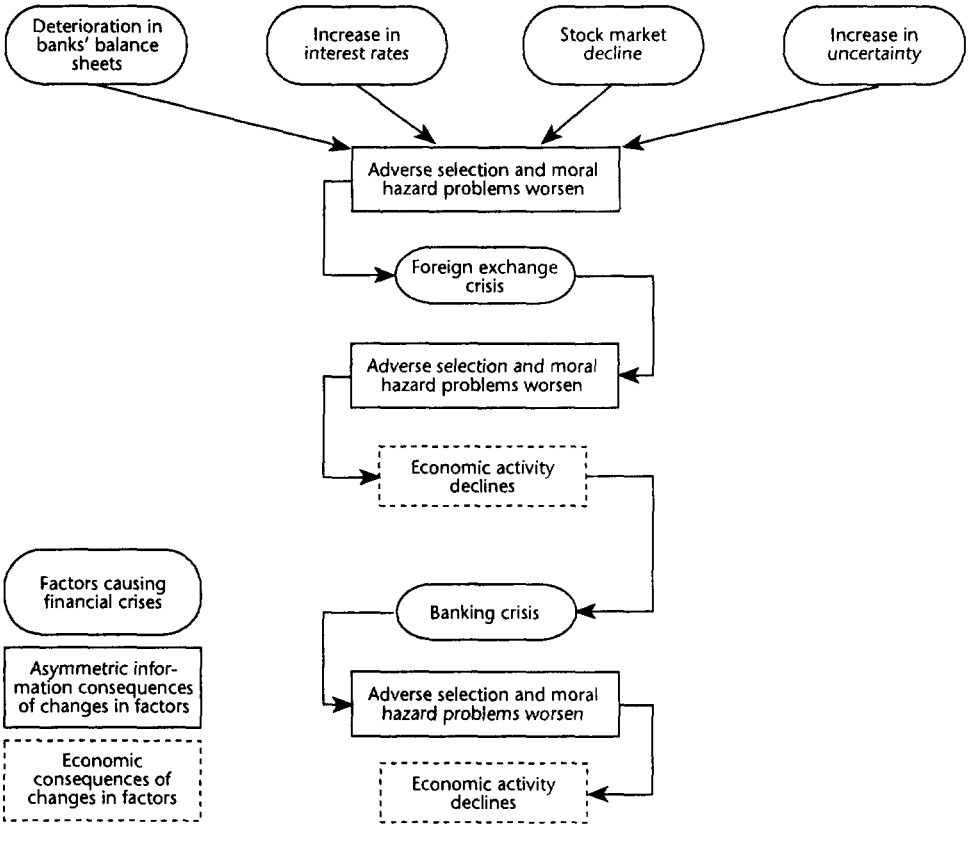


cial crises the increase in uncertainty, the rise in interest rates, and the stock market crash increased the severity of adverse selection problems in credit markets, while the decline in net worth stemming from the stock market crash also increased moral hazard problems. The increase in adverse selection and moral hazard problems made it less attractive for lenders to lend and led to a decline in investment and aggregate economic activity.

Because of the worsening business conditions and uncertainty about their banks' health, depositors began to withdraw their funds. The resulting bank panic and decline in the number of banks pushed interest rates even higher and reduced the amount of financial intermediation by banks. The spread between interest rates on low- and high-quality bonds widened, indicating a worsening of the problems created by adverse selection and moral hazard (Mishkin 1991), which led to yet greater economic contraction.

At this point in many of the episodes, including the panics of 1857, 1884, 1890, 1893, and 1907, there would come a sorting out of insolvent firms from healthy firms through bankruptcy proceedings. The same process would occur for banks, often with the help of public and private authorities. Once the sorting out was com-

Figure 2. Sequence of Events in the Mexican Financial Crisis of 1994–95



plete, uncertainty in financial markets would subside, the stock market would begin its recovery, and interest rates would fall. Adverse selection and moral hazard problems would diminish, the spread between interest rates on low- and high-quality bonds would narrow, and the financial crisis would be over. With financial markets able to operate well again, the stage would be set for economic recovery.

In a few episodes, however, such as the 1873 panic and the Great Depression, the economic downturn and contraction of the money supply resulting from the bank panic led to a sharp decline in prices. This unanticipated deflation—described by Irving Fisher (1933) as debt deflation—short-circuited the recovery process. It led to a further deterioration in firms' net worth because of the increased burden of indebtedness. When debt deflation set in, adverse selection and moral hazard problems worsened. As a result investment spending and aggregate economic activity remained depressed for a long time.

Several conclusions can be drawn from the U.S. episodes of financial crisis. The timing and pattern of the episodes seem to fit an asymmetric information interpretation of financial crises. Rather than starting with bank panics, most financial crises began with a rise in interest rates, a stock market crash, and a widening of the interest rate spread. Financial panics were frequently immediately preceded by a major failure of a financial firm and the beginning of a recession, which increased uncertainty in the marketplace. That increase in uncertainty and the rise in interest rates magnified the adverse selection problem in credit markets, while the stock market decline increased adverse selection and moral hazard problems. The worsening of adverse selection and moral hazard problems then led to a decline in investment activity and in aggregate economic activity.

Only after these problems have manifested themselves in financial markets did bank panics occur. Thus the theory outlined above helps to explain the timing of bank panics, because it sees them as a consequence of high interest rates, a major failure of a corporation or a nonbank financial institution, or weak business conditions stemming from recession, which make depositors nervous about the health of the banks that hold their deposits. Since depositors cannot easily screen good banks from bad when this adverse aggregate information appears, they worry about potential losses on their deposits and withdraw funds from the banking system, precipitating a panic. The facts about the crisis episodes discussed in Mishkin (1991) are thus entirely consistent with Gorton's (1988) view that bank panics are predictable and with asymmetric information theory. (The same timing of bank crises is typically found in developing countries; see Hausmann and Gavin 1995 and Kaminsky and Reinhart 1996.)

Once a bank panic sets in, the resulting loss of liquidity causes interest rates to rise further, the stock market to decline further, and adverse selection and moral hazard to worsen. Finally, with the sorting out of solvent from insolvent firms and banks, the crisis subsides, the stock market recovers, interest rates fall, and if the economic uncertainty and unanticipated deflation were not too severe, adverse selection and moral hazard problems diminish. If there has been no substantial deflation, we then see, as expected, a rapid decline in the spread between interest rates

for low- and high-quality borrowers. But if substantial deflation has set in, we see evidence of a debt-deflation process in which aggregate economic activity is depressed for a prolonged period.

Mexico's Experience in 1994–95

Because the institutional features of Mexico's debt markets differ from those of the United States in the nineteenth and early twentieth centuries, the sequence of events in the Mexican banking and financial crisis that began in December 1994 differs as well, although there are some similarities. A diagrammatic exposition of the sequence of events in the Mexican case is shown in figure 2.

An important factor leading up to the Mexican financial crisis was the deterioration in banks' balance sheets because of increasing loan losses. As the U.S. S&L debacle showed, deregulation of a financial system and rapid credit growth can be disastrous when banking institutions and their regulators lack the resources and skills to keep risk taking in check. A similar situation occurred in Mexico. Mexican banks had been nationalized in 1982 and, not surprisingly, they directed a large share of their lending to government—on the order of 50 percent. When the banks were privatized in the early 1990s, they had no formal credit bureaus to monitor loans to households and small businesses to make sure that borrowers were not taking on excessive risk. Yet bank credit to private nonfinancial enterprises grew rapidly, going from 10 percent of GDP in 1988 to more than 40 percent of GDP in 1994 (figure 3). This lending boom, which exceeded the screening and monitoring capabilities of the banks, occurred as a result of both increased flows of savings into the banking sector and an increasing share of bank lending going to private firms.

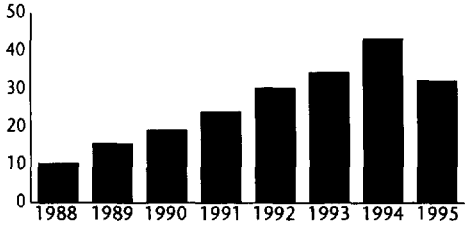
Nor did the primary regulator of banks in Mexico, the National Banking Commission, have the capability to monitor banks' loan portfolios and management practices to prevent inordinate risk taking. This lack of capabilities was exacerbated by the bank lending boom and by the tremendous expansion in lending by other financial institutions, such as credit unions, thrifts, and leasing companies. Not surprisingly, given what we have seen in similar situations in the United States and in the Nordic countries, when Mexican banks and nonbank financial institutions expanded their loan activities, they did indeed take on excessive risk. Many bad loans were the result (figure 4).

Also consistent with the U.S. experience in the nineteenth and early twentieth centuries was another precipitating factor in the Mexican crisis, a rise in interest rates abroad. In February 1994 the U.S. Federal Reserve began to raise the interbank federal funds rate to prevent incipient inflationary pressures from taking hold. Although the policy succeeded in keeping inflation in check in the United States, the upward pressure it put on Mexican interest rates (figure 5) increased asymmetric information problems in the Mexican financial system.[5] Central bank actions to protect the value of the peso also contributed to the rise in Mexican interest rates. The rise in interest rates directly increased adverse selection in Mexican financial mar-

Origins of Mexico's Financial Crisis

Bank credit to private firms rose...

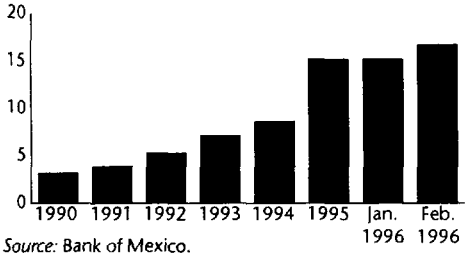
Figure 3. Bank Credit to Private Enterprises, 1988–95 (percentage of GDP)



Source: Bank of Mexico.

...and many of the loans were bad

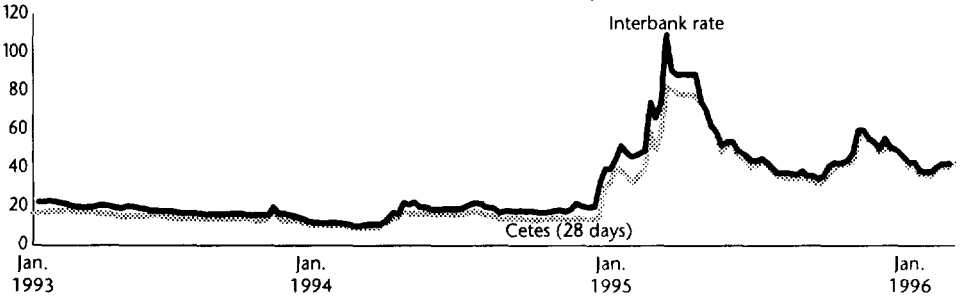
Figure 4. Nonperforming Loans as a Share of Total Loans, 1990–96 (percent)



Source: Bank of Mexico.

Interest rates spiked...

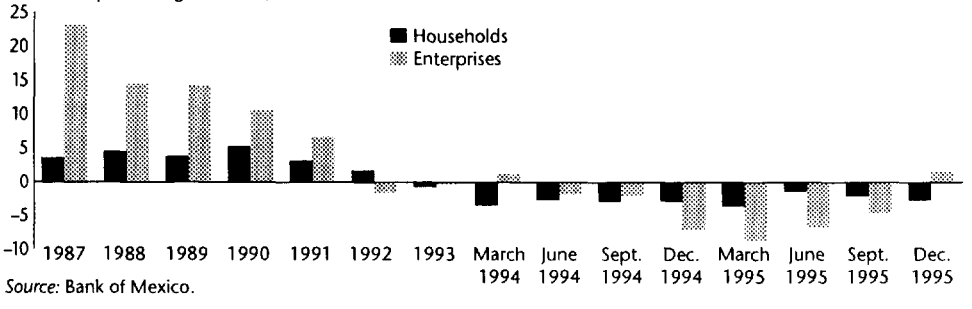
Figure 5. Interest Rates on Cetes and Interbank Loans, 1993–96 (percent)



Source: Bank of Mexico; IMF, *International Financial Statistics.*

...contributing to the deterioration of household and enterprise balance sheets

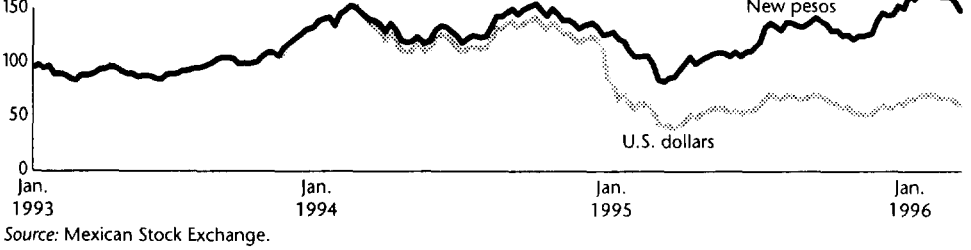
Figure 6. Net Creditor or Debtor Position with the Domestic Financial System, 1987–95 (stocks as percentage of GDP)



Source: Bank of Mexico.

The stock market declined as uncertainty increased

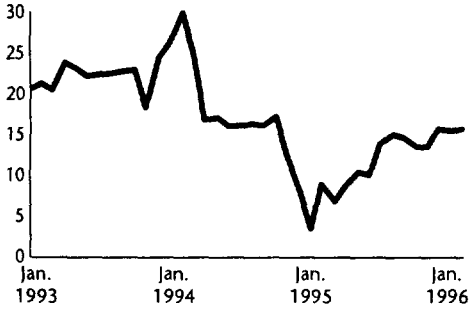
Figure 7. Stock Market Prices on the Bolsa, 1993–96, index (1993 = 100)



Source: Mexican Stock Exchange.

International reserves shrank to support the peso

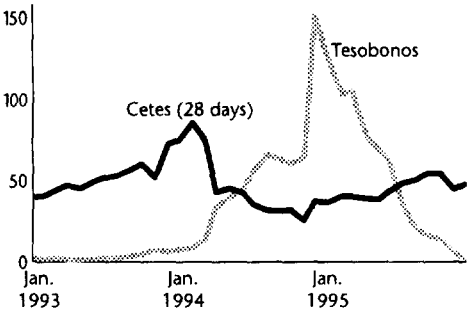
Figure 8. Mexico's International Reserves, 1993–96 (billions of U.S. dollars)



Source: Bank of Mexico.

Dollar-denominated bonds outpaced peso-denominated bonds

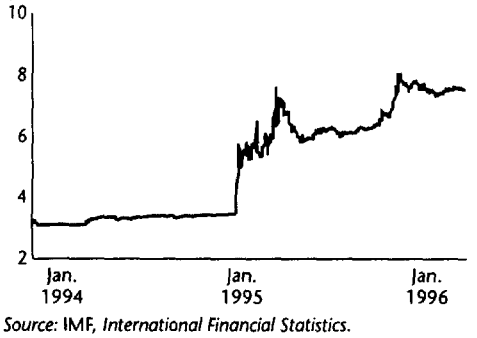
Figure 9. Cetes and Tesobonos Amounts, 1993–95 (billions of new pesos)



Source: Bank of Mexico.

The peso halved in value...

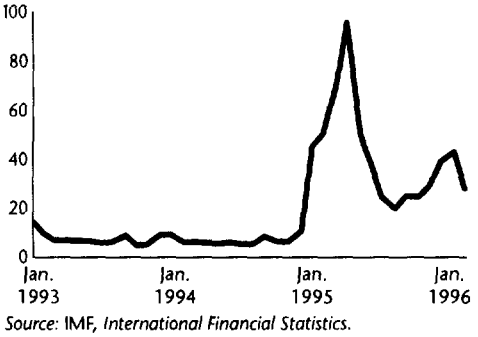
Figure 10. Peso-Dollar Exchange Rate, 1993–96 (pesos per U.S. dollar)



Source: IMF, *International Financial Statistics.*

...leading to a dramatic rise in inflation

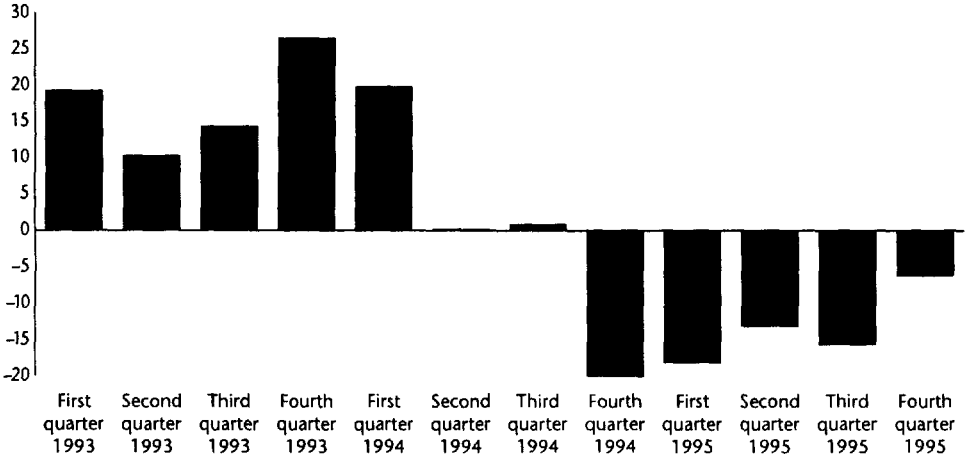
Figure 11. Consumer Price Inflation, 1993–96 (percent)



Source: IMF, *International Financial Statistics.*

Asymmetric information problems intensified, and foreign lenders pulled out funds

Figure 12. Quarterly Flows of Foreign Portfolio Investment, 1993–95, Annualized (billions of U.S. dollars)



Note: Balance of payments data; portfolio investment is portfolio equity and money market investment.
Source: Bank of Mexico.

kets by making it more likely that those willing to take on the most risk would seek loans. Even more important, increased interest payments lowered households' and firms' cash flow, causing a deterioration in their balance sheets (figure 6). Because debt contracts in Mexican financial markets are very short in duration, the rise in short-term interest rates had a substantial effect on the cash flow and balance sheets of households and firms. As asymmetric information theory suggests, this deterioration in balance sheets increased adverse selection and moral hazard problems in Mexican financial markets, making it less desirable for lenders to lend.

Also consistent with the U.S. experience in the nineteenth and early twentieth centuries, increases in uncertainty in Mexican financial markets and a stock market decline (figure 7) precipitated the full-blown financial crisis. The Mexican economy was hit by political shocks in 1994—specifically the Colosio assassination and the uprising in Chiapas—that increased general uncertainty in financial markets, while stock prices on the Bolsa fell nearly 20 percent between their peak in September 1994 and the middle of December 1994. The increase in uncertainty and decline in net worth increased asymmetric information problems by making it harder to distinguish good borrowers from bad and by increasing firms' incentives to make risky investments (since they had less to lose if the investments failed).

The increase in interest rates and uncertainty, the stock market decline, and the deterioration in banks' balance sheets were the initial conditions that worsened adverse selection and moral hazard problems (as shown at the top of figure 2). They made the Mexican economy ripe for a serious financial crisis when the foreign exchange market went into a full-blown crisis.

The political shocks of 1994 brought the Mexican peso under attack. Intervention in the foreign exchange market by the Bank of Mexico to support the commitment to a pegged exchange rate resulted in a substantial loss of international reserves (figure 8). With the growing uncertainty in the foreign exchange market, the government found it harder to finance its debt with peso-denominated bonds (cetes) and so dramatically increased its issues of dollar-denominated bonds (tesobonos) (figure 9). Even though the Mexican central bank raised interest rates sharply, the hemorrhaging of international reserves forced the Mexican authorities to devalue the peso on December 20, 1994.

The institutional structure of debt markets in Mexico then interacted with the peso devaluation to propel the economy into a full-fledged financial crisis. Devaluations are a highly nonlinear event because they result in a sharp change in the exchange rate. With the peso halved in value by March 1995 (figure 10), there was a dramatic rise in both actual inflation (figure 11) and expected inflation. Combined with the desire of the Bank of Mexico and the Mexican government to limit the peso depreciation, this rise in inflation caused interest rates on debt denominated in pesos to skyrocket to over 100 percent on an annual basis. The Mexican stock market crashed, falling another 30 percent in peso terms and more than 60 percent in dollar terms. And because many firms had debt denominated in dollars, the peso depreciation resulted in an immediate sharp increase in their indebtedness in pesos, while the value of their assets remained unchanged. The depreciation of

the peso that began in December 1994 led to an especially sharp negative shock to the net worth of private firms (see figure 6), causing a dramatic increase in adverse selection and moral hazard problems.

These asymmetric information problems were severe for both domestic and foreign lenders. Foreign lenders, who had the additional burden of difficulty in obtaining information about what was going on in the Mexican economy, quickly pulled their funds out of Mexico. Foreign portfolio investment inflows to Mexico, which had been on the order of $20 billion a year in 1993 and early 1994, reversed course. By the beginning of the fourth quarter of 1994 outflows exceeded $10 billion at an annualized rate (figure 12). Consistent with the theory of financial crises outlined in this article, the sharp decline in lending helped lead to a collapse in economic activity, with real GDP growth falling from around 4–4.5 percent a year in the second half of 1994 to rates of around –10 percent in the second and third quarters of 1995.

Further deterioration of the economy now occurred because the collapse in economic activity and the deterioration in the cash flow and balance sheets of firms and households led to a worsening of the banking crisis (see figure 2). Many firms and households were no longer able to pay off their debts, resulting in substantial loan losses for banks. In addition, the depreciation of the peso had a direct negative impact on banks' balance sheets. The foreign currency–denominated liabilities of Mexican banks jumped from 116.4 billion pesos at the end of December 1993 to 213.9 billion pesos at the end of December 1994, primarily as a result of the decline in the value of the peso from 3.1 to the dollar to 5.3 to the dollar (table 1). Although Mexican banks did have offsetting foreign currency–denominated assets, their likelihood of being paid off in full was substantially lowered because of the worsening business conditions and the negative effect of the increases in the peso value of these foreign currency–denominated loans on the balance sheets of the borrowing firms. Even if banks have a matched portfolio of foreign currency–denominated assets and liabilities and so appear to avoid the foreign exchange risk, a devaluation can nonetheless cause substantial harm to bank balance sheets. The resultant mismatch between foreign currency–denominated assets and liabilities on borrowers' balance sheets can force them to default on their loans, thereby converting a market risk for borrowers to a credit risk for banks with foreign currency–denominated loans.

Even more problematic for the Mexican banks is that many of their foreign currency–denominated liabilities were very short term, so that the sharp increase in the value of these liabilities led to liquidity problems because they had to be paid back quickly. Collapse of the banking system would then have been inevitable in the absence of a government safety net. But the Mexican government intervened, providing the funds to protect depositors and thereby avoiding a bank panic. However, as banks lost capital and began needing government support, they became less willing and able to lend. As we have seen, a banking crisis that hinders the ability of banks to lend also makes adverse selection and moral hazard problems worse in financial markets because banks are no longer as capable of playing their traditional financial intermediation role.

Table 1. Foreign Currency Exposure of Mexican Banks
(billions of pesos)

	December 1993	December 1994	March 1995	June 1995	September 1995	December 1995
Foreign currency–denominated assets						
Tesobonos	0.4	3.6	1.4	0.8	0.4	0
Securities	1.5	6.3	6.4	9.1	11.5	22.5
Loans	54.0	99.2	130.5	126.7	124.4	146.3
Assets of branches and agencies	67.3	116.8	132.7	123.7	121.9	142.4
Total	123.2	225.9	271.0	260.3	258.2	311.2
Foreign currency–denominated liabilities						
In Mexico	49.1	97.1	136.2	121.2	119.9	154.9
In branches and agencies	67.3	116.8	132.7	123.7	121.9	142.4
Total	116.4	213.9	268.9	244.9	241.8	297.3
Exposure						
Foreign assets minus foreign liabilities	6.8	12.0	2.1	15.4	16.4	13.9
Foreign assets/total assets (percent)	15.6	28.6	29.7	27.7	27.3	29.9
Foreign liabilities/total liabilities (percent)	14.7	27.0	29.5	26.0	25.6	28.6
Exchange rate (peso/dollar)	3.106	5.325	6.818	6.309	6.420	7.643

Source: Bank of Mexico.

The theory of banking and financial crises outlined in this article provides a cohesive explanation for the sequence of events that developed during the Mexican crisis of 1994–95. In addition, it explains the puzzle mentioned in the introduction to this article: it shows how a developing economy can shift dramatically from a path of reasonable growth before a financial crisis, as was the case in Mexico in 1994, to a sharp decline in economic activity after a crisis. The answer is that the financial crisis leads to such a substantial worsening of adverse selection and moral hazard problems in financial markets that there is a collapse in lending and hence in economic activity.[6]

Policy Measures for Recovering from Financial Crises

Since the consequences of financial crises are so disastrous, knowing what steps to take to expedite recovery from financial crises and minimize their effects on the economy is very important. Not surprisingly, the institutional structure of the financial system affects the measures that can be used by the authorities to stimulate recovery from a financial crisis. This section looks first at the policy measures that can be taken in industrial countries and than shows how the different institutional

framework of developing countries means that some of these measures may not work in these countries and may even be counterproductive.

Industrial Countries

Two features of the institutional structure of financial systems in most industrial countries are very important for enabling the central bank to stimulate a recovery from financial crises: debt contracts are denominated almost solely in domestic currency, and many debt contracts are of fairly long duration because inflation has tended to be moderate.

With this kind of institutional structure one method for a central bank to promote recovery from a financial crisis is to inject liquidity (reserves) into the financial system through open market operations or through lending to the banking sector. This expansionary monetary policy causes the money supply to increase, leading to a higher price level. Since debt contracts are denominated in domestic currency and many are of long duration, the higher price level produces the opposite effect from that of the unanticipated disinflation mechanism described earlier. The reflation of the economy lowers the debt burden of households and firms, thereby increasing their net worth. In turn, the increase in adverse selection and moral hazard problems induced by the financial crisis is undone. Injecting liquidity into the economy also raises the prices of such assets as land and stocks, further improving net worth and reducing adverse selection and moral hazard problems. Also, as discussed in Mishkin (1996), an expansionary monetary policy promotes economic recovery through other mechanisms involving the stock market and the foreign exchange market.[7]

The Federal Reserve's expansionary monetary policy and reflation of the economy were important elements in the recovery from the Great Depression in the United States after 1933 (Romer 1992). By increasing reserves and the money supply, the Federal Reserve stimulated the price level and induced other asset prices to rise. As a result financial markets recovered and the economy grew rapidly. Japan's recent experience also demonstrates the ability of an expansionary monetary policy to stimulate the economy in the face of a banking crisis. In 1995 the Bank of Japan began pursuing a more expansionary monetary policy that kept nominal interest rates low and increased the growth rate of the monetary base. The result was a sharp increase in stock market prices, a slowdown in the deflationary process, and some encouraging signs of an economic recovery in 1996.

A second method for promoting recovery from a financial crisis is for the central bank to stand ready to lend freely to banks during a financial crisis—the so-called lender-of-last-resort role. A textbook example of this use occurred in the United States in the aftermath of the stock market crash of October 19, 1987 (Mishkin 1995). Although "Black Monday" will go down in history as the largest one-day decline in stock prices to date, it was on Tuesday, October 20, 1987, that financial markets received their worst threat. To keep the stock market and the related index futures market functioning in an orderly fashion, brokers needed to extend massive amounts of credit on behalf of their customers for their margin calls. To get a sense

of the magnitude of the problem consider that just two brokerage firms, Kidder, Peabody and Goldman, Sachs, had advanced $1.5 billion in response to margin calls on their customers by noon on October 20. Clearly, brokerage firms as well as specialists (dealer-brokers who match buy and sell orders at a stock exchange) were severely in need of additional funds to finance their activities. Banks, however, were understandably growing very nervous about the financial health of securities firms and so were reluctant to lend to them just when they needed it most.

According to Brimmer (1989), Alan Greenspan, chairman of the Federal Reserve, and E. Gerald Corrigan, president of the New York Federal Reserve Bank and the Fed official most closely in touch with Wall Street, began to fear a breakdown in the clearing and settlement systems and the collapse of securities firms. To prevent this catastrophe, Greenspan announced before the market opened on October 20 the Federal Reserve System's "readiness to serve as a source of liquidity to support the economic and financial system." In addition to this extraordinary announcement, the Fed encouraged key money center banks to lend to their brokerage firm customers and made it clear that sufficient liquidity would be available to enable these banks to make such loans. (As it turned out the Federal Reserve made liquidity available through open market operations rather than through discount loans.) The banks did as they were encouraged and by October 21 had increased by $7.7 billion their loans to brokers and to individuals to purchase or hold securities. As a result the markets kept functioning on October 20 and a market rally that day raised the Dow Jones Industrial Average by more than 100 points (over 5 percent).

Evidence in Mishkin (1991) demonstrates that the stock market crash of October 1987 had the potential to produce a financial crisis but that the prompt action of the Federal Reserve prevented the crisis from spinning out of control. Indeed, one lesson from this episode is that quick action by the central bank often means that the amount of actual lending needed can be very small. The quicker the intervention, the smaller the amount of lending that is required.

Developing Countries

It may be far more difficult for the central bank to promote recovery from a financial crisis in developing countries than in industrial countries because of two institutional features of the financial system that preclude the use of expansionary monetary policy. Many developing countries have much of their debt denominated in foreign currency, and for many a past record of high and variable inflation has resulted in debt contracts of very short duration. As a result of these institutional features, pursuing an expansionary monetary policy to reflate the economy is likely to cause the domestic currency to depreciate sharply and expected inflation to rise dramatically.

As noted earlier, depreciation of the domestic currency leads to a deterioration in firms' and banks' balance sheets because much of their debt is denominated in foreign currency. Their burden of indebtedness rises and their net worth falls. The

upward jump in expected inflation is likely to cause interest rates to rise because lenders need to be protected against a loss of purchasing power when they lend. The resulting rise in interest rates causes interest payments to soar and the cash flow of households and firms to decline, resulting in a deterioration in households' and firms' balance sheets and potentially even greater loan losses to banking institutions. Also, the short duration of debt contracts means that the rise in the price level from expansionary monetary policy does not appreciably affect the value of households' and firms' debts, so there is little benefit to their balance sheets from this mechanism, unlike the case in industrial countries.

Thus the net result of an expansionary monetary policy in a developing country with the institutional structure described above is the opposite of that in an industrial country after a financial crisis: it causes a deterioration in balance sheets and therefore amplifies rather than ameliorates the adverse selection and moral hazard problems in financial markets caused by a financial crisis.

For similar reasons lender-of-last-resort activities by a central bank in a developing country may not be as successful as they are in an industrial country. When the Federal Reserve pursued a lender-of-last-resort role during the 1987 stock market crash, there was almost no sentiment in the market that this would lead to substantially higher inflation. That is much less likely to be the case in a developing country with a history of high and variable inflation. There, central bank lending to the financial system to expand domestic credit in the wake of a financial crisis might arouse fears of inflation spiraling out of control. The resultant rise in interest rates and depreciation of the exchange rate will cause cash flow and balance sheets to deteriorate, making recovery from the financial crisis less likely. The lender-of-last-resort role must be used far more cautiously by a central bank in a developing country with the institutional structure outlined here because central bank lending is now a two-edged sword.

Thus recovery from a financial crisis is a much more complicated exercise for many developing countries than it is for industrial countries. Expansionary monetary policy is not an option for stimulating recovery in most developing countries. Rather, monetary policy must be restricted to keeping inflation low and restoring confidence in the domestic currency. Indeed, a speedy recovery may require foreign assistance because liquidity from foreign sources does not lead to the undesirable consequences that result from the provision of liquidity by domestic authorities. Obviously there are other policy measures to stimulate recovery from a financial crisis when expansionary monetary policy is not an option. But the point being emphasized here is that certain institutional structures make a developing country's economy more prone to financial crisis and make it more difficult to extricate the economy from a financial crisis if it occurs.

Policy Implications for Developing Countries

Although the lessons for policy derived from the asymmetric information analysis developed here are not entirely new, this analytical framework provides a means of

assessing which policy measures are effective for reducing both the likelihood of banking and financial crises and their negative effects on the economies of developing countries.

The asymmetric information framework suggests the importance of a government safety net for the banking system—the key feature of which is not deposit insurance—in order to prevent bank panics. But because a safety net increases the moral hazard incentives for excessive risk taking by banks—a problem intensified in developing countries, where information may be more difficult to acquire—developing countries need to create and sustain a strong bank regulatory and supervisory system.

Forging a Strong Bank Supervisory System

A strong bank regulatory and supervisory system has several characteristics. First, bank regulatory and supervisory agencies need adequate resources to do their job effectively. Absent these resources, as the S&L supervisors in the United States and the National Banking Commission in Mexico found out, a bank supervisory agency will be unable to monitor banks well enough to keep them from engaging in inappropriately risky activities or to see that they have the appropriate management expertise and controls to manage risk and sufficient capital to keep in check the moral hazard incentives to take on excessive risk.

Second, proper accounting standards and disclosure requirements are crucial to a healthy banking system. Such standards are often lacking in developing countries and need to be beefed up considerably. Without appropriate information both markets and bank supervisors will be unable to adequately monitor banks to deter excessive risk taking.[8]

Third, bank supervisors need to take prompt corrective action to stop undesirable bank activities and, even more important, to close down institutions that do not have sufficient net worth, making sure that stockholders and managers of these insolvent institutions are appropriately punished. The asymmetric information framework indicates that prompt corrective action is particularly important. Most immediately, it prevents banks from "betting the bank" in order to restore the value of the institution. It also creates incentives for banks not to take on too much risk in the first place because they know that if they do, they are likely to be punished.

Fourth, because prompt corrective action is so important, the bank supervisory agency requires sufficient independence from the political process so that it is not encouraged to sweep problems under the rug by engaging in regulatory forbearance. One way to do this is to give the bank supervisory role to a politically independent central bank. This has desirable elements (Mishkin 1992), but some central banks might not want the supervisory task for fear that it would politicize them and impinge on their independence. Bank supervisory activities could instead be assigned to a separate agency that is independent of the government. Another way to improve incentives for bank supervisors is to hold them accountable if they engage in regulatory forbearance. Opening up their actions to public scrutiny makes regulatory forbearance less attractive, thereby reducing the principal-agent problem.

It also reduces the incentives of politicians to lean on bank supervisors to relax their supervision. An important but often overlooked part of the 1991 Federal Deposit Insurance Corporation (FDIC) Improvement Act in the United States that has helped make this legislation effective is the requirement that supervisory agencies produce a report whenever a bank failure imposes costs on the FDIC. The report is available to any member of Congress or the general public.

Liberalizing the Financial System

Deregulation and liberalization of the financial system have been strongly promoted for developing countries in recent years. Although these are desirable objectives, deregulation and liberalization can be disastrous if not managed properly, as the asymmetric information framework developed in this article shows. If the proper bank regulatory and supervisory structure is not in place before liberalization, risk-taking behavior will not be adequately constrained. Bad loans are the likely outcome, with potentially calamitous consequences for bank balance sheets at some point in the future. Financial deregulation and liberalization also often lead to a lending boom, both because of increased opportunities for bank lending and because of a financial deepening that brings more funds into the banking system, opening up new lending opportunities that may result in loan portfolios of very poor quality. And although financial deepening is a positive development for the economy in the long run, in the short run the lending boom may outstrip the information resources of the financial system, helping to promote an eventual financial collapse.

These dangers do not mean that developing countries should avoid liberalization. To the contrary, financial liberalization is critical to the efficient functioning of financial markets. Getting funds to borrowers with the most productive investment opportunities is particularly important in developing countries, where the right investments can have especially high returns, thereby stimulating rapid economic growth. Financial deregulation and liberalization thus need to be pursued actively and managed carefully. That means getting the proper bank regulatory and supervisory structure in place before liberalizing the financial system. It means providing sufficient resources to bank supervisors, installing adequate accounting and disclosure requirements, encouraging bank supervisors to take prompt corrective action, and insulating bank supervision from the political process. Developing countries may need to move slowly in financial liberalization in order to keep a lending boom from getting out of hand and exceeding the capabilities of bank management and supervisors.

Policy Implications of Other Institutional Features

The asymmetric information framework also sheds light on institutional features of the financial system besides bank regulation that can determine how prone a developing economy is to a financial crisis and how severe the economic effects of such a crisis are likely to be.

One such feature is the legal and judicial framework, which is very important to an efficient financial system. The inadequacies of legal systems in many developing countries are a serious problem for financial markets. When property rights are unclear or hard to enforce, financial intermediation is severely hampered. In many developing countries the legal system makes attaching collateral a costly and time-consuming process, thereby reducing the effectiveness of collateral in solving adverse selection and moral hazard problems in credit markets. Similarly, cumbersome bankruptcy procedures frequently cause long delays in resolving conflicting claims and so delay recovery from a financial crisis. Resolving bankruptcies by opening up the books of insolvent firms and redistributing their assets decreases asymmetric information in the marketplace and, as the discussion of U.S. financial crises shows, is an important part of the recovery from a financial crisis. Not until bankruptcies have been resolved is there enough information in the financial system to restore it to healthy operation.

The short duration of debt contracts is another important institutional feature of financial systems in many developing countries. By increasing cash-flow and liquidity problems for nonfinancial firms and banks when interest rates rise or the domestic currency depreciates, an institutional structure of short-duration debt renders the financial system more prone to financial crises. Developing countries need to encourage the development of longer-term debt markets to promote financial stability. One way is to use issues of longer-term government debt to increase liquidity in markets for long-term debt.[9]

An impediment to this strategy of issuing long-term government debt is that financial institutions in developing countries might be able to finance purchases of these securities in the money market only with very short-term liabilities. If there are substantial fluctuations in interest rates, a common problem in developing countries, the resulting maturity mismatch for these institutions can result in substantial market (interest-rate) risk. A sharp rise in interest rates, by decreasing the value of the long-term securities banks hold as assets while having little effect on the value of their short-term liabilities, can cause a sharp deterioration in banks' net worth and lead to insolvencies. This problem suggests that regulatory requirements or bank supervisory procedures to limit market risk may be especially important in developing countries. Pursuing price stability is also important because without it there will be substantial fluctuations in interest rates, and the resultant high interest-rate risk will make it harder to induce private markets to issue long-term debt.

Denominating a large amount of debt in foreign currencies is another institutional feature of many developing countries. Mexico's experience and that of other countries illustrate how dangerous that practice can be. Denominating debt in foreign currencies is very tempting for firms, banks, and governments in developing countries because of the difficulty of attracting capital, particularly foreign capital, for debt denominated in domestic currency, whose value is more unstable. It is important that this temptation be resisted. Denominating debt in foreign currencies not only makes a developing country's financial system more prone to financial crisis, it also intensifies the negative consequences of such a crisis. Furthermore, if a

financial crisis does occur, a developing country with a lot of its debt denominated in foreign currencies has limited policy options for extricating itself from the crisis. The dangers of denominating debt in foreign currencies is one of the most important policy lessons derived from the asymmetric information theory of banking and financial crises in developing countries.

Another is the grave danger of a pegged exchange rate regime for a developing country with a fragile banking system, short-duration debt contracts, and substantial amounts of debt denominated in foreign currencies. Some developing countries have chosen to peg their currency to a stable currency like the dollar because they are seeking a nominal anchor that will promote price stability. But if the domestic currency depreciates, interest rates rise dramatically, indebtedness worsens, and the resultant sharp deterioration in firms' and banks' balance sheets tips the country into full-scale financial crisis, with devastating effects on the economy. Thus combining a pegged exchange rate regime with the institutional features outlined above is like putting the economy on a knife's edge: one slip and the economy comes crashing down.[10]

A flexible exchange rate regime has the advantage that movements in the exchange rate are much less nonlinear than in a pegged exchange rate regime. Indeed, in a flexible exchange rate regime the daily fluctuations in the exchange rate have the advantage of making clear to private firms, banks, and governments the substantial risk involved in issuing liabilities denominated in foreign currencies. Furthermore, a depreciation of the exchange rate may provide an early warning that policies need to be adjusted in order to limit the potential for a financial crisis.

In Pursuit of Price Stability

I have already touched on the topic of price stability as a worthy goal. Being a central banker, I cannot resist turning back to a central banker's favorite topic (or, as some would say, obsession). I will not go into all the reasons why price stability is so desirable and why it is usually considered the primary goal for central banks in industrial countries. This is well-covered territory. But what the analysis of banking and financial crises here makes even clearer is how imperative the pursuit of price stability is for developing countries. Price stability can help promote financial stability because it leads to longer-duration debt contracts. And achieving price stability is a necessary condition for having a sound currency. With a sound currency it is far easier for banks, nonfinancial firms, and government to raise capital with debt denominated in domestic currency, which makes the financial system less fragile and reduces the negative impact of a financial crisis on the economy. Price stability also promotes the ability of policymakers to extricate a country from a financial crisis if it occurs.

Furthermore, countries with highly variable inflation have a credibility problem that limits what policymakers can do to promote recovery from a financial crisis. Without credibility, a central bank in a developing country that tries to use expansionary monetary policy to speed recovery from a financial crisis may do more harm than good. Instead of shoring up weakened balance sheets, an expansionary monetary policy is likely to lead to a rapid rise in expected inflation and interest rates, as

well as to depreciation of the exchange rate, all of which cause balance sheets to deteriorate and make the financial crisis worse. Similarly, engaging in a lender-of-last-resort rescue might backfire because it leads to worries about the commitment to low inflation. With a credible commitment to price stability, this vicious cycle will not occur. Expansionary monetary policy and the lender-of-last-resort role can be used to shore up balance sheets and nip a financial crisis in the bud or to promote rapid recovery from a financial crisis that has already occurred, as examples from U.S. history suggest. The analysis of banking and financial crises in developing countries using the asymmetric information framework thus leads to the following conclusion: having an independent central bank with a clear mandate for price stability is possibly even more important for developing countries than it is for industrial countries.

But as with the worthy goal of financial liberalization, single-minded pursuit of price stability can be dangerous. Rapid disinflation leading to high real interest rates has adverse cash-flow consequences for financial institutions. If the financial system is fragile because of already weakened balance sheets, the disinflation could result in a major financial crisis and a blow to the economy. Thus, before engaging in an anti-inflation stabilization program, developing countries need to attend to the health of their financial system, making sure that the regulatory and supervisory process has been effective in promoting strong balance sheets for financial institutions. Otherwise, financial institutions may not survive the stresses of an anti-inflation stabilization program. Successful monetary policy in developing countries, therefore, requires successful regulation and prudential supervision of the financial system.

Notes

1. Gertler (1988b) provides an excellent survey of the literature on asymmetric information and financial structure. Transactions costs also play a role in explaining why banks are such important players in the financial system, but asymmetric information issues are emphasized in this article because they help explain banks' unique position in the financial system and also help to understand phenomena such as financial crises.

2. Banks cannot entirely eliminate the free-rider problem. Knowing that a bank has made a loan to a particular company reveals information to other parties that the company is more likely to be creditworthy and will be undergoing monitoring by the bank. Thus some of the benefits of information collection produced by the bank will accrue to others.

3. As Edwards and Mishkin (1995) point out, the traditional financial intermediation role of banking has been in decline in both the United States and other industrial countries, where improved information technology has made it easier to issue securities. Although this trend suggests that traditional banking may eventually diminish in importance in developing countries as well, the barriers to information collection are so great in developing countries that the dominance of banks will continue for the foreseeable future.

4. However, an unanticipated decline in inflation during periods when an anti-inflation program is under way in developing countries has often been associated with very high real interest rates. Thus an unanticipated decline in inflation can hurt the balance sheets of firms in developing countries through the cash-flow mechanism discussed above.

5. Bank of Mexico officials believe that the interbank interest rate is a better measure of money market interest rates than the rate on cetes, the peso-denominated treasury bills. Cetes rates might have been artificially low because certain domestic investment funds were required to hold cetes and foreigners were discouraged by regulations in their own countries from purchasing money market instruments other than cetes, making cetes the instrument generally used for repurchase operations.

6. Of course, traditional transmission mechanisms operating through higher interest rates and the fiscal policy responses also helped to worsen the economic downturn.

7. Not all industrial countries are alike in their ability to use expansionary monetary policy to recover from a financial crisis. A country with a commitment to a pegged exchange rate might not be able to pur-

sue an expansionary monetary policy because that might force a devaluation of its currency. Even a country with a flexible exchange rate might not be able to pursue an expansionary monetary policy to promote recovery because that might cause a depreciation of the domestic currency that the authorities consider intolerable, particularly in smaller countries. Clearly, a large reserve currency country like the United States has the most flexibility in using expansionary monetary policy to reflate the economy to recover from or reduce the probability of a financial crisis.

8. A recent paper by Garber and Lall (1996) suggests that undisclosed off-balance sheet and offshore derivatives contracts played an important role in the Mexican crisis.

9. However, one advantage of governments issuing short-term debt is that it reduces the incentive to pursue inflationary policies to reduce the real value of their liabilities. Similarly, denominating government debt in foreign currency eliminates any incentive to lower the value of the debt through inflation and depreciation of the domestic currency. It is not clear how important these advantages are in practice, however.

10. See Obstfeld and Rogoff (1995) for additional reasons pegged exchange rate regimes may be undesirable.

References

Akerlof, George A. 1970. "The Market for Lemons: Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84: 488–500.

Berger, Allen N., and Gregory F. Udell. 1994. "Do Risk-Based Capital Requirements Allocate Bank Credit and Cause a 'Credit Crunch' in the United States?" *Journal of Money, Credit, and Banking* 26 (3): 585–628.

Bernanke, Ben S. 1983. "Non-Monetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review* 73: 257–76.

Bernanke, Ben S., and Mark Gertler. 1989. "Agency Costs, Collateral, and Business Fluctuations." *American Economic Review* 79: 14–31.

———. 1995. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives* 9 (fall): 27–48.

Bernanke, Ben S., and Cara S. Lown. 1991. "The Credit Crunch." *Brookings Papers on Economic Activity* 2. Washington, D.C.: Brookings Institution.

Boyd, James, and Mark Gertler. 1993. "U.S. Commercial Banking: Trends, Cycles, and Policy." *NBER Macroeconomics Annual 1993*. Cambridge, Mass.: National Bureau of Economic Research.

Brimmer, Andrew F. 1989. "Distinguished Lecture on Economics in Government: Central Banking and Systemic Risks in Capital Markets." *Journal of Economic Perspectives* 3: 3–16.

Calomiris, Charles W., and Gary Gorton. 1991. "The Origins of Banking Panics: Models, Facts, and Bank Regulation." In R. G. Hubbard, ed., *Financial Markets and Financial Crises*. Chicago: University of Chicago Press.

Calomiris, Charles W., and R. Glenn Hubbard. 1990. "Firm Heterogeneity, Internal Finance, and 'Credit Rationing.'" *Economic Journal* 100: 90–104.

Cecchetti, Stephen G. 1995. "Distinguishing Theories of the Monetary Transmission Mechanism." Federal Reserve Bank of St. Louis *Review* 77 (May/June): 83–97.

Diamond, Douglas W. 1984. "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies* 51: 393–414.

Drees, Burkhard, and Ceyla Pazarbasioglu. 1995. "The Nordic Banking Crises: Pitfalls in Financial Liberalization?" IMF Working Paper WP/95/61. International Monetary Fund, Washington, D.C.

Edwards, Franklin R., and Frederic S. Mishkin. 1995. "The Decline of Traditional Banking: Implications for Financial Stability and Regulatory Policy." NBER Working Paper 4993. National Bureau of Economic Research, Cambridge, Mass.

Eurocurrency Standing Committee of the Central Banks of the Group of 10 Countries (Fisher Group). 1994. "Public Disclosure of Markets and Credit Risks by Financial Intermediaries." Discussion paper. Bank for International Settlements, Basle.

Fisher, Irving. 1933. "The Debt-Deflation Theory of Great Depressions." *Econometrica* 1: 337–57.

FRBNY (Federal Reserve Bank of New York). 1993. "The Role of the Credit Slowdown in the Recent Recession." Federal Reserve Bank of New York *Quarterly Review* 18 (1): 1–114.

———. 1994. "Public Disclosure of Risks Related to Market Activity." Discussion paper. New York.

Garber, Peter M., and Subir Lall. 1996. "The Role and Operation of Derivative Markets in Foreign Exchange Market Crises." Brown University, Department of Economics, Providence, R.I.

Gertler, Mark. 1988a. "Financial Capacity, Reliquification, and Production in an Economy with Long-Term Financial Arrangements." Working Paper 2758. National Bureau of Economic Research (NBER), Cambridge, Mass.

———. 1988b. "Financial Structure and Aggregate Economic Activity: An Overview." *Journal of Money Credit and Banking* 20 (part 2): 559–88.

Gorton, Gary. 1988. "Banking Panics and Business Cycles." *Oxford Economic Papers* 40: 751–81.

Greenwald, Bruce C., and Joseph E. Stiglitz. 1988. "Imperfect Information, Finance Constraints, and Business Fluctuations." In M. Kahn and S. C. Tsiang, eds., *Finance Constraints, Expectations, and Macroeconomics.* Oxford: Oxford University Press.

Greenwald, Bruce C., Joseph E. Stiglitz, and Andrew Weiss. 1984. "Information Imperfections in the Capital Market and Macroeconomic Fluctuations." *American Economic Review* 74: 194–99.

Hancock, Diana, Andrew J. Laing, and James A. Wilcox. 1995. "Bank Capital Shocks: Dynamic Effects on Securities, Loans and Capital." *Journal of Banking and Finance* 19 (3–4): 661–77.

Hausmann, Ricardo, and Michael Gavin. 1995. "The Roots of Banking Crises: The Macroeconomic Context." Inter-American Development Bank, Washington, D.C.

Hubbard, R. Glenn. 1995. "Is There a 'Credit Channel' for Monetary Policy?" Federal Reserve Bank of St. Louis *Review* 77 (May/June): 63–74.

Kaminsky, Graciela L., and Carmen M. Reinhart. 1996. "The Twin Crises: The Causes of Banking and Balance of Payments Problems." International Finance Discussion Paper 544. Board of Governors of the Federal Reserve System, Washington, D.C.

Kane, Edward J. 1989. *The S&L Insurance Mess: How Did it Happen?* Washington, D.C.: Urban Institute Press.

Mankiw, N. Gregory. 1986. "The Allocation of Credit and Financial Collapse." *Quarterly Journal of Economics* 101: 455–70.

Mayer, Colin. 1990. "Financial Systems, Corporate Finance, and Economic Development." In R. G. Hubbard, ed., *Asymmetric Information, Corporate Finance, and Investment.* Chicago: University of Chicago Press.

Mishkin, Frederic S. 1991. "Asymmetric Information and Financial Crises: A Historical Perspective." In R. G. Hubbard, ed., *Financial Markets and Financial Crises.* Chicago: University of Chicago Press.

———. 1992. " An Evaluation of the Treasury Plan for Banking Reform." *Journal of Economic Perspectives* 6 (winter): 133–53.

———. 1994. "Preventing Financial Crises: An International Perspective." *Manchester School* 62: 1–40.

———. 1995. *The Economics of Money, Banking, and Financial Markets,* 4th ed. New York: HarperCollins.

———. 1996. "The Channels of Monetary Transmission: Lessons for Policy." Banque de France *Bulletin: Digest* 27 (March): 33–44.

Myers, Stewart C., and Nicolas S. Majluf. 1984. "Corporate Financing and Investment Decisions: When Firms Have Information That Investors Do Not Have." *Journal of Financial Economics* 13: 187–221.

Obstfeld, Maurice, and Kenneth Rogoff. 1995. "The Mirage of Fixed Exchange Rates." *Journal of Economic Perspectives* 9 (4): 73–96.

Peek, Joe, and Eric S. Rosengren. 1995. "Bank Regulation and the Credit Crunch." *Journal of Banking and Finance* 19 (2–4): 679–92.

Reuters Library Report. 1991. "Central Bank Chief Says U.S. Economy Has Turned Sluggish." October 28.

Rojas-Suarez, Liliana, and Steven R. Weisbrod. 1994. "Financial Market Fragilities in Latin America: From Banking Crisis Resolution to Current Policy Challenges." IMF Working Paper WP/94/117. International Monetary Fund, Washington, D.C.

Romer, Christina D. 1992. "What Ended the Great Depression?" *Journal of Economic History* 52 (4): 757–84.

Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71: 393–410.

———. 1983. "Incentive Effects of Terminations: Applications to Credit and Labor Markets." *American Economic Review* 73: 912–27.

Comment on "Understanding Financial Crises: A Developing Country Perspective," by Frederic S. Mishkin

Masahiko Aoki

Frederic Mishkin focuses on adverse selection and moral hazard problems arising from asymmetric information in investor-firm relationships, presenting reasons why banks, rather than security markets, might be a (imperfect) solution to these problems in developing economies. I agree with most of his argument. I did wonder, however, what we could learn by extending his asymmetric information framework to deal with the possible monitoring role of banks in developing economies.

Information in the Financial System

As Mishkin notes, adverse selection (hidden information) problems arise from the asymmetry of information about the riskiness of investment projects before investment occurs. Financial intermediaries may be able to cope with such problems by accumulating expertise in project evaluation and credit analysis (ex ante monitoring). Moral hazard (hidden action) problems arise because investors cannot distinguish the effects of events that management cannot control from the effects of management actions taken in implementing an investment project. Financial intermediaries may be able to reduce these problems by monitoring management activities (interim monitoring).

But there is another monitoring activity undertaken by investors and financial intermediaries that does not fall precisely into these two categories but that has a profound impact on the performance of management. This kind of monitoring includes such activities as verifying the outcome of investment projects (the financial state of the firm) and taking (controlling) action contingent on the outcome. This activity is related to the notion of "costly state verification" investigated by Diamond (1984) and may be referred to as ex post monitoring. Management may try to hide the true state of an outcome to avoid unfavorable actions by investors, but I submit that possible problems arising from the asymmetry of information about the out-

Masahiko Aoki is professor of economics at Stanford University.

Annual World Bank Conference on Development Economics 1996

come of management behavior can be distinguished from the problems arising from ongoing management behavior.

Diamond (1984) deals with a one-period model, so that ex post action contingent on the outcome involves only the distribution of investment returns and the imposition of penalties on defaulting borrowers. The penalties will make the net benefit for borrowers null, concealing the true outcome. In the real world, however, investor-manager relationships are repeatable, and the kind of ex post actions (rescues, dismissals, bankruptcies, takeovers) that are expected will have a significant impact on the subsequent behavior of managers. If a bad outcome is not followed by an appropriate punitive action, the surviving management (and future borrowers) may come to expect that their moral hazard behavior and hidden adverse attributes will not be punished. A good example is the so-called soft budget problem that was rampant in planned economies. The inefficiency of state-owned enterprises went unchecked because everyone expected the state bank to rescue financially distressed enterprises.

Moral Hazard and the Monitoring Role of Banks in Developing Economies

One important reason to distinguish the three stages of monitoring is that economies differ in their institutional arrangements for allocating the three monitoring functions among various intermediaries, with possibly different implications for their real and financial performance (Aoki 1994). From this perspective I want to amplify Mishkin's point that banks may be a good institutional device for coping with the information problems associated with financing in developing economies.

In industrial countries, where the securitization of financing is well advanced, the three stages of monitoring are performed separately by specialized intermediaries and agents. For example, depending on the types of firms and financial instruments involved, ex ante monitoring can be performed by investment banks, commercial banks, venture capital firms, or rating agencies; interim monitoring by boards of directors, commercial banks, fund managers, security analysts, or venture capital firms; and ex post monitoring by accounting firms, bankruptcy courts, takeover raiders, or reorganization specialists. In developing economies there is, at best, very little accumulation of specialized expertise. Under those conditions the three stages of monitoring may be integrated and delegated to banks as a way of stretching scarce monitoring resources.

Developing economies are not at the technological frontier. For them, therefore, the importance of ex ante monitoring lies not in assessing the potential of emerging engineering and commercial technologies but in assessing the organizational and personal capabilities of firms and entrepreneurs for digesting and improving on known technology and penetrating existing markets. For this type of ex ante monitoring the knowledge that a bank accumulates through the management of payments settlement accounts (interim monitoring) may be valuable.

Where the three stages of monitoring are separated, there can be no prior commitment by investors to be responsible for ex post monitoring. Security underwriters and initial investors can dispose of their holdings at any time. The bad ex post outcome is therefore punished by a yet-unknown takeover, application of the rules of bankruptcy procedures, or state intervention. By contrast, the bank, as an integrative delegated monitor, can commit ex ante to a wider menu of ex post actions, ranging from management interventions, rescue operations (rescheduling of loan contracts), and reorganization to liquidation. And it can choose an appropriate action from this menu ex post, contingent on how the situation evolves. Where the effects of management mistakes and inexperience are easily confused with those of a worsening market or technological environment, as is often the case in developing economies, the mechanical application of bankruptcy rules may not necessarily be conducive to the accumulation of management and organizational resources. On the other hand, the credibility of ex post punishment by banks may have better ex ante and interim incentive effects on management.

The problem of information imperfection in developing economies is not limited to that of information asymmetry between investors (lenders) and firms (borrowers). Borrowers may not necessarily have an information advantage regarding the profitability and riskiness of an investment project when there are substantial technological complementarities among various investment projects (among transportation, utilities, and materials, for example), yet the development of capital markets is imperfect. The probability distribution of profitability in those projects may be significantly affected by investment actions in complementary projects in other industries. Without the transmission of relevant information, coordination failures may result. The delegation of ex ante monitoring to the relatively concentrated banking sector may be helpful for information and investment coordination through the sector's wider scope of financing activities.

However, implied relationships between the bank and the firm may involve the risk of moral hazard behavior on the bank's part as well. After all, the bank is just another firm. Mishkin argues that banks are particularly well suited to solving adverse selection and moral hazard problems because they make private loans, thereby avoiding the free-rider problem. But if a bank and a group of firms develop an exclusive, long-term credit relationship (relational banking), there is no guarantee that the bank will act tough later if it discovers financial problems with the affiliated firms, particularly if the bank can expect assistance from the state. One way to make the bank a responsible and accountable ex ante and interim monitor is to make it liable for any costs incurred in possible ex post actions, such as rescue or liquidation. But discipline on banks is often softened by state assistance.

Another point in Mishkin's article about which I have reservations is his claim that collateral can be an effective mechanism for reducing adverse selection and moral hazard problems. It may be, as Stiglitz and Weiss (1981) argue, that borrowers who are willing to place collateral are gamblers (reckless real estate speculators, for example). It may also be that a reliance on collateral deprives banks of the incen-

tive for ex ante and interim monitoring. Such a hazard was dramatically manifested in the real estate bubbles in Japan and the Nordic countries in the late 1980s.

Financial Crises As a Consequence of the Adverse Selection and Moral Hazard Problems of Banks

Mishkin appears to regard the factors leading to financial crisis—an increase in interest rates, stock market declines, the consequential worsening of the balance sheet of both nonfinancial firms and banks, an increase in uncertainty—as exogenous shocks that trigger a worsening of adverse selection and moral hazard problems (see Mishkin's figures 1 and 2). But this reasoning broadens the term adverse selection beyond its established usage. For example, Mishkin argues that "a higher interest rate leads to even greater adverse selection by increasing the likelihood that the lender is lending to a bad credit risk. If the lender cannot discriminate among borrowers and identify those with the riskier investment projects, it may reduce the number of loans it makes" (p. 40). This implies that a higher interest rate (and the other factors mentioned above) leads to a decline in lending as banks seek to avoid the worsened adverse selection problems.

However, the original argument of Stiglitz and Weiss (1981) was that the (Walrasian) interest rate ceases to be the allocator of credit because the potential borrowers who are bidding a higher interest rate may be the ones with higher default possibilities. If the bank emerged as an institutional response to the failure of the Walrasian market mechanism, it would rely on its monitoring capability to ascertain riskiness among potential borrowers. It would not respond passively to the interest rate, as Mishkin argues.

My reading of history, though not as extensive as Mishkin's, is that the factors that he posits as triggers of adverse selection and moral hazard problems are often the consequences of the moral hazard of banks and firms. Before interest rates rise and the stock market falls, the economy is likely to be booming. In that phase of the business cycle economic agents tend to become optimistic about their prospects, and banks may not be an exception. Banks may become more generous in their credit analysis, mutually competing for borrowers, and become permissive about managerial excesses at borrowing firms. Mishkin's recounting of the savings and loan crisis in the United States and the bank crisis in Mexico indicates that this was indeed what happened before those crises. Another revealing story concerns the origins of the current Jusen (specialized corporations that provide housing financing) crisis in Japan. An important lesson of these episodes is that moral hazard and adverse selection problems with financial institutions are a major source, rather than a consequence, of financial crisis.

Thus I would substitute the following characterization for Mishkin's definition of a financial crisis: a financial crisis is a nonlinear disruption to financial markets that is fostered by the moral hazard behavior of financial intermediaries that fail to control adverse selection and moral hazard problems in the real sector, so that financial markets are unable to efficiently channel funds to borrowers who have the most pro-

ductive investment opportunities. The moral hazard behavior of banks is most likely to be generated and sustained by mismanagement of policies for macroeconomic stability. I therefore concur with Mishkin's reminder of the crucial importance of price stability.

Moral Hazard of Regulators: The Political Economy of Regulation

A sticky question in the theory of finance is, who monitors the monitor? This question may be asked in sequence. The management of firms is monitored by investors and their agents. In some economies investors may be represented by stockholders and their agents (the board of directors). In others a bank may be the principal monitor. More generally, in any economy a specific corporate governance structure is defined by a particular combination of debt instruments and equity holdings.

Looking next at who monitors the board of directors, one might say that competition in the market for corporate control provides the ultimate discipline. But in the case of a bank-dominated financial system, who monitors the banks? Does market competition among banks, and banks' concern with their reputation, provide sufficient discipline? Probably not. As Mishkin argues, free competition among banks may lead to frequent runs on banks and disrupt the efficient working of the financial system. What about depositors? In most cases small depositors lack the incentive and the expertise to monitor the bank. It is to cope with these problems that governments provide the safety net of deposit insurance.

To curb possible moral hazard problems arising from deposit insurance, governments impose various regulatory measures, such as capital requirements, on banks. Dewatripont and Tirole (1993) note an interesting analogy between capital requirements and the disciplinary role of external debt in corporate governance. A firm's failure to repay its external debt triggers the automatic transfer of control from stockholders to debt holders. This prospect provides an incentive for stockholders to monitor management. The violation of capital requirements and the consequential stepwise intensification of regulatory interventions are analogous to the transfer of control to debt holders. According to Dewatripont and Tirole, the regulator (the government) acts as a representative for depositors.

Mishkin, along with Dewatripont and Tirole, recognizes, however, that there is also a potential for regulators to engage in moral hazard. When the banking sector is rife with bad debts, regulators often try to cover up the problem, whether to hide their responsibility and mistakes or to respond to political pressure to rescue ailing financial institutions. Mishkin's suggestion that bank supervisory activities be housed in a bank regulatory authority that is independent of the government is one option for coping with such problems. The difficulty lies in how to approximate such an idealized solution in concrete political-economic contexts. Further, the shift from discretionary to rule-oriented regulation may inadvertently penalize bank management for systemic macroeconomic risks for which individual management may not be directly responsible. Mishkin considers that a countercyclical expansionary policy in a financial crisis could mitigate such problems in industrial

economies, but he is skeptical about its effectiveness in developing economies. An alternative measure, suggested by Dewatripont and Tirole (1993), would use procyclical deposit insurance premiums to help banks in a downturn and tax them in a boom.

References

Aoki, Masahiko. 1994. "Monitoring Characteristics of the Main Bank System: An Analytical and Development View." In Masahiko Aoki and Hugh Patrick, eds., *The Japanese Main Bank System: Its Relevance for Developing and Transforming Economies*. Oxford: Oxford University Press.

Dewatripont, Mathias, and Jean Tirole. 1993. *The Prudential Regulation of Banks*. Cambridge, Mass: MIT Press.

Diamond, Douglas. 1984. "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies* 59: 393–414.

Stiglitz, Joseph, and Andrew Weiss. 1981. "Credit Rationing with Imperfect Information." *American Economic Review* 71: 393–410.

Comment on "Understanding Financial Crises: A Developing Country Perspective," by Frederic S. Mishkin

Edward J. Kane

Frederic Mishkin has given us a richly instructive study of financial crises that draws on asymmetric information theory. He derives from it the recommendation that developing countries build a strong system for regulating and supervising banks.

Mishkin's theory and recommendation are broadly consistent with my own analysis of periodic financial breakdown. Although I want to celebrate the points of agreement, I must emphasize one important point of difference: I understand a weak system for regulating and supervising banks to be a stubborn political-economic equilibrium in most countries. Politicians' penchant for weak supervisory systems is a consequence of a category of asymmetric information that Mishkin does not address.

This additional informational asymmetry comes from half-true disinformation that is deliberately introduced into the flow of data that government regulators send to taxpayers about how well regulation is working. No matter how much economists may exhort government leaders, a nation will not build a financial regulatory system strong enough to weather the stress of deep and widespread insolvency among financial institutions unless information flows on regulatory performance can be reworked to align the incentives of financial regulators more closely with those of ordinary taxpayers.

I see financial institutions as information factories that create information networks among customers and other financial firms. Through these networks the institutions collect, verify, store, process, and transmit information for themselves and for customer accounts. Individual institutions and their information networks benefit from the coordination and contract enforcement services of outside regulators. These regulatory services transform rather than solve the vulnerability problems of the financial system. Conflicts in risk management incentives for managers of regulatory enterprises parallel those that face managers of financial services firms. At times of crisis, incentive conflicts in regulation typically induce top regulators to

Edward J. Kane is the Cleary Professor in Finance at Boston College.

Annual World Bank Conference on Development Economics 1996

reinforce rather than to offset the disposition of insolvent "zombie" institutions to take socially harmful risks.

Baseball players and jazz musicians stress the importance of bringing their efforts home. I believe that although Mishkin has given us a deft analysis of the role of asymmetric information in financial crisis, he has stopped his work one base (or one melody) short of home.

Mishkin rushes by the task of establishing accountability for regulatory performance in an environment of asymmetric information as if this were the easy part of the problem to solve. He presumes that incentives to forbear are improved by giving banking authorities nominal independence from the political process. But historical research on state-sponsored deposit insurance systems (Bodenhorn 1996; Calomiris 1990; White 1981) and on the politics of the U.S. Federal Reserve System suggests that independence becomes increasingly hard to sustain when politicians come under stress.

Economic analysis of regulators' incentive conflicts can explain both why regulators the world over resist accountability and what regulatory patterns and truth-telling requirements could improve accountability for regulatory performance (Boot and Thakor 1993; Kane 1995a, b). Before discussing the phenomenon of governmental resistance to accountability, let me reframe Mishkin's theory of financial crisis.

Asymmetric Information Theory of Financial Crisis

Like a batter who hits a triple in baseball, Mishkin touches three valuable bases. But rather than running to home plate, he projects his analysis home by drawing loose policy insights about how to avoid financial crises in developing countries.

The article's first base is a review of the post-1970 literature on asymmetric information and principal-agent conflict that seeks to explain why banks and banking regulation exist. Banks emerge as an efficient solution to society's need to observe and analyze what would-be borrowers take to be hard-to-communicate "inside information" about their future earnings. Banks solve these problems by establishing information-based relationships with customers (Kane and Malkiel 1965). These relationships provide a two-way flow of continuing benefits to banks and customers alike.

Government regulators are introduced as a way of overcoming a bank's incentive to engage in excessive risk taking. The article notes that incentives for government regulators to control bank risk taking may be distorted by careerism and by influence peddling. But it offers no solution to the age-old question of how society ought to monitor its monitors.

At second base the article explains how the adverse selection and moral hazard problems that regulators prove unable to control serve as a foundation for poor credit decisions. Each bad loan a bank puts on its books increases the bank's exposure to loss—and in a nontransparent way. The nontransparency sets up a game of hide and seek with creditors, outside auditors, and government examiners. As hidden bad loans multiply at a single institution, that *institution* becomes vulnerable to

macroeconomic shocks and customer runs. When hidden bad loans multiply across the banking system, that *system* becomes vulnerable to crises and panic.

At third base the model becomes an instrument for understanding a series of real-world crises: in U.S. history and in contemporary Mexico. The model tells a story of how banks contribute to macroeconomic fluctuations and crises that can be made plainer by resorting to metaphor.

I propose that we look at hidden bad loans as termites that eat away surreptitiously at a bank's net worth. To complete the metaphor, we may view a bank's true economic net worth as the beams supporting a roof that is supposed to protect taxpayers, customers, and other bank stakeholders from being soaked by occasional financial storms.

An institution falls into crisis when the weakened roof of its particular enterprise collapses in a storm. A wider financial crisis occurs when many bank roofs give way at the same time. The more important that the roofless banks are individually, the more extensive is the destruction that their troubles pass on to the real economy. That is so because damaged banks shift their priorities from servicing their customers' financial needs to rebuilding their own roofs. This reordering of bank priorities hampers the ability of household and business customers to lay their hands quickly on the resources they need to preserve or improve their own houses.

Even for banks that eventually rebuild their roofs and save their buildings, the flow of credit to information-relationship customers is impaired until the storm has passed. Customers whose relationship banks are trashed beyond repair are hurt more and hurt longer because they have to establish an information relationship with a new bank before they can fill their credit needs.

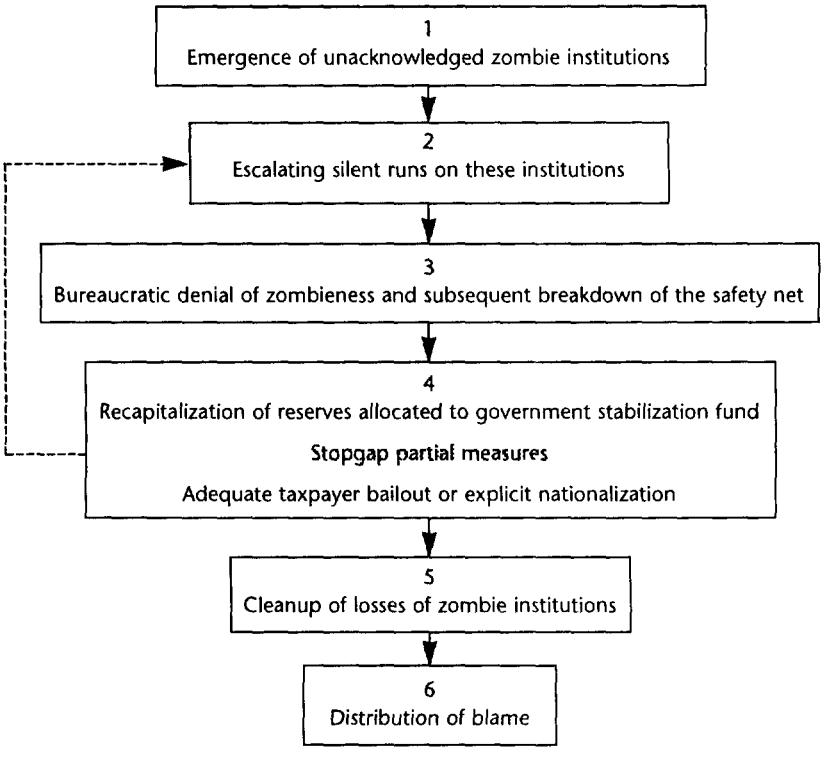
Any event that can sharply decrease the perceived value of financial assets can qualify as a metaphorical storm. Mishkin emphasizes sudden sharp increases in interest rates or in aggregate uncertainty. He offers a catalogue of unanticipated events that might have those postulated effects: failure of a prominent institution, recession, political instability, stock market crash, deterioration of corporate balance sheets, sharp unanticipated disinflation, unanticipated devaluation or exchange rate depreciation, and sharp decline in the demand for a region's real products.

Incentive Conflict in Financial Regulation

Jane Jacobs (1961) revolutionized planning for urban safety by emphasizing that the safest areas in city parks and streets were those that provided no opportunities for bad guys to hide. Dark places, picturesque nooks, and shrubs that block sight lines encourage evil doing by making crime hard to observe.

Economic analysis supports the view that when hidden insolvencies develop, the incentive incompatibilities inherent in representative democracy make government financial regulators inordinately quick to deny the importance of emerging problems and inordinately slow to take the actions needed to protect taxpayers (figure 1). The longer and deeper an institution's insolvency runs, the greater is the opportunity for well-informed uninsured depositors to move their funds away before the bank's

Figure 1. Six Stages of a Deposit Institution Mess



losses are resolved and charged against bank stakeholders. These "silent runs" allow sophisticated and well-connected private stakeholders to shift losses to other depositors and to unwary taxpayers. In time, silent runs create pressure on regulators to act. But the first round of regulatory action is usually too timid to stop the deposit outflow for long. Eventually, the pressure resumes and becomes strong enough to reveal the regulatory disarray.

Although silent runs act as forcing events, the central problem in forestalling bank insolvencies does not lie in making government guarantees credible. It lies in measuring and controlling the losses that credible guarantees and expedient delays pass on to healthy institutions and taxpayers. Around the world, but especially in developing countries, officials enjoy private benefits that conflict with the goal of minimizing taxpayer losses. Government officials have discretionary control over their reporting frameworks and a propensity to use this control to cover up emerging problems and to postpone effective treatment. As a crisis unfolds, officials can and do routinely distort information flows about the quality of their performance and repeatedly put off at least some of the painful adjustments that banks need and taxpayers deserve.

Getting better supervisory performance requires making the costs generated by regulatory forbearance observable so that regulators can be suitably disciplined in the press and in the labor market for post-government employment. As taxpayers'

exposure to losses mount, the press and academia must be able to observe and communicate to the public the facts about the accumulating exposures and whatever influence peddling has occurred. What is needed to complete Mishkin's framework is a scheme that establishes accountability for policy delays and other kinds of costly regulatory mistakes. Perfect accountability exists when the authorities are immediately and completely answerable for their actions. This requires that their decisions, their motives, and the consequences for taxpayers be transparent enough for outsiders to monitor and for labor markets and budgetary control systems to discipline.

Defects in public sector accountability do not occur by happenstance. Far from being embraced by real-world policymakers, accountability is systematically resisted. The economic function of defects in accountability is to insure the reputations of top government officials against the quick emergence of embarrassing facts and to permit unacknowledged tradeoffs to occur between overt policy goals and covert reasons for following policies that benefit politically powerful financial institutions at the expense of taxpayers. High officials and top institutional managers value opportunities to report on their performance in self-serving ways. They use these opportunities to alibi their way out of embarrassment by creating disinformational cover that convincingly misrepresents their true motives and the consequences they anticipate.

When facing an incipient financial crisis, government regulators (including elected politicians) resemble medical doctors who are subject to inadequate social controls against engaging in regulatory malpractice. Like doctors, the first responsibility of regulators should be to find out what is going wrong and to treat it promptly. But the instinctive response of most regulatory doctors is to ignore important symptoms, to rely on superficial accounting tests that are bound to yield inadequate information, and to prescribe only timid therapies.

Reducing Incentive Conflict in Financial Regulation

Three complementary mechanisms can lessen incentive conflict among banks, regulators, and taxpayers. One mechanism would extend limits on stockholder liability for creditor claims in bank insolvencies to two or three times the capital stockholders have invested in the bank. Extended stockholder liability was widely and successfully used in the United States in the years before federal deposit insurance was enacted.

A second mechanism would make greater use of opportunities for private participation in proactive supervisory processes for controlling bank loss exposure. The idea is to use co-insurance provisions to reassign to the private sector any responsibilities for loss control monitoring and loss resolution that can safely be stripped from government (Kane 1995b). Such schemes seek to unbundle risks that are privately insurable from the catastrophic risks that banks inevitably shift to society. Because free-rider problems are mitigated in private enterprises, stakeholder monitoring of co-insuring private monitors is bound to be more intense than taxpayer monitoring of government supervisory enterprises. Private co-insurance can also help to reduce and simplify the monitoring and loss control activities in the catastrophic risks that remain with the government and taxpayers. Poor private moni-

toring sends market signals that poor government monitoring does not and cannot trigger. These observable signals can alert society to mistakes in the backup government supervision of catastrophic risks.

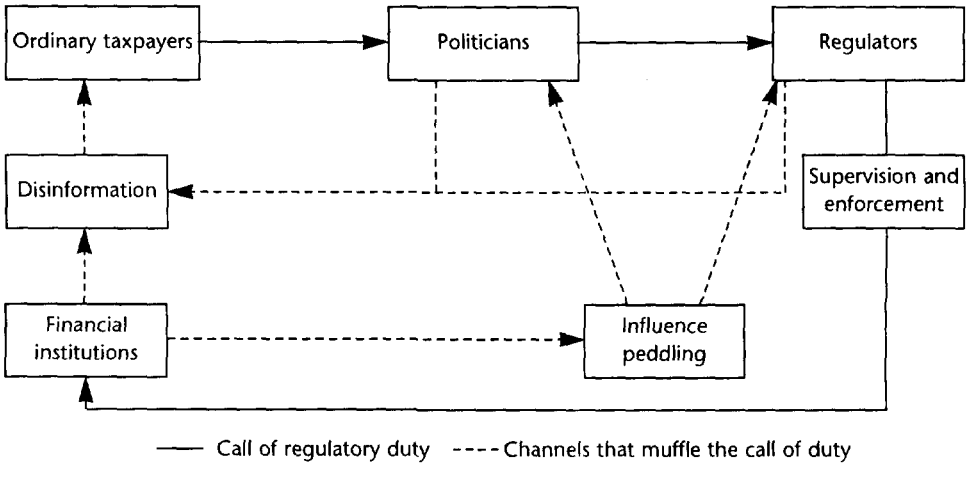
A third mechanism would make official measures of regulatory performance more informative, thereby increasing the accountability within government for supervisory decisions that affect taxpayers' exposure to loss in financial institutions. In countries with explicit deposit insurance, this can be done by establishing market-value accounting for the exposures that pass through to taxpayers from the insurance fund. And with or without explicit deposit insurance ethical reforms can be promulgated that aim simultaneously at clarifying the duties of loyalty, care, and competence that regulatory officials owe to taxpayers and at increasing the size and enforceability of penalties for violating regulatory duties.

A Summary Perspective

Informationally efficient capital and labor markets may be presumed to discipline bank creditors, stockholders, and managers. But capital markets do not trade debt or stock whose payoffs are based solely on the performance of government regulatory and supervisory enterprises. Nor do the contracts that labor markets offer to government officials reliably discipline officials after the fact for hiding evidence of poor performance.

A conscientious regulator routinely faces incentive conflict (figure 2). This conflict has intensified financial crises in the United States and Mexico alike. The solid lines in figure 2 indicate the call of regulatory duty; the dotted lines indicate the subtle and unsubtle channels through which this call can be muffled and even shouted down by disinformation and influence peddling. Whether regulators operate in an industrial or a developing country, the first step in weakening the lines of influence peddling is to acknowledge that they exist.

Figure 2. Chain of Delegation in Financial Regulation



References

Bodenhorn, H. Diran. 1996. "Zombie Banks and the Demise of New York's Safety Fund." *Eastern Economic Journal* 22: 21–33.

Boot, Arnould W. A., and Anjan V. Thakor. 1993. "Self-Interested Bank Regulation." *American Economic Review* 83: 206–12.

Calomiris, Charles. 1990. "Is Deposit Insurance Necessary?" *Journal of Economic History* 50: 283–95.

Jacobs, Jane. 1961. *The Death and Life of Great American Cities.* New York: Random House.

Kane, Edward J. 1995a. "Difficulties of Transferring Risk-Based Capital Requirements to Developing Countries." *Pacific-Basin Finance Journal* 3: 193–216.

———. 1995b. "Three Paradigms for the Role of Capitalization Requirements in Insured Financial Institutions." *Journal of Banking and Finance* 19: 431–59.

Kane, Edward J., and Burton G. Malkiel. 1965. "Bank Portfolio Allocation, Deposit Variability, and the Availability Doctrine." *Quarterly Journal of Economics* 79: 113–34.

White, Eugene N. 1981. "State-Sponsored Insurance of Bank Deposits in the United States." *Journal of Economic History* 41: 537–57.

Floor Discussion of "Understanding Financial Crises: A Developing Country Perspective," by Frederic S. Mishkin

Could it not be argued, asked a participant from the World Bank, that although deposit insurance might be inadvisable for industrial countries, it is important for mobilizing savings in transition and developing economies, where people need assurance that they can rely on the formal banking system?

Mishkin responded that the main reason governments provide deposit insurance is to have a safety net for banks and depositors. But deposit insurance is not the only kind of safety net. The problem with deposit insurance is that it tends to protect for idiosyncratic risk as well as for systemic risk. What society really wants in a safety net is protection mainly against systemic risk, and even then protection creates moral hazard.

Regarding Mishkin's comment that an increase in interest rates could worsen adverse selection, the participant from the World Bank asked whether Mishkin was talking about a generalized increase in rates or a sharp increase. A slow, general increase in interest rates can raise banks' profits, the participant said, because interest rates on loans generally increase faster than interest rates on deposits. But there is also a risk that higher interest rates will attract worse creditors. Mishkin responded that it depends on the model being examined. With some models there are problems once interest rates hit a certain threshold. With other models problems arise only when there is a sharp increase in rates.

Another participant from the World Bank asked whether a sudden increase in disclosure and information might not create a serious disturbance. People would realize just how bad things had been and would suspect that things were actually worse than they were being told. Mishkin said that when there is a problem, people tend to think things are worse than they actually are. Thus when there is disclosure—particularly during a crisis, when people already believe everything is going to crash—that usually helps ease the crisis. Only rarely would disclosure make things worse.

This session was chaired by Rachel Lomax, vice president and chief of staff, Office of the President, at the World Bank.

Annual World Bank Conference on Development Economics 1996

Edward Kane (discussant) added that the main benefit of disclosure is that it puts pressure on the authorities to act in a timely fashion. Returning to a metaphor he used in his comments, Kane said that if termites are destroying a house, the sooner they are stopped, the better things are for the structure of the house. The problem of time inconsistency, added Mishkin, is important in macroeconomics and even more important in regulation. Even though it might sometimes seem as though waiting is the better approach in a particular situation, that creates expectations that you will wait before taking action in the future as well. So one benefit of quick disclosure is that people know that the authorities will react quickly, which encourages them to behave prudently.

A consultant asked Mishkin if he thought that banks exhibit a herd instinct. The participant recalled a Marxist professor who had once asked him why a country would need more than one bank if all banks end up doing the same thing. In some cases yes, responded Mishkin, government regulation can produce a herd instinct. Banks supported by a strong government safety net know that as long as all the other banks in the system are in trouble, they will all get bailed out, which makes them more likely to offer the same kinds of loans that other banks offer. It was no accident, Mishkin said, that all the money center banks started making highly leveraged transaction loans and loans on developing country debt when they did. It was a consequence of their governments' adherence to the too-big-to-fail policy.

The consultant then asked Mishkin to comment on the tendency of banking systems in developing countries (such as Argentina and Mexico) to give large borrowers more favorable rates than small borrowers, despite the fact that in most countries (including the United States) small industry is the main generator of economic growth. According to Mishkin, asymmetric information theory explains this phenomenon very nicely: The large firms are the ones for which information is most readily available. For that reason, in fact, only the largest firms in industrial countries can issue securities. In developing countries the problem of large firms getting the lowest interest rates while small firms generate the most economic growth only adds to the urgency of the need to solve information problems. Developing countries must create financial systems with the right legal and disclosure infrastructure in order to produce economic growth.

Another participant asked whether the decisions of the currency board had made it easier or more difficult for Argentina to recover from its banking crisis. Mishkin said that it had helped, because maintaining convertibility—thereby preventing inflation or hyperinflation—was exactly the right approach. Argentina had a lot of dollar liabilities and a system that was based on a firm commitment to a fixed exchange rate. Given that system, if Argentina had tried to inflate out of the crisis, it would have run into the problems that Mexico had.

Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?

Gerard Caprio Jr. and Daniela Klingebiel

Since the late 1970s bank insolvencies have become increasingly common. Where these failures are systemic, they can drain a country's financial, institutional, and policy resources—resulting in large losses, misallocated resources, and slower growth. Using a new database covering some eighty-six episodes of insolvency, this article examines the causes and effects of these crises and how governments have responded. It finds that both macroeconomic and microeconomic factors have figured in bank crises and that, based on the criteria developed here, few governments have responded well to these episodes. To better manage insolvencies, policymakers must develop a regulatory framework that allows banks to respond more robustly to shocks and ensures proper management and oversight. That bankers have not regularly planned for shocks suggests that they have not had the incentive to do so.

> No degree of regulatory wisdom could or should have made the 1920s a profitable time for banks in [U.S.] agricultural regions affected by drastic declines in prices and land values. In the face of these shocks, some failures were inevitable. What regulation could have done, but did not do, was make the system as a whole less susceptible to shocks and more resilient in its response to failures.
>
> —Charles Calomiris (1992, p. 312)

The past ten to fifteen years have been to economists interested in banking and incentive issues what the 1840s and 1850s were to gold prospectors in California. After several decades—most of the post–World War II era—of relatively calm financial markets, in recent years there has been a profusion of banking

Gerard Caprio Jr. is lead economist in the Finance and Private Sector Development Division of the Policy Research Department at the World Bank. Daniela Klingebiel is in the Young Professionals Program at the World Bank. The authors thank Phil Brock, Charles Calomiris, Elena Folkerts-Landau, Patrick Honohan, Philip Keefer, Homi Kharas, Ross Levine, Millard Long, Herminia Martinez, Andrew Sheng, Mary Shirley, and Shahid Yusuf for comments. Conversations with the late Herb Baer helped initiate this line of research. Colleagues throughout the World Bank provided information on country cases. Alexander Tanzi supplied competent research assistance, and Daniele Evans kept the various drafts (and the authors) organized.

Annual World Bank Conference on Development Economics 1996

crises in a variety of industrial and developing countries, in many cases widespread enough to qualify as systemic. In absolute terms Japan will likely suffer the largest losses, with official estimates putting nonperforming loans in 1995 at about $400 billion (unofficial estimates reach $1 trillion, or about 25 percent of GDP). Argentina in the early 1980s likely saw the largest relative loss (estimated variously at 20–55 percent of GDP), with Chile not far behind (13–42 percent of GDP). In both cases losses appear to have been closer to the upper end of these ranges.[1] The lower end, typical of losses in many developing countries, does not capture the slowdown in economic growth when resources are driven out of the formal financial sector (and into less efficient uses) and stabilization programs are derailed.

Based on a first-ever compilation of data on insolvencies around the world since the late 1970s, we analyze the factors that have caused these crises, and how governments have responded.[2] Extending the line of research started by Sundararajan and Balino (1992), Brock (1992), and Sheng (1996), we examine a wider variety of cases and link macroeconomic and microeconomic factors in the debate about how bank insolvency arises and spreads. Most episodes of insolvency are caused by a mixture of bad luck, bad policies—both microeconomic (regulatory) and macroeconomic—and bad banking. But these factors are neither independent nor immutable. Although establishing an absolutely fail-safe banking system should not be the goal of reform—even if it were possible, it likely would be too expensive—banking systems can, as Calomiris (1992) suggests, be made more robust in their responses to bad luck and bad policies through a regulatory framework that rewards prudent and diversified risk taking by bankers.

Bank Solvency: Does It Matter?

Although many financial historians mark the rise of financial houses in thirteenth-century Italy as the origin of modern banking, one of the earliest crises on record goes back to 33 A.D., when a confluence of factors—the sinking of some ships loaded with uninsured commodities, a slave revolt, fraud, defaults on foreign debt, liquidity-draining government policies, and a bout of domestic and international contagion—shut down several banking houses in Rome (Calomiris 1989). Tiberius Caesar resolved the crisis by providing government funds to *reliable* bankers and certain debtors, forgiving some interest, and suspending government policies that had temporarily drained liquidity. Most of the institutions recovered.

This case notwithstanding, government safety nets were uncommon before the nineteenth century. Indeed, in many countries central banks were formally instituted only in the twentieth century. Until then market forces of one form or another were relied on to deal with bank insolvency, and depositors generally had no government safety net protecting their funds. Widespread bank failures were fairly common in the United States, in large part because many states had unit banks (banks with only one office). Their lack of diversification made these banks especially susceptible to failure, and indeed it was in states where such banks were common that deposit insurance schemes first became popular—and routinely failed. At the end of the

twentieth century government intervention (before and especially after episodes of bank insolvency) has become commonplace, suggesting a belief that bank solvency is important. Why is this the case, and how does bank failure differ from the failure of a nonfinancial firm?

Differences between Bank and Nonbank Firms

Insolvent firms in market economies generally find it difficult or impossible to raise new funds. This lack of capital precludes their acting on profitable investment opportunities and may force them to sell important assets. Insolvency also may distort the incentives of managers, making them more susceptible to fraud and moral hazard. At the very least insolvency reduces the incentives for owners or managers to exert effort consistent with the long-run health of the firm. Although small firms have a high failure rate in market economies, major nonfinancial firms seldom cease operations when they become insolvent because they usually have both a core group of activities that remains profitable and new investment opportunities. In addition, most large firms have sizable investments in fixed capital goods that often are difficult to resell. As long as the net present value of the firm's operating profits and tax loss carryforwards exceeds its liquidation value, it makes sense to continue operations—albeit with minimal new investment, divestiture of noncore activities, and installation of new management. Insolvent firms deal with the imbalance between assets and liabilities by marking down liabilities and equity to the new, lower value of the assets and the future cash flow. Equity holders usually see most of their claims wiped out, while debt holders have a portion of their claims converted into equity. Individual funding is provided only after assets have been marked down.

The imposition of losses on creditors plays an important role in this process. Creditors do not necessarily have the best information about the troubled firm. Nevertheless, they have a strong incentive to respond to a serious deterioration in the firm's performance by seeking its restructuring or liquidation. If creditors bore none of the losses, they would have little incentive to allocate credit to its best use or to monitor the performance of borrowers. Unprofitable firms would be able to continue operating undisturbed.

When banks become insolvent, many of these adverse consequences can be deferred. The most important factor accounting for this difference is that the output and production processes of nonfinancial firms often are more transparent than those of banks, reflecting both the information-intensive nature of banking and its intertemporal quality—most bank products or services include a promise to pay in the future, meaning that it can take time for a bank's inability to fulfill its contracts to become evident. Banks can conceal problems by rolling over bad loans or by raising more deposits and increasing the size of their balance sheets. Especially when depositors enjoy explicit or implicit protection, banks often can attract new depositors with the promise of high interest rates so as to increase their bets with current clients; look for new, high-risk, high-return areas; or work a Ponzi scheme.

The opacity of bank loans also means that they are harder to sell than the typical inventories of nonfinancial firms, so it is commensurately more difficult for banks to raise liquidity to restructure when a negative shock hits. Moreover, in selling off its (hard to market) assets the bank might be forced to accept "fire sale" offers—sales of assets at a lower price than they could command given a normal search time for the highest bidder or adequate time to convey sufficient information about an asset's real quality.[3] Thus banks that are in trouble have the incentive and the ability to delay loss recognition. When banks with a low or negative net worth remain open, bank owners or senior managers are less motivated to monitor them, so bank staff and officers have an ideal opportunity to engage in a variety of defalcations. The resulting tendency for insolvent banks to increase their losses has been widely noted and suggests the need for prompt, corrective action.

Thus banks differ from nonfinancial firms because of a combination of information imperfections and intertemporal contracting. Not only can banks have trouble evaluating borrowers, but the health of a bank (or a banking system) is difficult to discern because depositors, supervisors, and other outsiders are unable to see through the veil surrounding banks' balance sheets until it is too late (Simons and Cross 1991). This information problem, coupled with banks' demandable debt and sequential servicing features, makes banking inherently fragile and susceptible to runs: depositors' first indication of trouble can be a line of other depositors waiting to collect their funds. Since banks are part of the payments system, contagion could lead to a halting of payments and a return to barter, to the detriment of overall economic activity. The simultaneous selling of assets by many banks, prompted by common shocks or even rumors, can lead to large declines in quite visible asset prices (such as that of land) and thereby increase political pressure for the government to act. Money market mutual funds, which offer a par guarantee but hold only highly liquid and riskless (or low-risk) short-term assets, do not suffer from panics or runs, except possibly in the case of fraud. Mutual funds, which do not offer a par guarantee but rather are equity-like instruments, also are less prone to runs.

The possibility of contagion means that a single bank failure has not only a direct negative effect on GDP associated with the loss of the bank's profits and wages (as with any bankruptcy) but also an indirect and potentially larger effect to the extent that the bank failure leads to or is associated with other bank failures and the shutdown of the payments system. Widespread bank failure also can indirectly affect economic activity to the extent that bank closures lead to the drying up of information, as Bernanke (1983) argues was the case during the U.S. Depression. Benston and others (1986) contend that bank failures and limited (nonsystemic) bank runs have for the most part had no greater effect on the U.S. economy than have the failures of nonfinancial firms of comparable size, and that the losses generally were low in the pre–deposit insurance era.[4] Since the advent of a federal safety net, however, matters appear to have changed: fewer banks failed in the 1980s than during the Depression, but depositor losses per dollar of deposits were higher (White forthcoming). Even in the United States, however, and even before 1914 (when the

Federal Reserve System was created), the instability of individual banks has at times developed into concerns about the banking system as a whole.

Thus the importance of bank insolvency relative to that of nonbanks can be distinguished by the possibility of a systemic crisis. The real cost to the economy of bad loans, whether as part of a generalized crisis or as isolated problems, is the misallocation of resources. Although much of the lending supported by insolvent banks is thought to be underwriting productive investments, these banks' losses are evidence that this is not always the case. But since estimates of the real return on projects—both those undertaken and those crowded out by bad loans—are often lacking, observers rely on information about nonperforming loans, loan losses, or the cost of restoring insolvent institutions to solvency as a proxy for the cost of resource misallocation.[5]

Systemic bank insolvency also drives resources out of the formal financial sector and into less productive uses. In addition to a direct budgetary impact, widespread bank insolvencies can have the added cost of changing government policy in a negative direction, such as by derailing stabilization programs or by retarding or reversing financial and nonfinancial sector reforms. Developing and transition economies in particular are hard pressed to deal with bank insolvencies because they lack deep capital markets, which can spread the costs of insolvency over a number of years. Without this buffer, and with a more limited tax base, developing countries are more likely to resort to an inflation tax to finance banks' losses.

There is no objective, generally accepted definition of when a problem in the banking sector becomes systemic. Central bank governors tend to behave as though "they know a systemic problem when they see one" or as though a problem becomes systemic when those at the central bank think that an event could develop into a systemwide problem. This view is not as arbitrary as it first appears: contagion can occur when creditors believe that there is a problem, and central bankers may be in the best position to discern if a crisis is likely to become systemic, since they should have good information (through their links with commercial banks) on the likelihood of a run by the liability holders of the banking system. Moreover, the presence or absence of central bank support can determine the likelihood that a crisis will occur, though support that is too readily or too cheaply available may encourage excessive risk taking and larger losses. This uncertainty and the possibility of contagion suggest that a problem can have systemic implications even if only a small portion of banking system assets is impaired.

Types of Insolvency

There are three general types of bank insolvency: those limited to a single bank or a small number of banks, which clearly are not systemic; overt banking system runs; and a more silent form of financial distress. Overt runs are easy to recognize: a banking panic (that is, a systemic run) occurs when "bank debt holders at all, or many, banks in the banking system suddenly demand that banks convert their debt claims into cash (at par) to such an extent that the banks suspend convertibility of their

debt into cash" (Calomiris and Gorton 1991, p. 5) or take other actions to deal with the crisis, such as securing support from a lender of last resort. Overt runs happen suddenly and end quickly. Financial distress of the banking system, when a significant portion of the system is insolvent but remains open, is perhaps the most pernicious type of insolvency. This problem is relatively common in developing and transition economies, where bank runs are averted by explicit or implicit (for example, when the state owns a large segment of the banking sector) deposit insurance. Financial distress can persist for years, overlooked by weak supervisory and regulatory systems and obscured by bankers' ability to make bad loans look good by granting new loans (de Juan 1987). Distress can continue indefinitely, but it may progress into overt runs if the public begins to doubt the validity of a government guarantee or the authorities come to recognize the costs of misallocating resources and intervene to restructure or otherwise resolve distressed institutions.

Acknowledging the uncertainties in separating systemic from more limited banking problems, Bartholomew, Mote, and Whalen (1995) define systemic risk as "the *likelihood* [emphasis added] of a sudden, usually unexpected, collapse of confidence in a significant portion of the banking or financial system with potentially large real economic effects" (p. 9). This definition avoids any quantification and leaves scope for central bankers' judgment, as alluded to earlier. When is the collapse likely, or what probability level is sufficiently high? Since one important reason sudden crises occur is that banks are opaque, defining what constitutes a crisis or systemic event is inherently subjective.

For the purpose of this article the definition is narrowed further, to cases in which the net worth of the banking system has been almost or entirely eliminated. Even though systemic problems clearly can arise when a banking system has positive net worth, it should be easy to agree that if the banking system is insolvent—that it is, if loan losses are sufficient to wipe out the system's capital—then the problem is systemic. During the 1980s the capital in many developing countries' banks was less than 5 percent of assets. Under those circumstances, if nonperforming loans net of provisions were 10 percent of assets and if banks generally collected 50 percent on these loans, then losses would be sufficient to eliminate the banking system's capital. Since this collection rate and capital ratio likely are higher than the developing country average for the 1980s, and loan loss provisioning was limited, even a lower ratio of nonperforming loans could eradicate capital.[6] Moreover, estimates of nonperforming loans usually are biased downward, suggesting that this definition of systemic crisis is quite conservative. Thus our inclusion in the systemic crisis group of countries with ratios of nonperforming loans to total loans of 5–10 percent also appears conservative. Finally, erring to include a few banking systems with a small but positive net worth is sensible because low net worth appears routinely to be translated into negative net worth, no doubt reflecting incentives to "loot" the bank (Akerlof and Romer 1994).

Under this definition bank insolvency is widespread, with as many as eighty-six episodes in sixty-nine countries (box 1). Some judgment has gone into the compilation of the list, because data on losses are not available for many countries and

because official estimates often understate the size of the problem.[7] Moreover, virtually every transition economy belongs on this list at some stage in the transition process; to limit the number of countries with missing information, however, some were excluded. Including all the transition economies—for which we hope eventu-

Box 1. Major bank insolvencies

Systemic cases (most or all bank capital exhausted)

Africa
Benin (1988–90)
Burkina Faso (late 1980s)
Cameroon (1987–present)
Central African Republic (1980s and 1994)
Chad (1980s and 1990s)
Congo (1980s and 1991)
Côte d'Ivoire (1988–91)
Eritrea (1993)
Ghana (1982–89)
Guinea (1985, 1993–94)
Kenya (1985–89, 1992, 1993–95)
Madagascar (1988)
Mauritania (1983–93)
Mozambique (1987–present)
Nigeria (1990s)
Senegal (1988–91)
South Africa (1977)
Tanzania (1987, 1995)
Togo (1993–present)
Uganda (1994)
Zaire (1991–92)
Zambia (1995)

Asia
Bangladesh (late 1980s–present)
India (1994–95)
Nepal (1988)
Philippines (1981–87)
Sri Lanka (1989–93)
Thailand (1983–87)

Latin America
Argentina (1980–82, 1989–90, 1995)
Bolivia (1986–87)
Brazil (1990, 1994–95)
Chile (1976, 1981–83)
Colombia (1982–87)
Costa Rica (several instances)
Ecuador (early 1980s)
Mexico (1981–82, perhaps through 1990–91, 1995)
Paraguay (1995)
Uruguay (1981–84)
Venezuela (1980, 1994–95)

Middle East and North Africa
Egypt (early 1980s, 1990–91)
Israel (1977–83)
Kuwait (1980s)
Morocco (early 1980s)

Europe and Central Asia
Turkey (1982–85)

Transition economies
Bulgaria (1990s)
Estonia (1992–94)
Hungary (1991–95)
Latvia (1995)
Lithuania (1995–96)
Poland (1990s)
Romania (1990–93)
Russia (1995)
Slovenia (1990s)

Industrial countries
Finland (1991–93)
Japan (1990s)
Norway (1987–89)
Spain (1977–85)
Sweden (1991)

Borderline or smaller cases

Asia
Hong Kong (1982–83, 1983–86)
Indonesia (1994)
Malaysia (1985–88)
Singapore (1982)
Taiwan, China (1983–84, 1995)

Industrial countries
Australia (1989–90)
France (1994–95)
Germany (late 1970s)
New Zealand (1987–90)
United Kingdom (1974–76)
United States (1984–91)

Source: Caprio and Klingebiel 1996.

ally to have adequate data—would bring the number of countries covered to about ninety and the episodes to well over one-hundred.

Why Banking Systems Go Bust

A long and well-developed literature has been devoted to identifying a single cause of banking crises. This section reviews the debate, including some historical evidence, and sees what can be learned from our global database. The evidence both undermines single-cause theories of insolvency and finds a more important role for microeconomic factors than is commonly conceded.

Macroeconomic Explanations

Although many analysts have found microeconomic or financial causes for bank crises, a recent study of crises in Latin America emphasizes their macroeconomic roots (Gavin and Hausmann 1996). The authors do not say that banking crises are always and everywhere a macroeconomic phenomenon, but they do use the following analogy:

> Chains break at their weakest link, but that does not mean that the specific flaws in the weakest link fully explain why the chain broke; one needs also to understand what caused the tension on the chain. Indeed, strengthening weak links in the chain only works if one succeeds in identifying the weakest link before it snaps. . . . Tension is placed on the chain by economy-wide factors, including, in particular, macroeconomic developments. When macroeconomic forces place great strain on the banking system, the weakest banks are the ones most likely to fail, but it is the macroeconomic tension, as much as the weakness of individual banks, that causes the failures. (Gavin and Hausmann 1996, pp. 27–28)

This view echoes that of Gorton (1988), who finds that in the United States recessions cause bank panics but bank panics do not cause recessions—a conclusion that applies to overt forms of bank insolvency, is based on the pre–Federal Reserve period (when there was no federal safety net for banks), and depends on which episodes one defined as panics. By using an overt run as the defining event of a bank crisis, this approach might merely identify the denouement of a tragedy, as when a terminally ill patient checks into a hospital before dying. If instead the disease itself—unsafe and unsound banking—is defined as the crisis, then it is possible that crises begin long before a system collapses and might even cause or contribute to macroeconomic problems, as Bernanke (1983) contends was the case during the U.S. Depression. Since prolonged periods of financial distress are quite common in developing countries, this article explores the more general problem of bank insolvency in its various manifestations, with bank runs as a subset.

Whether macroeconomic pressures are uniformly the main cause of bank crises can be elucidated by considering two economies with similar macroeconomic

shocks and structures but with different outcomes in the banking sector. Perhaps the best examples are Canada and the United States. Both endured the severe shocks associated with the Great Depression, yet the United States saw 15,000 banks fail during the 1920s and 1930s, and Canada saw only one bank fail (in 1923).[8] The depression was less severe in Canada than in the United States, suggesting that bank failures make a difference (Haubrich 1990). The banking crises of 1873, 1893, and 1907 caused large losses in the United States, but no crises occurred in Canada (Bordo forthcoming). At the very least this example suggests that forces in addition to macroeconomic factors are worth analyzing in predicting, diagnosing, and assessing the costs of bank failures. Also, Benston and others (1986) note that between 1875 and 1933 the correlation between the failure rates of U.S. banks and those of businesses was a mere 0.24. They conclude that while "bank failures and the state of the national economy are intertwined . . . [they are] not very closely [intertwined] at all times" (p. 59). If macroeconomic explanations are not sufficient to explain bank crises, what other factors might lie behind bank insolvency?

Other Possibilities

A number of analysts have made theoretical attempts to uncover the main cause of banking crises. Guttentag (1994) and Bartholomew, Mote, and Whalen (1995) produced the most succinct summaries of the factors behind banking crises, including the well-known debt-deflation school of Kindleberger (1978), Minsky (1982), and Bernanke (1983), who contend that although bank crises may not cause recessions, they certainly make them worse. Poor information markets and weak banks make this type of credit crunch more likely in developing countries, particularly after major swings in relative prices (Caprio 1992). General uncertainty, asymmetric information, and speculative bubbles led by excessive credit growth are other commonly cited factors. Increasingly, financial liberalization is mentioned as a cause of bank insolvency, likely because many recent insolvencies came on the heels of reform attempts. And in many cases a variety of regulatory and bank-specific management factors—including asset-liability mismatches, insufficient diversification, connected lending, and fraud—are cited (see below).

It is often claimed that bank insolvency in transition economies is different from that in market economies, and certainly this assertion can be justified in the transition economy banking systems that experienced massive initial losses as a result of the bad debts—most stemming from loans made to uncompetitive state-owned enterprises—inherited from pre-transition regimes. All the transition economies dealt with these early losses, however, most often by inflating them away. Succeeding rounds of insolvency have appeared in many transition economies as a result of political interference, lending to state-owned enterprises, poor management, and a variety of other factors similar to those cited worldwide.

In all economies it is entirely possible and indeed likely that there is not one but a variety of causes of bank insolvency. For example, a macroeconomic shock or

heavily concentrated loan portfolio might cause one (or more) bank(s) to have low or negative net worth. Private owners then have less incentive to manage the bank well, and indeed where their liability is limited they have every incentive to gamble or loot the bank, as noted earlier. This tendency helps explain why losses in the U.S. savings and loan institutions skyrocketed from an initial $30 billion to a final $150 billion (Barth and Bartholomew 1992).[9] Where the supervisory and regulatory regime is not particularly stringent or the government intervenes directly in allocating and pricing credit, as in many developing countries, initial losses might stem from politically motivated lending, connected lending, or outright fraud. Governments often fail to confront the problem early on, however, and delays routinely lead to much larger losses. The common link between different types of bank failure is that initial losses, whatever their cause, often multiply when prompt corrective action is not taken.

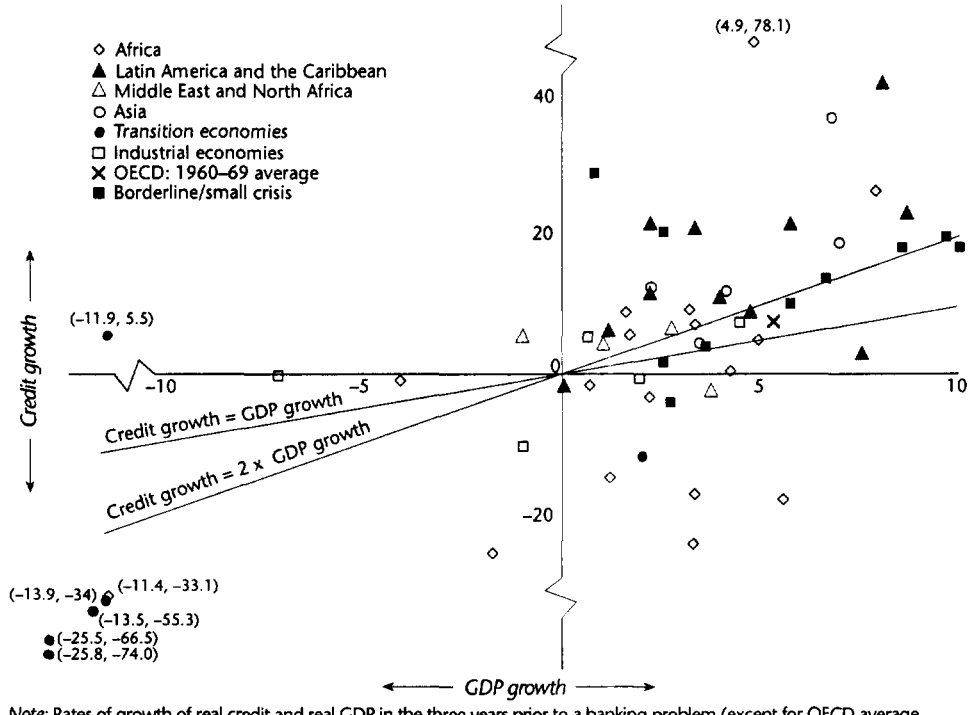
What the Data Say

The rest of this section examines what our data on a global set of banking crises can tell us about their origins. Because macroeconomic factors are more readily quantified, we look there first. Gavin and Hausmann (1996) argue that excessive credit growth is a primary factor behind banking crises. Although this appears to be true in Latin American crises, the link between credit growth and insolvency disappears when a broader sample is considered (figure 1).

Rapid credit growth often leads to or reflects a decline in the credit standards of individual banks. When a banking system grows rapidly, it is difficult for supervisors (or even bankers) to keep abreast of loan quality, since their information usually arrives late. The rate at which credit growth becomes excessive and the amount of time it needs to persist before being clearly indicative of a problem are difficult to gauge. But given the tendency of financial systems to deepen gradually as countries grow, real credit growth of one to two times GDP growth might be expected in normal times. A more definitive test of the role of credit would be possible if it were easier to identify with confidence a set of countries without banking problems. Doing so is a daunting challenge, however—consider that just a few years ago many observers viewed Japan's banking system as a model for others to follow; fewer would do so today.

With somewhat greater certainty, given both the favorable macroeconomic climate and the higher franchise value of banking, we selected OECD countries during the 1960s as a crisis-free sample. These countries not only fall within the "safe" zone in figure 1 (real credit growth of one to two times real GDP growth), they also do so with so little variation that it is not possible to display all twenty observations. Many of the borderline or less significant episodes of bank insolvency also lie near this zone. In addition to the Latin American countries above the safe zone of credit growth, many countries, particularly African and transition economies, show signs of a credit crunch. This crunch may reflect factors such as tight monetary policy, an aversion to lending on the part of banks because of a high-risk environment, or an episode of insolvency that began before the data shown here.

Figure 1. Real Credit and GDP Growth in the Three Years Preceding Bank Insolvency
(percent)



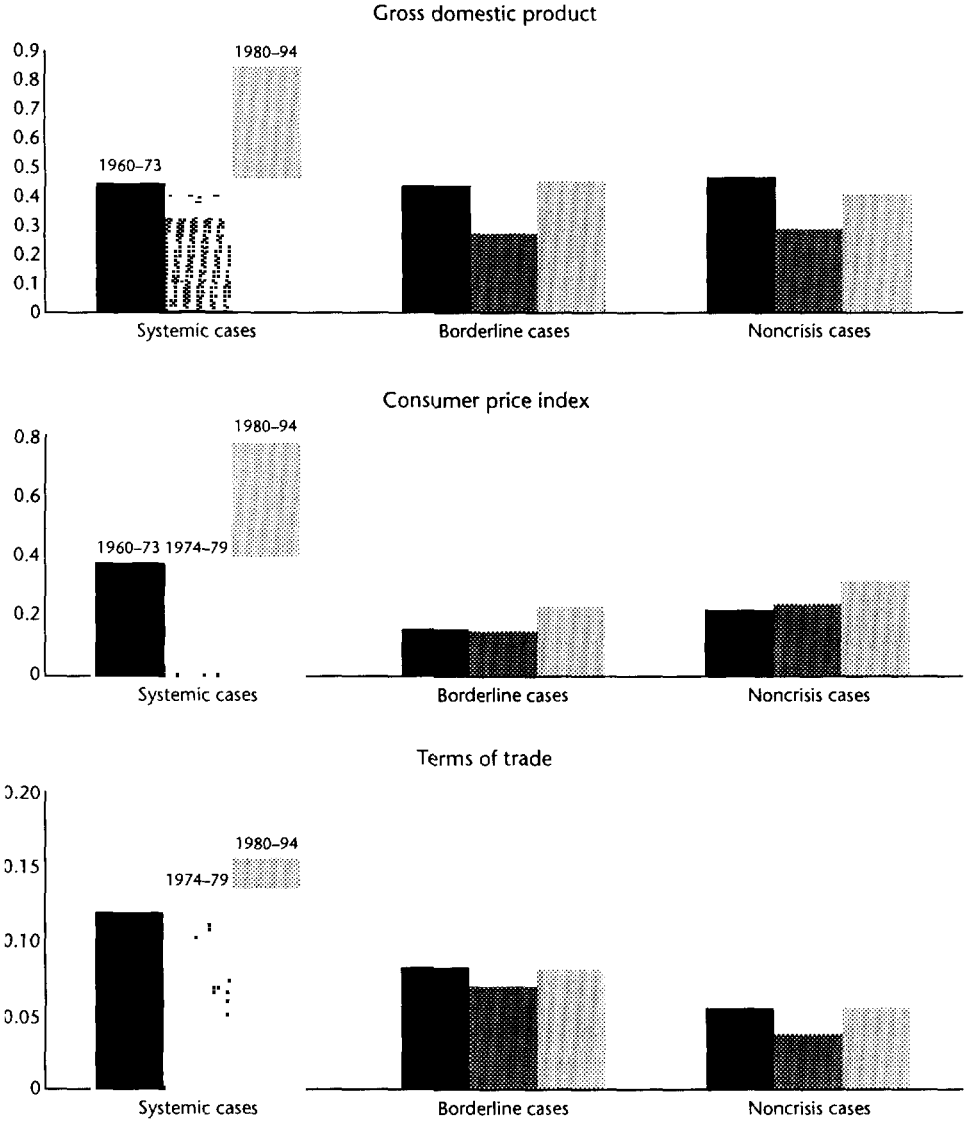
Note: Rates of growth of real credit and real GDP in the three years prior to a banking problem (except for OECD average in the 1960s). Credit growth refers to growth in bank credit to the nongovernment sector.
Source: Caprio and Klingebiel 1996.

From figure 1 one might conclude that officials should pay closer attention to credit growth and sound macroeconomic policy, though it is notable that what appeared to be excessive credit growth in Indonesia (and Japan) occurred at a time of single-digit inflation. It may well be that the monetary authorities should look at broader indicators of asset prices (which have little if any weight in consumer and producer prices) as well as at credit developments. But as the figure suggests, other factors also play a role in banking problems.

Volatility is another possible cause of crisis, since large variations in output and prices make it more difficult to pick good borrowers. Since 1980 crisis countries have seen greater volatility in output, inflation, and terms of trade (figure 2). To be sure, part of the volatility in GDP and inflation was caused by the banking crises. Less feedback effect might be expected in the volatility of terms of trade since relative prices are determined mainly by international factors. Also notable is that crisis countries experienced not only significantly greater volatility in terms of trade than did the other two groups but also significant shocks in the years prior to the episode. In 75 percent of the crisis countries the terms of trade fell by more that 10 percent in the years preceding the episode, with an average fall of 17 percent. Borderline crisis countries saw a 4 percent decline, while the noncrisis countries saw a 4 percent rise.[10] Many of the developing countries in our sample are highly concentrated in a

Figure 2. Bank Crises and Volatility

Average coefficient of variation



Source: Caprio and Klingebiel 1996.

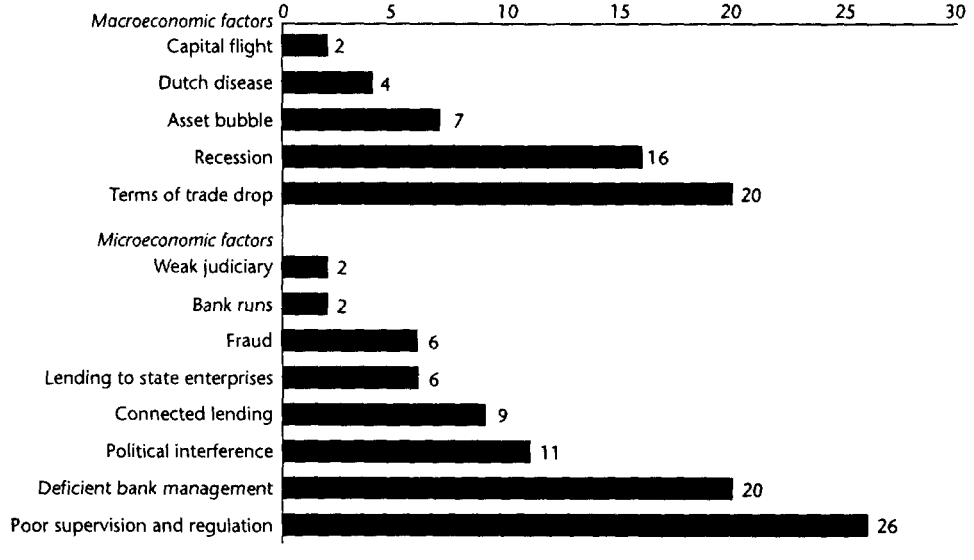
few commodities (Caprio and Klingebiel 1996) and, as suggested by the earlier comparison of U.S. states with no branching restrictions with states having undiversified unit banks, banks in small economies are highly vulnerable to external shocks.

In sum, a combination of greater volatility and shallow financial markets is particularly harmful, since it means that the shocks are more damaging and the means for spreading their costs over time are more limited. Figure 2 also suggests that the regulatory framework (such as Bank for International Settlements guidelines)

devised for noncrisis countries is likely to be inadequate for developing countries, as well as for some industrial countries that have experienced greater volatility.

Macroeconomic disturbances of any kind can weaken the nonfinancial sector and then show up in the portfolios of banks. But a poor (or wildly expansionary) macroeconomic climate need not be fatal to banks. The regulatory framework—which, along with a variety of institutional variables, plays a crucial role in defining the incentive system in banks and the environment in which they function—determines how resilient a banking system is in the face of shocks. Obtaining information on how incentives operate is difficult. But for the subset of twenty-nine episodes for which we could obtain better information (from published reports or interviews with experts) and which includes the most prominent and severe cases of banking sector insolvencies of the past fifteen years, the primary causes of bank insolvency are considered to be deficient management, faulty supervision and regulation, government intervention, or some degree of connected or politically motivated lending (figure 3).[11] Although large terms of trade declines figured in about two-thirds of the sample and recessions in about half, microeconomic factors were even more prevalent, indicating the presence of incentive distortions. Although events like capital flight or bank runs may have occurred more than twice in this subset, neither factor was thought to be important in most cases. And though many countries suffer from weak judicial systems—and improvements in this area likely would help instill better incentives—these deficiencies did not occur overnight and were rarely considered to be a main cause of insolvency.

Figure 3. Factors behind Twenty-Nine Bank Insolvencies



Note: Shows the number of times each factor was cited in twenty-nine country cases; twenty-nine is the maximum number of citations possible.
Source: Caprio and Klingebiel 1996.

This assessment does not help settle the issue of whether macroeconomic or microeconomic causes are more important, but it does suggest that even when macroeconomic disturbances are the main factors precipitating a banking crisis, it is by no means clear that they are the most important causes. The Gavin-Hausmann "chain" analogy at the beginning of this section breaks down to the extent that the links in the chain—the banks themselves—know that pressures on the links will occur frequently and unpredictably. In such a world putting all the blame on external forces is more an excuse than a justification (Fetter 1931). Macroeconomic factors often play an important but indirect role: a strong macroeconomic climate can help erode incentives for prudent banking—why sink resources into assessment or monitoring if repayment rates are high?—whereas a downturn exposes the results of poor management. And a regulatory framework designed to cope with routine shocks can prove inadequate if the shocks are sufficiently large. Thus authorities have to be continuously alert to the erosion of incentives in good times and prepared to act as a lender of last resort (only) in inordinately bad times.

A thorough analysis of why bank insolvency has increased in the past fifteen to twenty years and of why losses appear to be far larger than even during the Great Depression is beyond the scope of this article. But several factors have likely contributed. The last wave of banking problems was in the 1930s. During the thirty years following World War II systemic bank insolvency was for the most part thought to be a matter of historical interest—no doubt reflecting a good macroeconomic environment. Also, the spread of state ownership of banks allowed problems to be disguised. It is certainly plausible that greater macroeconomic volatility has contributed to the increased incidence of banking problems, particularly since the end of the pegged exchange rate system in 1973. Although figure 2 shows that crisis countries experienced less volatility in terms of trade than in inflation and output, many developing countries saw their terms of trade rise during the 1960s, whereas during the more recent period the trend was very much negative.

The increase in the number of separate banking systems during the postwar period—resulting from the end of colonization in many countries, the dissolution of several empires, and the breakup of the Soviet Union—has meant that there are more banking systems and more new banks but fewer skilled bankers or supervisors and relatively low capital levels, especially compared with the pre-Depression era—a good recipe for insolvency. Additionally, at the microeconomic level incentives for prudent risk taking in banking systems have weakened. In part this may be due to Kindleberger's (1978) warning that each generation needs its own follies—bankers who survive recessions or the bursting of speculative bubbles remember the lessons and become more conservative, but each new generation of bankers has to relearn them. Indeed, the recent wave of bank insolvencies might also be attributed to the good economic climate of the postwar period. By making it easier for borrowers to repay, this climate may have encouraged weaker credit standards and less investment in monitoring.

In addition, the lifespan of tightened credit standards can depend on the franchise value of bank licenses (Caprio 1996), which has declined in several industrial coun-

tries since the 1950s largely as a result of increased nonbank competition (Keeley 1990; Weisbrod, Lee, and Rojas-Suáraz 1992); in all likelihood greater profitability would have induced more prudent conduct. Moreover, bankers' willingness to take risk must have been increased by the spread of deposit insurance schemes, both implicit and explicit (Garcia 1995); that is, by the unwillingness or inability of governments to treat banks the way they treat (or should treat) nonfinancial firms. Thus it is also plausible that in addition to increased volatility, an increase in bankers' incentives to take risk explains why banking no longer appears to be so safe.

How Countries Have Responded

In contrast to the laissez-faire practices of the nineteenth century, governments in recent years (as in the time of Tiberius Caesar) have felt compelled to step in when bank insolvencies become widespread. Not all government interventions have been successful, and there are lessons to be derived from the more successful performers relating both to how the problems arose and to how they were resolved.

Evaluating the Success of the Restructuring Exercise

Measuring success and failure in finance is difficult because no single indicator reflects how well a financial system performs its various functions. Here we use four criteria to evaluate attempts to deal with bank insolvency:

- *Financial deepening.* Financial system deepening, as signaled by a rising ratio of M2 (or a broader monetary aggregate) to GDP, indicates that the banking system has stabilized in the aftermath of the crisis. A flat or falling ratio shows that problems remain in the banking sector.
- *Development of real credit.* Moderate growth of real credit indicates that the crisis has passed. Negative real credit growth may reflect a credit crunch,[12] while exploding credit growth (well in excess of twice real GDP growth and sustained for several years) may reflect distress borrowing from those in default.
- *Real deposit interest rates.* Real deposit interest rates above 10 percent are a strong indicator that banks are bidding up rates in order to stay afloat and signal that there is financial distress in the banking system (Brock 1995).[13] On the other hand significantly negative real rates on a sustained basis often signify a credit crunch or government intervention.
- *Recurrent problems in the banking system after restructuring.* The ratio of nonperforming loans to total loans should fall in the aftermath of a banking crisis. A flat or rising ratio (relative to the level during the crisis period) implies that banks' health did not improve on a long-term basis after restructuring. Unsustained recovery might have different causes—for example, banks did not stop lending to their borrowers in default or the government did not manage to bring fiscal, monetary, or external imbalances under control. A repeated episode of widespread insolvency is the surest sign that

attempts to cure the previous episode did not succeed; isolated episodes of bank failure might indicate success.

Of the eighty-six episodes of bank insolvency listed in box 1, data on these four indicators of success are available for sixty-four. We scored a country's banking sector performance in each of these areas as a one (performance consistent with a sound financial system) or a zero (there may still be problems). A country that scores a four after restructuring is considered a clear success, at least for now; just as central bankers can never eliminate inflation but have to be ever vigilant about its revival, regulating a banking system and responding to insolvency requires constant attention. A score of two or three indicates a mixed result, which could go either way, while a zero or one means either that the response was deficient or that there has not been sufficient time since the episode to see clearly the fruits of the policy response. This last point should be stressed: given that many countries have encountered insolvency problems during the 1990s, it may well be too early to judge success or failure. For that reason we focus below on four cases that occurred in the 1980s.

Very few of the sixty-four cases are clear successes, and among developing countries only Chile and Malaysia make this category (table 1). Twenty-eight restructuring exercises achieved mixed results, indicating that some problems persisted in the banking system. Among these cases Benin, New Zealand, the Philippines, and Thailand score a three (see Caprio and Klingebiel 1996 for detailed results). An equal number of cases have yet to be resolved or were unsuccessful. Of this group five countries—the Central African Republic, Chad, the Congo, Guinea, and Tanzania—have very low scores, signaling that governments still need to tackle fundamental problems in the financial sector. In some of the cases yet to be resolved the success of the restructuring efforts is at least questionable. For example, in Venezuela a comprehensive plan that addresses deficiencies in the banking system is still being adopted. Hungary's repeated recapitalization efforts may prove to be effective but at this point constitute a warning sign that underlying problems may not yet be resolved.

By contrast, Estonia's situation looks more promising, with the government sending the signal that losses will not be readily covered by the authorities. A new external and internal incentive system has depositors bearing the burden of losses and imposes tough exit requirements for banks. Among other transition economies Poland has made noteworthy improvements, especially evident in the increased technical capacity of Polish banks (partly a result of a twinning program with Western commercial banks) and in double-digit capital-asset ratios and relatively sophisticated workout units. Among developing countries Argentina has made considerable progress in bracing its banking system against shocks. This increased stability became evident during last year's "tequila" crisis, which the Argentine banking system weathered relatively well. Besides making progress on the macroeconomic front, the Argentine government has put in place a regulatory system with firm exit policies (seventy-six banks closed or merged in the past two years) and moved toward more market-based monitoring of the banking system by subjecting banks to greater disclosure requirements. In both Argentina and Poland the establishment and

Table 1. Evaluating the Restructuring Exercises

Successful	*Mixed results*	*Unsuccessful/not yet resolved/insufficient time*
Africa		
	Benin (1988–90)	Burkina Faso (late 1980s)
	Côte d'Ivoire (1988–91)	Cameroon (1987–)
	Ghana (1982–89)	Central African Republic (1980s and 1991)
	Madagascar (1988)	Chad (1980s and 1990s)
	Mauritania (1983–93)	Congo (1980s and 1991–)
	Senegal (1988–91)	Guinea (1985)
	Zaire (1991–92)	Kenya (1985–89)
		Mozambique (1987–)
		Nigeria (1990s)
		Tanzania (1987, 1995)
		Togo (1993–)
		Uganda (1994)
		Zambia (1995)
Asia		
Malaysia (1985–88)	Bangladesh (late 1980s)	India (1994–)
	Nepal (1988)	Indonesia (1994–)
	Philippines (1981–87)	Japan (1990s)
	Singapore (1982)	
	Sri Lanka (1989–93)	
	Taiwan, China (1983–84)	
	Thailand (1983–87)	
Latin America		
Chile (1981-83)	Argentina (1980–82)	Brazil (1994–)
	Argentina (1989–90)	Costa Rica (several instances)
	Bolivia (1986–87)	Mexico (1994–)
	Colombia (1982–87)	Paraguay (1995–)
	Mexico (1981–91)	Venezuela (1994–)
	Uruguay (1981–84)	
Other		
Australia (1989–90)	Egypt (early 1980s)	Bulgaria (1990s)
Germany (late 1970s)	Finland (1991–93)	Estonia (1992–94)
United Kingdom	New Zealand (1987–90)	France (1994–95)
(1974–76)	Norway (1987–89)	Hungary (1991–)
United States	Romania (1990–93)	Latvia (1995)
(1984–91)	Spain (1977–85)	Lithuania (1995–96)
	Sweden (1991–93)	Poland (1991–)
	Turkey (1982–85)	Russia (1995)
		Slovenia (1990s)

Note: Countries were rated on the basis of four criteria: financial deepening, development of real credit, real deposit interest rates, and recurrence of problems after banking sector restructuring.
Source: Caprio and Klingebiel 1996.

maintenance of incentive-compatible regulatory frameworks would likely put these countries among the ranks of our more successful cases.

Highlights of Four Cases

To draw solid lessons from countries' experiences with banking crisis and reform, enough time needs to have passed since the crisis for policy changes to have taken root. Here we look at four cases that range from clear success to clear failure. Chile and Malaysia can be considered among the more successful cases. Both countries experienced financial system deepening, moderate real credit growth, reasonable real deposit rates (that is, consistent with a healthy financial system), and no recurrent banking problems. Ghana's bank restructuring exercise achieved mixed results. The financial system did not deepen in the aftermath of the restructuring, and real credit growth was below real GDP growth—both indicators of possible ongoing problems. But real deposit interest rates were moderately positive and therefore at a level consistent with a healthy financial system. The less positive cases can be attributed mainly to the government's failure to restructure state-owned enterprises, which accounted for many of the nonperforming loans, or to maintain macroeconomic discipline. Guinea's banking system restructuring in the 1980s was a clear failure. The ratio of M2 to GDP remained flat, real credit growth was excessive, real deposit rates reached double-digit levels in 1994, and the banking system experienced recurrent problems. The restructuring was unsuccessful because the government failed to implement an adequate accounting and regulatory framework and to bring the budget deficit under control.

While Malaysia's restructuring affected only a small portion of the banking system, almost all the banks in Guinea were bankrupt in 1985, with just one institution—accounting for 1 percent of financial system deposits—remaining open. In Ghana and Chile insolvent institutions controlled almost half of total assets. Interestingly, the two successes represent one of the smallest (Malaysia) and one of the largest (Chile) insolvency episodes in terms of resolution cost (Caprio and Klingebiel 1996).

In Guinea the resolution mechanism chosen was liquidation by the government, with large depositors forced to bear some of the losses. In Malaysia the authorities opted to merge insolvent deposit institutions with healthier ones and to inject capital into marginally solvent institutions (with support from current shareholders). In Chile and Ghana the government decided to restructure banks on a case-by-case basis by transferring a large part of nonperforming assets to the central bank (Chile) or to a special loan recovery agency (Ghana). In Chile the banks remained responsible for recovering loans and were required to buy back their loans from the central bank on a face-value basis.

By resolving incentive problems in the external and internal framework of banks, the Chilean and Malaysian authorities managed to stabilize their banking systems on a long-term basis. Increased disclosure figured prominently in Chile, including the requirement for two private ratings of banks each year and timely publication of

information on their financial condition, with significant penalties for noncompliance (Brock 1992, pp. 199–216). In Malaysia, which has not featured greater disclosure requirements, the burden on supervisors is correspondingly greater. In both countries shareholders incurred losses and managers were fired, emphasizing the message that poor performance would be costly. Both Chile and Malaysia made efforts to contain or shrink macroeconomic imbalances—bringing the budget deficit under control, lowering inflation, and devaluing their currencies. Both countries also significantly improved their regulatory frameworks, brought accounting rules up to international standards, and strictly enforced the changes. To improve banks' internal incentives, the authorities in both countries barred banks from lending to borrowers in default, changed senior management in the restructured banks, and made attempts to collect on written-off loans.

By contrast, the authorities in Guinea and to a lesser degree in Ghana failed to address problems in the incentive structure of banks. In Guinea this inaction led to a second crisis less than a decade later, and the country still has not managed to reduce its budget deficit significantly. In Ghana macroeconomic control has eased since 1992, with potential adverse consequences for the banking system. Although Ghana was more successful than Guinea in adopting a stronger regulatory framework and tougher accounting standards, it is not yet clear how well these changes are being enforced. Ghana does, however, require its banks to undergo annual audits by outside auditors; the Guinean authorities have made no such provisions. It is not clear that Ghanaian or Guinean banks have halted lending to borrowers in default or made serious attempts to collect written-off loans. Although some change in management occurred in the Ghanaian banks, at least one bank manager just switched banks. A twinning arrangement with foreign banks was adopted, however. In Guinea the banks were liquidated, and no information is available on former managers.

An important question for future research, in particular given the crucial role of financial systems in economic development, is why so few countries have managed to cope successfully with bank insolvency. Answering this question will require in-depth case studies of the political economy of bank resolution and restructuring. As with public enterprise reform, it will be important to see whether factors that influenced the desirability, feasibility, and credibility of reform can be identified in more and less successful cases (World Bank 1995).

Dealing with Crises: The Case for Inoculation

Though we have noted several factors, both macroeconomic and microeconomic, that have played a role in generating the large losses incurred by banking systems worldwide, it remains difficult to know when a banking system is having problems. Data often are poor and bank problems are easily disguised, so dating crises or engaging in causality testing is necessarily precarious. Moreover, bank insolvency often occurs when macroeconomic volatility is on the rise and when incentives appear to have weakened—not a good time to try to determine causality. When financial deregulation occurs at the same time it becomes a convenient target,

though in many countries deregulation proceeded faster than improvements in financial infrastructure and incentives, suggesting that weak financial infrastructure and poor incentives led to the crises. Thus, although macroeconomic factors are important, microeconomic and incentive factors likely are key to determining the magnitude of banking problems and in some cases are even the main cause.

Developing a More Robust Regulatory Framework

The challenge, then, is to devise a regulatory framework that enables the banking system to be a more resilient absorber of shocks. Just as modern medicine cannot cure the flu but can lessen the consequences of severe cases, modern financial economics cannot, without risking the patient, devise an absolutely fail-safe banking system. Banks will face shocks, some banks will fail, and when the shocks are large and policy errors compound them, failures will become systemic. Moreover, the goal is not to have an absolutely safe banking system, but rather to have one that performs its intermediation function well, that is, that undertakes the funding of a diversified bundle of prudent risks. What a regulatory framework can and should do is support this function by providing bankers, shareholders, and perhaps depositors with the incentives to plan for shocks, in effect raising the shock threshold for producing a systemic banking problem. Once a problem occurs, the framework should encourage prompt recognition and early resolution, since delay usually leads to larger losses. That bankers have not routinely planned for larger shocks suggests that they have not had the incentive to do so. As the comparison between the Canadian and U.S. experiences highlights, it is quite possible to have a banking system that absorbs shocks rather than magnifies them.

There is no single model for a banking system that encourages bankers to engage in prudent risk taking. However, most countries need to move toward increasing the loss exposure of bankers through higher capital ratios, higher liability limits for bank owners, mutual liability (making groups of banks liable for member losses), or higher franchise values (making greater bank profitability the carrot that induces greater prudence)—all of which needs to be coupled with better information and better disclosure so that outsiders can effectively monitor banks. Bank managers who face a credible threat of removal for poor performance (and legal action, where warranted) are more likely to behave prudently, as are bank owners who risk seeing their capital wiped out and their ability to control failed institutions ended. In addition, two plans developed by Chicago economists (some from the University of Chicago, some from the Chicago Federal Reserve) have been suggested to cure banking problems. The narrow banking plan, developed in 1933, would limit banks' investments to safe, highly liquid instruments, leaving other intermediation to less regulated institutions whose liabilities are not guaranteed. A more recent plan would require banks to issue concentrated amounts of subordinated debt, thereby creating a group of highly motivated investors to oversee banks and effectively serve as owners. Although these different proposals are interesting to debate (see Caprio 1996), no single recommendation is likely to be consistent with the

institutions and political forces of all countries. But the goal—a regulatory framework that places greater emphasis on market forces—is the common thread of all these plans and is, in our view, the indispensable ingredient of a better banking system.

Are any of these options politically realistic? Large banking crises tend to make such options more appealing; the authorities in Argentina, Chile, Hong Kong, and the United States all improved incentive frameworks following banking crises. One way to improve the political appeal of better incentive systems is to make sure that some classes of bank liability holders have their own funds at risk. It may be unrealistic to limit deposit insurance to small deposits, since large deposit holders often wield significant political influence. But it may be possible to put some part of all depositors' funds at risk, thereby creating an interest group in favor of better supervision and regulation. Without such a motivated group, improvements in incentives may be temporary, since the memory of large crises often fades quickly and "dynamic" bankers become popular figures.

Even with an incentive-compatible regulatory framework, bad luck and bad macroeconomic policies cannot be eliminated, and they will occasionally threaten even the strongest banking system. Prompt action by current or potential owners, creditors, and supervisors will help limit losses and lead to a healthier financial system. But if a country's banking system is still threatened by even small shocks after incentives have been improved, its residents will avoid the formal banking system, a common occurrence in many small developing countries. Here the goal of reform should be to increase the size of the shock that will destabilize the banking system by allowing or even requiring banks in small economies to diversify. Most developing countries are small in terms of market size, and the smallest countries had the largest terms of trade declines (on the order of 20–86 percent, the largest in our sample) and the greatest volatility. Thus requiring banks to invest solely in their domestic economy weakens the banking system. This factor is more important for smaller and more concentrated economies, meaning that a greater role for foreign and regional banks must be fostered. Small economies need even better incentive frameworks and enhanced diversification.

A Role for the World Bank

Developing country banking systems are experiencing widespread problems that need to be dealt with since finance, particularly banking, is so important to long-run economic development (King and Levine 1993; Schiantarelli and others 1994). The World Bank and other multilateral institutions clearly must do something. The size of banking system losses shows that resources are being misallocated on a large scale and that growth is being commensurately reduced. And because governments are devoting funds to fixing banking problems, these funds are not available for health, education, or other programs. Because incentives are a large part of the problem, however, major financial sector lending is not recommended unless developing country authorities adopt a more incentive-compatible regulatory framework. Thus

the World Bank may well have to reject requests for financial sector loans when countries' banks are insolvent.

For countries not yet ready to significantly upgrade their incentive framework—and here there is room for judgment in determining how much action is enough—more emphasis is needed on building financial infrastructure (accounting, auditing, legal systems, banking and supervisory skills) and educating member countries on the need for incentive reform. These efforts will allow banks to better screen and monitor their clients and more easily enforce contracts. Encouraging the elimination of real sector distortions, a major goal of the past decade of World Bank lending, will reduce erroneous signals about the long-term viability of projects and borrowers.

Constructing a long-term (at least ten-year) plan for a country's financial system should be an integral part of the World Bank's financial sector lending strategy. If all World Bank work in the financial sector began with a review of how incentives are working in the financial system and how any proposed program would alter them (an incentives impact statement, if you will), more effective assistance over the medium term likely would result. Moreover, the World Bank needs to be more effective in gathering data on banking and on banks' clients, not just to predict upcoming crises but also to evaluate the impact of its financial sector interventions. Without firm-level data on banks' clients, accurate evaluation of financial sector interventions is impossible.

To be sure, no one endorses lending into a poor incentive framework. The point here is that merely encouraging governments to adopt an 8 percent capital adequacy ratio and to improve bank supervision is not sufficient. Developing countries face larger volatility than industrial countries, their economies are more concentrated, and they face sizable hurdles (as do industrial countries) in effectively supervising banks. If the World Bank limited itself to a few financial sector operations in countries with significantly improved incentive frameworks, it likely would increase the number of clear success cases both directly and, through imitation, indirectly.

A difficult but necessary change is that the World Bank must be willing to tell a country that it is simply too small to have its own banking system, if by that the authorities mean one solely or principally composed of domestic banks investing in the home economy. In very small economies the required incentive system would be difficult to achieve without foreign diversification. Thus promoting a greater role for foreign banks, including regional banking, is needed—a difficult idea to sell in many countries.

More research would help to determine how much diversification is needed and how to diversify safely. Another interesting research question is how much the regulatory system needs to favor more conservative banking before the growth-reducing effects of this approach outweigh the savings from sounder banking. At present there is neither sufficient variation in regulatory practice nor accurate enough information on banking to evaluate this issue. It appears safe to say, however, that most countries suffer from excessively imprudent banking because of governments' willingness to underwrite losses, making the exposure of bank owners quite low. Research on how outcomes in the banking sector vary with the scope of

deposit insurance coverage would be particularly useful, as would an effort to better understand the political economy of bank restructuring. With reliable data, research also could shed light on the interactions among concentration of bank portfolios, volatility of asset prices, and prudent capital-asset ratios.

Achieving better banking will require the use of both carrot—in the form of profitable opportunities for banking—and stick—such as prompt replacement of bad managers, substantial losses for (and replacement of) owners, and more mobile deposits. Such mechanisms will ensure that bankers take risks, but only risks that are prudent. The breadth and depth of banking problems reviewed in this article suggest that bigger carrots and bigger sticks are needed. If there is a bright side to this experience, it is that authorities and citizens, being better informed about the scope of the problems, might be more apt to reconsider the rules of the game in banking now than at any time in recent history. The challenge to the development community is to not let this opportunity pass without an effective response.

Notes

1. Figures for losses are hard to come by and estimates range considerably. In many cases there are no clear, uniform accounting standards; banks can disguise losses for some time by granting additional credit, losses can be apportioned to a variety of groups, and governments can pay for losses in a variety of ways, from injecting funds directly into insolvent banks to helping the borrowers (such as with cash, import licenses, monopoly power, and so on). See Marshall and Schmidt-Hebbel (1994) and Rodriguez (1994) for the higher estimates and Rojas-Suárez and Weisbrod (1995) for the lower ones. The main difference is that the higher estimates attempt to capture the entire fiscal cost, including exchange rate subsidies, while the lower estimates include only funds, credit, and bonds injected directly into the commercial banks.

2. Background data for this article is presented in Caprio and Klingebiel (1996).

3. A fire sale is more likely if the bank or banks in trouble hold a significant portion of an asset, such as in the case of a small town bank's real estate portfolio or Japanese banks' holdings of land and buildings, since the asset sale will then have a significant impact on prices. In the absence of explicit deposit insurance, depositors may even lose from fire sale losses—that is, initial losses can induce banks to dump good assets on the market to pay off early depositors, leading to lower prices for these good assets, more depositors demanding redemption of their funds, and so on.

4. Between 1865 and 1919 an average of 64 U.S. banks failed each year, with a standard deviation of 71 (Benston and others 1986, p. 58). These figures rose to 1,070 and 966, respectively, between 1920 and 1933. For the most part these were small banks, with only a tiny fraction of deposits affected. Limits on nationwide branching directly contributed to the large number of failed institutions—at their peak in 1920, there were 30,000—and their individual fragility. In other industrial countries and in developing economies the regulatory structure permits or encourages far fewer banks, so the failure of one or a few is likely to have more significant effects on the financial system.

5. Nonperforming loans are those for which interest payments are in arrears by a certain number of days. Although there is no internationally recognized standard for this time limit, bank supervisors often use ninety days. Loan losses are the actual loss on loans that are written off by banks, which is the outstanding loan amount less the value of what is recovered (including any collateral) minus collection expenses.

6. Because there are no comparable capital ratios or provisioning systems for developing country banks, this statement is based on impressions gathered from developing country officials, banking consultants, and World Bank staff. Note that not only is deciding on whether an episode is systemic difficult in a private banking system, but in countries with a significant proportion of bank assets in state-owned banks, capital in the banks may be little more than an assumed claim on the treasury. Although the incentive system may be different in public and private banks, bureaucrats often engage in the same cover-ups of bad loans that private bankers have employed. Finally, as bank capital ratios rise, as they have in the 1990s, this definition will have to be revised.

7. In cases where net capital was positive according to official data but negative based on statements of experts, we sided with the experts. Decisions on when one crisis ends and another begins are necessarily

arbitrary; hence it is possible to differ on whether some cases represent multiple crises or a continuation of the preceding one. Although it would be possible to use the indicators of success (see the next section) to determine the end of a crisis, that would mean that many cases would involve ongoing crisis.

8. Kryzanowski and Roberts (1993) argue that some or all Canadian banks were insolvent but were kept open by the government's implicit guarantee. This case is difficult to prove because banking operations generally are opaque and because Canada's banking system historically relied on mutual liability and a relatively small number of nationwide banks.

9. Thus the finding by the U.S. Comptroller of the Currency, the U.S. Federal Deposit Insurance Corporation, and others that fraud and violations of the law figured prominently in U.S. bank failures from the 1860s to the 1970s may merely capture the final stages of a drama in which initial losses were magnified.

10. Clearly, it would be optimal to have a longer time series of data on the precise size of bank crises and then to investigate the relationship between the size of the crisis and the change in and volatility of the terms of trade. It would then be possible to say whether greater volatility in the terms of trade led to more or more costly bank insolvency. Although we do not have such information, the data in figure 2 are at least persuasive.

11. Because the data underlying the descriptions in figure 3 are drawn from a wide variety of sources familiar with the sample countries and their financial systems, the descriptions it uses are somewhat generalized. Poor supervision or regulation is used to denote easy entry, little or no information gathering on bank solvency on the part of the supervisory agency, or no follow-up even if useful information was collected. In many cases deficient management refers to little or no monitoring of borrowers, few internal controls, or basic asset-liability mismatches. Connected or politically motivated lending and fraud need no elaboration.

12. A credit crunch may arise in response to several events: a run on deposits that, if not offset by monetary policy, forces a contraction in total assets and total lending; a postcrisis tightening of regulatory capital requirements; because the resolution mechanism reduces the banking system's information capital; or as a voluntary response to a reduction in the banking system's capital cushion or an increase in the risk of its current portfolio.

13. Positive but moderate real rates do not necessarily mean that there is no financial distress, since they may be restrained by official or informal controls.

References

Akerlof, George A., and Paul Romer. 1993. "Looting: The Economic Underworld of Bankruptcy for Profit." *Brookings Papers on Economic Activity* 2. Washington, D.C.: Brookings Institution.

Baer, Herbert, and Daniela Klingebiel. 1994. "Systemic Risk When Depositors Bear Losses: Five Case Studies." In George G. Kaufman, ed., *Banking, Financial Markets, and Systemic Risk.* Greenwich, Conn.: JAI Press.

Barth, James R., and Philip F. Bartholomew. 1992. "The Thrift Industry Crisis: Revealed Weaknesses in the Federal Deposit Insurance System." In James R. Barth and R. Dan Brumbaugh, eds., *The Reform of Federal Deposit Insurance.* New York: Harper.

Barth, James R., Philip F. Bartholomew., and Carol J. Labich. 1989. "Moral Hazard and the Thrift Crisis: An Analysis of the 1988 Resolutions." In Federal Reserve Bank of Chicago, ed., *Proceedings of the 25th Annual Conference of Bank Structure and Competition.* Chicago: Federal Reserve Bank of Chicago.

Bartholomew, Philip F., Larry R. Mote, and Gary Whalen. 1995. "The Definition of Systemic Risk." U.S. Office of the Comptroller of the Currency, Department of Economic and Policy Analysis, Bank Research Division, Washington, D.C.

Benston, George J., Robert A. Eisenbeis, Paul M. Horvitz, Edward J. Kane, and George G. Kaufman. 1986. *Safe and Sound Banking: Past, Present, and Future.* Cambridge, Mass.: MIT Press for the American Bankers Association.

Bernanke, Ben. 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review* 73(3): 257–76.

Blass, Asher A., and Richard S. Grossman. 1995. "A Costly Guarantee? The 1983 Israel Bank Shares Crisis Revisited." Discussion Paper 95.05. Maurice Falk Institute for Economic Research in Israel, Jerusalem.

Bordo, Michael. Forthcoming. "Regulation and Bank Stability: Canada and the United States, 1870–1980." In Gerard Caprio, Jr. and Dimitri Vittas, eds., *Reforming Finance: A Historical Perspective.* Cambridge: Cambridge University Press.

Borish, Michael S., Millard F. Long, and Michel Noel. *Enterprise and Bank Restructuring: Recent Losses from Transition Countries.* World Bank Discussion Paper 279. Washington, D.C.

Brock, Philip, ed. 1992. *If Chile Were Texas: A Primer on Banking Reform.* San Francisco, Calif.: Institute for Contemporary Studies Press.

———. 1995. "High Real Interest Rates and Bank Recapitalizations." University of Washington, Department of Economics, Seattle.

Calomiris, Charles W. 1989. "Deposit Insurance: Lessons from the Record." Federal Reserve Bank of Chicago *Economic Perspectives* (May/June): 10–30.

———. 1992. "Do 'Vulnerable' Economies Need Deposit Insurance? Lessons from U.S. Agriculture in the 1920s." In Philip L. Brock, ed., *If Chile Were Texas: A Primer on Banking Reform.* San Francisco, Calif.: Institute for Contemporary Studies Press.

Calomiris, Charles W., and Gary Gorton. 1991. "The Origins of Banking Panics: Models, Facts, and Policy Regulation." In R. Glenn Hubbard, ed., *Financial Markets and Financial Crises.* Chicago: University of Chicago Press for the National Bureau of Economic Research.

Caprio, Gerard Jr. 1992. "Policy Uncertainty, Information Asymmetries, and Financial Intermediation." Policy Research Working Paper 853. World Bank, Washington, D.C.

———. 1996. "Bank Regulation: The Case of the Missing Model." Policy Research Working Paper 1574. World Bank, Washington, D.C.

Caprio, Gerard Jr., and Daniela Klingebiel. 1996. "Bank Insolvencies: Cross-Country Experience." Policy Research Working Paper 1620. World Bank, Washington, D.C.

Caprio, Gerard Jr., Izak Atiyas, and James Hanson, eds. 1994. *Financial Reform: Theory and Experience.* Cambridge: Cambridge University Press.

Davis, E. P. 1992. *Debt, Financial Fragility, and Systemic Risk.* New York: Oxford University Press.

de Juan, Aristobulo. 1987. "From Good Bankers to Bad Bankers: Ineffective Supervision and Management Deterioration as Major Elements in Banking Crises." World Bank, Financial Policy and Systems Division, Washington, D.C.

Fetter, Frank W. 1931. *Monetary Inflation in Chile.* Princeton, N.J.: Princeton University Press.

Fleming, Alex, and Samuel Talley. 1996. "The Latvian Banking Crisis: Lessons Learned." Policy Research Working Paper 1590. World Bank, Washington, D.C.

Garcia, Gillian. 1995. "Deposit Insurance: Obtaining the Benefits and Avoiding the Pitfalls." International Monetary Fund, Monetary Affairs and Exchange Department, Washington, D.C.

Gavin, Michael, and Ricardo Hausmann. 1996. "The Roots of Banking Crises: The Macroeconomic Context." In Ricardo Hausmann and Liliana Rojas-Suárez, eds., *Banking Crises in Latin America.* Balitmore, Md.: Johns Hopkins Press.

Gorton, Gary. 1988. "Banking Panics and Business Cycles." *Oxford Economic Papers* 40(4): 221–55.

Guttentag, Jack. 1994. "Debt, Financial Fragility, and Systemic Risk: A Review." *Journal of Economic Literature* 32 (September): 1238–40.

Haubrich, Joseph G. 1990. "Nonmonetary Effects of Financial Crises: Lessons from the Great Depression in Canada." *Journal of Monetary Economics* 25: 223–52.

Keeley, Michael C. 1990. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review* 80(5): 1183–1200.

Kindleberger, Charles Poor. 1978. *Manias, Panics, and Crashes: A History of Financial Crises.* New York: Basic Books.

King, Robert, and Ross E. Levine. 1993. "Finance, Entrepreneurship, and Growth: Theory and Evidence." *Journal of Monetary Economics* 32: 513–42.

Kryzanowski, Lawrence, and Gordon S. Roberts. 1993. "Canadian Banking Solvency, 1922–1940." *Journal of Money, Credit, and Banking* 25 (August): 361–76.

Marshall, Jorge, and Klaus Schmidt-Hebbel. 1994. "Chile: Fiscal Adjustment and Successful Performance." In William Easterly, C. A. Rodriguez, and Klaus Schmidt-Hebbel, eds. *Public Sector Deficits and Macroeconomic Performance.* New York: Oxford University Press.

Minsky, Hyman P. 1982. *Can 'It' Happen Again: Essays on Instability and Finance.* Armonk, N.Y.: M.E. Sharpe.

Mishkin, Frederic S. 1991. "Asymmetric Information and Financial Crises: A Historical Perspective." In R. G. Hubbard, *Financial Markets and Financial Crises.* Chicago, Ill.: University of Chicago Press.

Morris, Felipe, Mark Dorfman, Jose Pedro Ortiz, and Maria Claudio Franco. 1990. *Latin America's Banking System in the 1980s.* World Bank Discussion Paper 81. Washington, D.C.

Rodriguez, Carlos Alfredo. 1994. "Argentina: Fiscal Disequilibria Leading to Hyperinflation." In William Easterly, C. A. Rodriguez, and Klaus Schmidt-Hebbel, eds., *Public Sector Deficits and Macroeconomic Performance.* New York: Oxford University Press.

Rojas-Suárez, Liliana, and Steven Weisbrod. 1995. "Banking Crises in Latin America: Experiences and Issues." In Ricardo Hausmann and Liliana Rojas-Suárez, eds., *Banking Crises in Latin America.* Baltimore, Md.: The Johns Hopkins University Press.

Schiantarelli, Fabio, Izak Atiyas, Gerard Caprio, Jr., John Harris, and Andrew Weiss. 1994. "Credit Where it is Due? A Review of the Macro and Microeconomic Evidence on the Real Effects of Financial Reform." In Gerard Caprio Jr., Izak Atiyas, and James Hanson, eds., *Financial Reform: Theory and Experience.* New York: Cambridge University Press.

Sheng, Andrew, ed. 1996. *Bank Restructuring: Lessons from the 1980s.* Washington, D.C.: World Bank.

Simons, Katerina, and Stephen Cross. 1991. "Do Capital Markets Predict Problems in Large Commercial Banks?" *New England Economic Review* (May/June): 51–6.

Sundararajan, Vasudevan, and Tomas Jose T. Balino, eds. 1991. *Banking Crises: Structural Weaknesses, Support Operations, and Economic Consequences.* Washington, D.C.: International Monetary Fund.

Weisbrod, Steven R., Howard Lee, and Liliana Rojas-Suárez. 1993. "Bank Risk and the Declining Franchise Value of the Banking Systems in the United States and Japan." IMF Working Paper 92/45. International Monetary Fund, Washington, D.C.

White, Eugene. Forthcoming. "Deposit Insurance." In Gerard Caprio Jr. and Dimitri Vittas, eds., *Reforming Finance: A Historical Perspective.* Washington, D.C.: World Bank.

World Bank. 1989. *World Development Report 1989: Financial Systems and Development.* New York: Oxford University Press.

———. 1995. *Bureaucrats in Business: The Economics and Politics of Government Ownership.* A Policy Research Report. New York: Oxford University Press.

Comment on "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?" by Gerard Caprio Jr. and Daniela Klingebiel

Donald J. Mathieson

Gerard Caprio and Daniela Klingebiel have written a trail-blazing piece on the global experience with financial crises since the 1970s. The database that they have developed on the timing, extent, and cost of various crises will be of great value to economists doing research on financial sector instability. The authors raise three major issues that I will address: the relative importance of macroeconomic and microeconomic sources of banking system instability, the difficulties involved in using market-based prudential supervision in developing countries, and the management of financial crises.

Causes of Crisis

Caprio and Klingebiel's data indicate that recent banking crises have generated some truly massive losses—often as much as 10–20 percent of GDP. No financial reform, no matter how effective or far-reaching, can generate sufficient efficiency gains to compensate losses on this scale. The authors attribute much of this instability to microeconomic factors, that is, to the lack of an incentive structure that encourages prudent risk taking. Although conclusive answers on the relative importance of macroeconomic and microeconomic sources of banking crises will have to wait for econometric analysis of the data the authors have collected, I believe macroeconomic factors are somewhat more important than the authors suggest.

Many of the banking system crises identified by the authors were preceded by either major episodes of macroeconomic instability or major stabilization programs. During periods of macroeconomic instability—especially when accompanied by high inflation—relative price variability can make it extremely difficult to differentiate between good and bad investments. It may be equally difficult to evaluate credit risks when a major stabilization program and extensive structural reforms create the expectation of a much higher rate of return on all domestic investments.

Donald J. Mathieson is chief of the Emerging Markets Studies Division in the Research Department at the International Monetary Fund.

Annual World Bank Conference on Development Economics 1996

Market-Based Supervision

For the most part I agree with the authors' views on the importance of market-based supervision. But it seems to me that implementing such a system is likely to be more difficult in developing than in industrial countries because of key structural differences between the two types of economies. Relative to the size of their GDPs, developing countries face larger internal and external economic shocks than do industrial countries. Thus, other things being equal, banking systems in developing countries are more likely to experience episodes of instability, even if they have the same market-based prudential supervision standards as industrial countries.

As noted by the authors, the vulnerability of banking systems in developing countries to external shocks has been increased by their often high concentration of credit risks. A number of factors contribute to this situation. First, in contrast to the situation in industrial countries, where banks have been losing their largest borrowers to securities markets, banks in developing countries often concentrate their lending on the largest corporate and state-owned borrowers. This approach makes banks vulnerable to the same shocks as the country's largest nonfinancial companies. Since the largest corporations are often in the export sector, large terms of trade shocks can have an extremely adverse impact on developing countries' banking systems.

Ownership links also contribute to the concentration of credit risks in bank loan portfolios. Some developing countries allow large holding groups, which means that much of the activity of the banks controlled by these groups is focused on lending to other nonfinancial companies within the group. Intergroup lending not only concentrates the credit risks in the bank's portfolio, it also creates special difficulties for prudential supervision because of the possibility of self-dealing financial transactions. Self-dealing transactions also can make it extremely difficult to gauge the bank's true capital position.

The authors suggest that this concentration of risks can be partly offset by allowing banks to hold foreign assets. One question is whether banks will have an incentive to hold foreign assets during periods when there is the greatest need for them to do so (for risk diversification purposes). For example, in recent years it has become common for countries to introduce a strong stabilization program at the same time as extensive structural reforms. As a result, domestic and foreign investors anticipate much higher rates of return on domestic claims, and high real interest rates stimulate a large capital flow. Banks participate in this process both by accepting foreign deposits and by borrowing abroad, and they perceive a high real return on domestic lending. Given the often high differential between what banks perceive that they can earn on domestic assets relative to foreign assets, they would not want to hold foreign assets even if they could.

Another structural characteristic that will make it more difficult to introduce market-based prudential supervision in developing countries is the difficult and costly (relative to the amount that can be recovered) enforcement of financial contracts in these countries. Judicial systems often work very slowly, and in a high-inflation environment the real value that can be recovered on a contract whose price is

specified in nominal terms is often so low that it is not worthwhile to pursue legal actions. Obtaining convictions for fraud is also extremely difficult.

In addition, disclosure requirements and accounting standards are usually much less developed than in industrial countries, making it difficult for outsiders to evaluate the position of banks. Indeed, many failed banks have received positive audit reports in the six months prior to their failure.

Taken together, these characteristics will make it more difficult to rely on market-based prudential supervision in developing countries because they increase the vulnerability of the banking system to large shocks, dull the incentives for prudent risk taking, and discourage monitoring of banks' activities.

Crisis Management

Finally, the authors argue that prompt closure rules are the key to containing the costs of financial crises. For most developing countries, however, the rule may have to be prompt control instead. Although prompt closure is a good rule for banks whose capital is eroding gradually (for example, because of fraud), it is unrealistic if a banking system has just experienced a large macroeconomic shock and virtually all banks are in trouble. If this rule were followed literally in such a situation, there would be no payments system. To contain costs, the emphasis must be on administrative controls that limit the risk-taking activities of the managers of insolvent institutions until the banks can be restructured or merged.

Comment on "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?" by Gerard Caprio Jr. and Daniela Klingebiel

Roberto Zahler

Gerard Caprio and Daniela Klingebiel's comprehensive study presents interesting research proposals and suggestions for World Bank policies on the financial sector. I agree with their basic conclusion that appropriate actions must be taken to improve the microeconomic policies that affect banking. But I think that the article minimizes the effects of macroeconomic policies—despite the fact that the authors discuss two related macroeconomic elements: excessive credit growth (more than twice the rate of growth of output) and volatility in output, inflation, and terms of trade. Here I consider some of the implications of overlooking macroeconomic causes of banking crises.

Prices

Caprio and Klingebiel's article fails to discuss the role of macroeconomic policies relative to key macroeconomic prices and the associated effects on banking solvency. In most Latin American countries during the late 1970s and early 1980s one of the most important difficulties the economic authorities faced was that key prices—especially exchange rates, interest rates, and asset prices, particularly nontradable assets—were clear outliers, in the sense of being far away from any reasonable fundamental, norm, or trend value that could be considered a long-term equilibrium value. And experience shows that a country with fiscal equilibrium, reasonable domestic credit growth, high interest rates, and, especially, expectations of exchange rate changes can see these elements come together in a way that stimulates overindebtedness both domestically and abroad. Furthermore, under those conditions a country may end up with huge current account deficits that originate in excess private rather than public spending, combined with domestic interest rates that are higher than foreign interest rates, even when country risk premium differentials and exchange rate expectations are taken into account. Those interest rates might be much higher than any reasonable rate of return in the nonfinancial sectors

Roberto Zahler is president of the Central Bank of Chile.

Annual World Bank Conference on Development Economics 1996

of the economy. Because those sectors are the main debtors of the financial system, such a distortion eventually may develop into crisis.

Another important issue not discussed in the article is that a sudden financial opening usually requires asset prices, especially for nontradable assets, to overshoot relative to any reasonable long-term equilibrium value, especially if denominated in foreign currency. This outcome tends to have two macroeconomic effects with clear implications for the financial sector: a wealth effect, which encourages domestic spending at rates well above what the authors call rapid credit growth, and an artificial increase in the value of banks' loan collateral, which quickly disappears once the asset price bubble crashes, leaving banks with much bigger loan losses.

Policies

In designing and implementing macroeconomic policies, one has to consider their impact on the health of the financial system. From Chile's experience in the late 1970s, I believe three lessons are extremely important. First, macroeconomic volatility must be reduced through coherent and consistent macroeconomic policies, because macroeconomic instability increases the vulnerability of bank debtors, leads to inappropriate and erroneous asset risk evaluation by banks, and ultimately makes the banking system more fragile. Although this important conclusion may appear to be rather obvious in theory, anyone who has had to design macroeconomic policy and implement it in central banks and ministries of finance knows that it is very hard to do in reality.

Second, policies should favor flexibility in order to distribute macroeconomic shocks among various markets and avoid concentrating the effects of a shock in a particular market in a way that will prove unbearable, unbalanced, and unsustainable. These shocks—as reflected by abnormal exchange rates, abnormal interest rates, or abnormal wages or unemployment rates—tend, directly or indirectly, to cause the financial system to deteriorate.

Finally, the speed and sequencing of reforms matter a lot. If reforms are poorly designed in terms of pacing or sequencing, they may generate huge distortions during the transition from a repressed to a liberalized situation. Thus the right speed at which to liberalize the domestic financial sector relative to the foreign financial sector or the right sequence in which to open trade and to liberalize the financial sector are not the same in every country or in a single country under different initial conditions. If reforms are not properly sequenced and paced, the situation may end up worse than it was initially—even to the point of total bankruptcy of the financial system.

Supervision and Regulation

In Chile during the 1970s banks tended to show a reasonable balance between assets and liabilities denominated in foreign currency. But even though, in a simple way, banks were complying with the rules, not enough care was taken to monitor the

exposure of banks' debtors to a change in the exchange rate. This lack of oversight proved extremely costly. The value of most loans eventually was downgraded because, even though assets were matched to liabilities, debtors were unable to absorb fluctuations in exchange rates. Thus regulators must be extremely cautious, exercising judgment in evaluating risk and not limiting themselves to one kind of accounting analysis.

A second point is that, by requiring bank directors to disclose information and to assume responsibility for their credit decisions, the banking law should play a major role in providing incentives for prudent risk taking by banks. In many countries the banking law does not clearly specify how, when, and under what circumstances regulators can intervene in banks.

Finally, the authors' assertion that it would be difficult for banks in very small economies to achieve the required incentive system without diversification abroad touches on an extremely delicate matter that requires careful distinctions and wise judgment. Although it is true that the risks of banking are greater in small economies than in economies where there are more opportunities to diversify, many small economies lack the capacity required to oversee bank lending or investments abroad. By abroad I mean other emerging markets, not Europe or the United States. And those markets may have a higher country risk, higher contagion risk, higher arbitrage risk, and different sorts of supervision. Moreover, big international banks could end up using a relatively low-risk country like Chile to intermediate funds and to pass these funds to higher-risk countries, increasing the risk to Chile abroad through its banking system's activities. Thus banking activities outside the home country may end up increasing rather than reducing banks' overall portfolio risk. This is an area that requires not only more research, but also a lot of experience and know-how, with careful studies of each country's specific situation.

Floor Discussion of "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?" by Gerard Caprio Jr. and Daniela Klingebiel

Brian Wilson (chair) opened the discussion by asking for Gerard Caprio's response to the discussants' comments. Caprio said that he agreed with many of them. On Roberto Zahler's (discussant) final point—whether diversifying abroad increases or lessens risk for banks in developing countries—they probably disagreed less than Zahler thought. Caprio had not necessarily meant that banks should engage in banking activity abroad but rather that they could hold a relatively well-diversified portfolio of securities. If Mexican banks had had half their assets invested in a global bond fund in 1994, for example, they would be in better shape today because not all of their assets would be concentrated in Mexico. Caprio said he was well aware that banks can get into trouble when they move away from markets they know, so he would not urge countries to accept that recommendation before considering the others, especially the one about strengthening incentives in the banking sector.

Caprio agreed with Donald Mathieson (discussant) that the environment in developing countries is more difficult in many respects, but he pointed out that industrial countries were once developing countries and that when they were, they had had far higher capital ratios, often not because they were required to but because the framework induced it. There was no government safety net, and bank owners had double, triple, or unlimited liability for their losses. In U.S. states with unit banks (single branch banks), banks failed regularly. But in states where banks were allowed to establish branches and were induced to hold capital ratios of 30–40 percent because of their liability, banks weathered the same macroeconomic shocks better than did the undiversified banks in other states.

As for the points made about macroeconomic shocks, said Caprio, whenever there are efforts to push consensus too far in one direction, there is a risk that other valid points will be underemphasized. Although he would not deny that macroeconomic factors are important—and having worked at a central bank he appreciated the

This session was chaired by Brian V. Wilson, vice president, Financial Policy and Institutional Strategy, at the World Bank.

Annual World Bank Conference on Development Economics 1996

importance of price stability and the need for officials to pay attention to it—he also thought that people had already learned that lesson. Thus, while it is important not to forget it, perhaps it is time to focus on another: the influence of incentives in the banking system. Daniela Klingebiel added that only part of the article's focus was on understanding microeconomic and macroeconomic factors. Another major concern was with how countries should deal with crises once they occur. And after looking at twenty-nine cases in more detail, she and Caprio found that governments that dealt with incentive distortions at the same time that they minimized or resolved macroeconomic imbalances had the highest probability of resolving the crisis.

A participant from the World Bank expressed strong reservations about the need for small, emerging countries to diversify abroad. The difficulty of a host country banking supervisor and a home country banking supervisor exchanging effective supervisory information has been well illustrated by the problems created by the actions of Bank of Commerce and Credit International in 1991 and Japan's Daiwa Bank in 1995. Moreover, the example of state banks diversifying within the United States does not hold because all the banks were under the supervision of the federal government. Caprio agreed that banks could get into trouble when they engaged in lending activities or dealt with businesses and localities with which they were not familiar. But that was not the only route to diversification, and some routes were easier to supervise than others.

For example, the World Bank could offer countries mutual funds in which their banks could invest, although some observers would argue that large investment firms could do that better than the World Bank. The point, Caprio said, is that banks concentrated solely in a small economy are going to get into trouble unless they impose a 50, 80, or 100 percent capital requirement. And that might be harder to sell than diversification abroad. Mathieson said that he did not see how banking systems in small economies could diversify, however, because most of them are too small to do so. For many developing countries the only way to achieve a diversified banking system would be on a regional basis.

A participant from the World Bank, returning to Caprio's point about strengthening banks at the microeconomic level to provide protection against macroeconomic shocks, asked the panelists how easy and practical it is to use regulations to provide that kind of protection. What exactly should regulators do if outliers become apparent while asset values are rising? Is it realistic to assume that regulators can tell these banks that such developments are out of line while still transferring the risk allocation to the banks? Zahler answered by saying that, in practical terms, it is difficult if not impossible for regulators to say that to banks. The banks' natural response is to ask why, if the authorities are not doing anything about the problem, banks should, say, downgrade the value of collateral or consider a change in exchange rates that is not part of formal macroeconomic policy. Zahler agreed that while regulation is crucial, it is practically impossible and even undesirable (because of costs) to make a financial system immune to shocks. He would insist, however, on reasonable macroeconomic policies in the sense that they should not create disequilibrium or distress in the financial system.

Countries that have reformed their regulatory framework, added Caprio, most often did so after substantial losses in their financial system. That had certainly been the case in Chile, New Zealand, and the United States. The real challenge lies in keeping the system relatively robust or even strengthening it during good times. In the United States, for example, some analysts are urging that capital requirements be strengthened. Although the Basle agreements stipulate an 8 percent capital adequacy ratio, 10 percent is the minimum for full banking power in the United States, and there have even been recommendations to go above that. Imposing stiffer requirements is difficult during good times, however. Moreover, added Mathieson, in developing countries that have a history of extensive bailouts when large shocks occur, it is hard to make credible any government plan not to provide a bailout when the next shock comes.

Zahler offered a concrete example. Chile is entering its thirteenth year of high economic growth—averaging more than 6 percent a year—so in theory things have been going so well that the banking system could be facing the risk of regulators relaxing their standards. But in fact Chile has been prudent, in the sense that it is carefully examining the current account deficit, the growth rate of nontraditional exports, unemployment, real wages, productivity, and so on—looking at both the financial system and macroeconomic policy. So, in fact, the authorities are trying to put the brakes on the economy when most people are encouraging them to loosen things up. We are lowering inflation, but we are proceeding gradually to avoid sudden changes in asset values. It is important that we not let a good macroeconomic environment cause problems in the financial system. When we examine indicators—in particular, interest rates, exchange rates, and asset prices—we need to be alert not only for trends that may signal macroeconomic problems, but also for those that may have adverse implications for the financial system.

In terms of macroeconomic and microeconomic causes, a participant said, it is impossible to distinguish between them because the tension on the weak link in the chain is what causes the problem. It is important to assess whether it is easier to prevent macroeconomic imbalances or to address microeconomic problems, but it is essential to work on both.

Sequencing is also important, said another participant. In Chile there was a period of macroeconomic adjustment and financial liberalization prior to the problems in 1981–83. Only during the mid- and late 1980s did Chile reform the prudential and supervisory framework. As in many other countries, the problems in Chile followed a pattern: extensive financial liberalization and financial deepening followed by crashes followed by prudential reform. Maybe an issue that needs to be addressed is what the institutional framework should be before the financial system is liberalized. Caprio admitted that it was hard to sell the message of good sequencing: that countries should build up their regulatory framework and supervisory capacity early on. When a new administration has only a brief window in which to make reforms, it is better that the reforms be done quickly than not at all.

The speed of sequencing is a broad subject, said Zahler. Based on Chile's experience, he thought an economy should open up commercially before it opens up

financially to the international economy. Moreover, financial opening should be gradual, to give the authorities more control over macroeconomic variables. As for banks diversifying by going abroad, of course they could buy bonds and AAA-rated instruments, but that is not the kind of business most bankers (at least in Chile) are looking for. They want real business, and in real business no one has a very good idea of the risks. Zahler favored allowing banks to invest or lend a small amount of capital abroad, but at a gradual pace set by regulators. That way banks and regulators would learn how to do business abroad but would not put too much at risk and would develop reasonable expectations of returns.

It is important, said a participant from the International Monetary Fund, to distinguish between financial crisis (such as a sudden deterioration in banks' portfolios) and financial distress (which represents structural problems in an economy). The effects of microeconomic and macroeconomic variables differ for cases of crisis and distress. The participant wondered whether grouping countries according to the severity of their banking crises (major or minor) was important in determining whether their responses to the crises were successful or unsuccessful. For example, except for Chile and perhaps the United States, most of the successful countries in Caprio and Klingebiel's sample did not experience major crises. That said something about the importance of the crisis and the implications of the mechanisms or restructuring exercises used to deal with the problem. Thus it might be more useful to understand the backgrounds of different countries than to lump them together in terms of major or minor crisis.

Caprio agreed, suggesting by analogy an analysis of the causes of death. In a group of people who died of either heart attacks or cancer, it would be possible to say that if they had died from a heart attack, it was probably rather sudden, but with cancer it was not as easy to say because it is harder to know when their illness became terminal. Similarly, in analyzing the causes of financial shocks, even if there were a sudden event, there might have been substantial distress for a long time before it happened. Take the case of the U.S. savings and loan institutions. When did they die? Most had died long before their death was acknowledged. Caprio believed that they started to die in the 1960s, when they were, in effect, constructed to go out of business. With the interest rate mismatch they began assuming, it was only a matter of time before problems developed. Caprio agreed that the situation would be easier to analyze if there were more serious econometric evaluations; the problem is that the data are slippery. It is difficult to establish with confidence when a crisis occurred. To do more serious econometric work, it will be important to determine some of those dates.

Reducing Poverty: Targeted Programs and Rural Finance

Political Economy of Alleviating Poverty: Theory and Institutions

Timothy Besley

Poverty alleviation programs usually develop from one of two approaches: the technocratic or the institutional. Economists tend to rely on the technocratic approach, which emphasizes designing policies that target the poor as effectively as possible. Noneconomists and nongovernmental organizations favor the institutional approach, which emphasizes the development of institutions and the improvement of government performance as means of alleviating poverty. This article bridges the gap between these two perspectives using a simple theoretical framework that shows how each approach can be modeled and incorporated into policy discussions. From this, a two-pronged assault on poverty is possible—one that effectively targets interventions while developing ways of delivering resources that defy or change the governing political mechanism.

> The political economy of targeting has to be concerned not just with the economic problems of selection, information, and incentives but also with political support for, and feasibility of, aiming public policy specifically at removing deprivation of particular groups.
>
> —Amartya Sen (1995, p. 21)

Understanding how to alleviate poverty is a central concern of development economics. Although there is ample evidence that policies designed to foster economic growth significantly reduce poverty (Bruno, Squire, and Ravallion 1995), policies aimed specifically at alleviating poverty are also important. For example, programs that provide credit and build human capital try to eliminate the causes of poverty.[1] Such programs can have a short- or a long-run perspective. This article's short-run perspective on program design assumes that income-earning abilities are fixed. The optimal form of intervention in this context is much debated in industrial and developing countries alike (Atkinson 1987; Lipton and Ravallion 1995; Dasgupta 1994).

Timothy Besley is professor of economics at the London School of Economics. Many of the ideas in this article were developed in joint work with Steve Coate. The author is also grateful to Arsenio M. Basilacan, Robin Burgess, Jeff Hammer, Karla Hoff, and Callisto E. Madavo for comments on an earlier draft.

Annual World Bank Conference on Development Economics 1996

Two somewhat disparate approaches to program design, what I call the technocratic and the institutional, are explored here. The technocratic approach, usually associated with economists, focuses on targeting, exploring the theoretical and empirical implications of trying to direct limited resources to people with the greatest need. These efforts emphasize the difficulties of identifying target groups and use creative approaches to program design that substitute for the detailed information required to achieve first-best results. Central to such explorations are the incentive effects of program design, which underscore the need to know key behavioral parameters, such as labor supply responses, in order to formulate policy. I call this approach technocratic because of its implicit model of the policymaking process. Its perspective is that of a "planner" charged with meeting a well-defined objective in the best way possible. As such, it parallels a long and distinguished tradition in public sector economics.

The institutional approach is more common among noneconomists. For them the question of why programs for the poor do or do not work has much more to do with social institutions than with policy design. In this view antipoverty policies fail because the poor lack political power or because administrative incompetence or corruption keeps governments from delivering services. Thus improving the lives of the poor requires developing institutions, improving government performance, and changing political structures and attitudes toward the poor.

The gulf between these two perspectives is evident in their views of the role of nongovernmental organizations (NGOs) in poverty alleviation programs. The technocratic approach rarely refers to NGOs, while the institutional approach considers them vital to the attack on poverty in developing countries. Institutionalists see a number of roles for NGOs. They see NGOs as policy actors, able to substitute for corrupt and inefficient governments in delivering services to the poor, and as an essential part of the political process, fostering empowerment among the poor and lobbying governments to change policies.

The institutional approach presents interesting challenges to economists. In one sense it suggests that economic analysis is of secondary importance in the design of poverty alleviation programs. Determining whether this claim is true would require analyzing how effective institutional approaches—such as handing over responsibility to NGOs—are in alleviating poverty. The literature on this subject is extremely thin. Studies of NGO performance tend to be case studies carried out with small and frequently unrepresentative samples (Edwards and Hulme 1995).

The Cost-Minimizing Approach

The increasing concern with better targeting in poverty alleviation programs stems from governments' desire to minimize the cost of achieving poverty alleviation objectives. This desire is an implication of models in which taxpayers, as financiers of transfer programs, seek fiscally efficient methods of helping the poor—that is, they want programs to be designed in a way that minimizes the financial burden imposed. Thus the insights from the technocratic literature are legitimate concerns in a well-defined decisionmaking model of antipoverty policy.

A commonly accepted model of program design, the cost-minimizing approach, addresses a number of salient features in current debates about transfers to the poor. Moreover, it is consistent with the desire for targeting. It is also a useful first step toward developing a positive theory of transfers to the poor. The model described here (based on Besley and Coate 1992, 1995) makes no pretense at realism. It is, however, a useful vehicle for clarifying thinking about a number of issues relating to poverty alleviation programs.

Basic Framework

The model views society as composed of two groups: those who make transfers (the rich) and those who receive them (the poor; see Zeckhauser 1971). There are n_R rich and n_P poor. Each group is assumed to be homogeneous, though this assumption is later relaxed. There are, however, differences between the income-earning abilities of members of the two groups that are captured in their wage rates.[2] Hence an individual of type i who supplies l units of labor will enjoy earnings of $y_i = a_i l$ ($i \in \{P, R\}$). Consumption is denoted by x_i. Preferences of the poor are given by:

$$u_P = x_P - h(l_P) \tag{1}$$

where $h(\cdot)$ is the disutility of labor.[3] The utility of the rich is

$$u_R = x_R - \xi(x_P, \lambda_R) \tag{2}$$

where λ_R is a parameter that represents the degree of sympathy the rich feel toward the poor.[4] (The role of this parameter will become clearer below.)

Equation 2 reflects the belief that the rich care about their own consumption but also get disutility from poverty.[5] The equation builds in, in a simple and transparent way, the notion that the rich care about the poor. This concern could be interpreted as altruism or as an attempt to avoid the social unrest that could arise if the poor suffered too great a consumption deficit (Boone 1996). In any event, whatever the reason, there is a motivation to redistribute income. There is, however, an implied nonwelfarism in the preferences of the rich; they are presumed to care only about the consumption of the poor, not their utility.

The government is assumed to be controlled by the rich. Its objective is to design a poverty alleviation program that is financed through taxes paid by the rich. The taxes finance transfers to the poor. Transfers to the poor can depend, in principle, on any characteristics observable to the government. Initially, however, I will focus on the case where transfers depend only on income—a function $B(y)$.

The government's choice of program is constrained in a number of ways. All the cases considered here impose a constraint of voluntary participation by the poor, which says that the poor must be willing to take any benefits intended for them.

This condition rules out a scheme in which the rich are able to tax the poor, and it captures the idea that taxes are compulsory in a way that transfers are not. It also reflects a concern that pervades transfer programs, about the failure of eligible people to take up benefits. Formally, the requirement of voluntary participation says that people not taking transfers must be at least as well-off as they would be if they were taking transfers:

$$\text{(3)} \qquad \underset{l}{Max}\ \{a_P l + B(a_P l) - h(l)\} \geq \underset{l}{Max}\ \{a_P l - h(l)\} \equiv u_P.$$

Before specifying the problem faced by the government, we must first specify the tax system used to finance this program. For simplicity, we suppose a head tax on the rich. Although such a tax is unrealistic, the general principles discussed here are not affected by this simplification.

In the model so far, the form of the optimal benefit schedule is simple. With no imperfect information about the poor's work ability, the rich can effectively choose the consumption-labor supply bundle of the poor, subject to voluntary participation. This is equivalent to setting a benefit schedule of the form:

$$\text{(4)} \qquad B(y) = \begin{cases} x_P - y_P & \text{if } y = y_P \\ 0 & \text{otherwise.} \end{cases}$$

Individuals receive benefits only if they have earnings of less than a certain level. The absence of imperfect information makes this distinction enforceable. Thus the problem to be solved by the optimal benefit schedule can be written as:

$$\underset{\{x_P,\, y_P\}}{Max}\ y_R - \left(\frac{n_P}{n_R}\right)(x_P - y_P) - \xi(x_P, \lambda_R)$$

subject to

$$\text{(5)} \qquad x_P - h(y_P/a_P) \geq u_P.$$

This is the canonical form of the optimal poverty alleviation program.

Although the point here is not to develop the poverty alleviation program that solves this problem—especially given how unrealistic the model is—it is useful to see the solution since it is the benchmark for illustrating the main ideas in this article. The only constraint the government faces is voluntary participation. Hence the transfer scheme stipulates an earnings level y^*_P just high enough that the poor are better off than they would be without the transfer.[6] The government then sets the transfer to fill the gap between this earnings level and the poverty line. Transfer recipients in such a scheme are "forced" to work harder than they would in the absence of transfers. But since they receive a transfer, they receive their pre-program utility level. (This type of transfer scheme is known in the income maintenance literature as a conditional earnings subsidy.) Their consumption is z, the minimum level that the rich want the poor to achieve.

Before complicating the model, we make an observation that is key to the analysis that follows. The problem of designing an optimal benefits scheme can be thought of as consisting of two stages. The first is the program design stage:

$$\underset{x_p}{Min}\ (n_P(x_P - y_P))$$

subject to

$$x_P \geq z$$

and

$$x_P - h(y_P/a_P) \geq u_P. \tag{6}$$

The term income maintenance comes from the fact that the income level to be attained by the poor (denoted by z) is specified as a constraint on the problem. The solution to this is denoted by $C(z)$, defined as the minimum cost of fulfilling a poverty alleviation objective of z. Solving this stage does not involve the preferences of the rich.

The second stage reflects the choice of a subsistence level. It involves solving:

$$\underset{z}{Max}\left\{ y_R - \frac{C(z)}{n_R} - \xi(z, \lambda_R) \right\}. \tag{7}$$

Given the cost of delivering a maintenance level of z, the rich decide how high they can afford to set z.

These two stages can be thought of as akin to designing a benefits scheme and then deciding how to finance it and how generous it should be. Viewing the design of poverty alleviation programs as comprising these two stages is a major attraction of the cost-minimizing approach. The rest of this section discusses some implications of this approach, leaving questions of benefit determination to the next section.

Heterogeneous Poor and Imperfect Information

Most discussions of the design of poverty alleviation programs assume heterogeneity among the poor, whether in their need for transfers or in their deservingness. These ideas and their implications are incorporated into the cost-minimizing approach.

Suppose that the poor fall into two groups: those with a wage rate sufficient to avoid poverty under normal labor supply conditions (meaning that, at their utility-maximizing labor supply, they will not be poor) and those with a wage rate so low that they will be poor even under such conditions. The first group is referred to as undeserving and the second as deserving.[7] With full information only the deserving will receive transfers under the cost-minimizing approach. Thus the optimal cost-minimizing approach is targeted toward the deserving group.

Now suppose that information is imperfect and that the government cannot tell who falls into each group, even though it can observe private sector earnings. If the

government wants to alleviate poverty among the deserving, it must make a common transfer to all poor individuals. The level of this transfer $(z - y^*_P)$ is the same as described above. The key difference is that the undeserving may choose to reduce their earnings in order to claim the transfer, thus raising the cost of the program. The government has several options for trying to redirect the program only to the deserving poor.

INCOME TESTS. If the government can observe earnings, then testing on this basis is an option (Besley and Coate 1995). The possibility that the undeserving may reduce their earnings in order to qualify for support is a standard concern in the literature. One solution is to offer a benefits system that provides some incentives for the undeserving by specifying a gradual withdrawal rate for benefits. Thus poor individuals receiving benefits may keep some fraction of any extra earnings they may generate without having their benefits reduced accordingly.

The notion that governments must design programs that respect incentive compatibility concerns is familiar to students of modern public economics. Loosely, it states that governments must encourage the undeserving to reveal themselves by offering an option that they prefer to the option offered to the deserving. Suppose that the government offers a tiered scheme with a benefit level b_D for the deserving and a benefit level b_U for the undeserving. The undeserving must find it preferable to claim b_U to claiming b_D.[8] Although this outcome is possible even if the government pays no transfer to the undeserving, here we focus on the case where the undeserving would rather work less and claim a benefit. In this case the government will optimally offer a benefit level $b^*_U < b^*_D$ and an earnings level at which it can be claimed so that the undeserving prefer the lower benefit level.[9] The government gets an incentive benefit by offering benefits to a group to whom it really does not want to make a transfer, since the transfer the group receives is smaller than it otherwise would be, and so costs are lower.

Two general lessons can be derived from this simple model. First, optimally targeted programs can include transfers to the nonpoor that are designed to preserve incentives. The nonpoor are encouraged not to take the benefit intended for the poor by being offered a transfer instead. Although some discussions of poverty alleviation programs argue that any level of transfer to the nonpoor is wasteful, the theory of incentives demonstrates that transfers to the nonpoor can be an essential part of a well-designed cost-minimizing program—one that attaches no weight to the well-being of the undeserving group. By making such transfers, the program prevents the undeserving from trying to participate more fully in the program. The idea that this is a flaw in the scheme is incorrect.

Second, cost minimization here generates a Pareto improvement. The surplus accrues to the taxpayers rather than to the poor, however. This outcome reflects the fact that the cost-minimizing approach develops from the perspective of the taxpayers (who finance the programs). Better program design benefits the people who are paying for it. The next section shows that taxpayers' willingness to finance increased benefits may rise with more efficiently targeted programs.

NONINCOME TESTS. Even if the government does not have access to income information, it is still possible to reduce the cost of poverty alleviation programs, using nonincome tests. Work requirements are the most common test of this type. The idea is that the undeserving in the target group are more likely to be deterred by having to work in exchange for benefits (Nichols and Zeckhauser 1982). Among the costs imposed on transfer recipients is the displaced private sector work, either in markets or in home production. This added cost implies that, if concerns about participation are to be respected, meeting program objectives will necessitate a higher transfer level to compensate for lost private sector earnings.

Besley and Coate (1992) consider the design of the cost-minimizing transfer program in this case using the two-type model of the poor described above with an incentive constraint for the case where income is not observable. Suppose that the government designs a scheme that demands that individuals of type i perform c_i hours of work in exchange for transfer of b_i. The work requirement may deter the undeserving from claiming a transfer.[10] Thus the choice for governments is between a universal scheme in which everyone receives a transfer and a scheme that imposes a work requirement so that only the deserving participate.

Either scheme can be cost minimizing, depending on the parameters of the problem. Essentially, the tradeoff is between the value of displaced earnings and the gain from reducing participation by the undeserving. Using this model, it is possible to make a case for a universal benefit scheme in one place and a targeted scheme in another—even if the objective is the same in both places. The case for targeting is contingent on the economic environment being studied.[11] This model also underscores the point that better targeting saves taxpayers money, unless it results in improved program generosity (see below).

This example illustrates a case in which the utility of the undeserving is reduced by a move to a targeted scheme. This happens because the use of a work requirement either lowers the utility of the undeserving if they choose to work to claim a benefit or it deters them from claiming any benefit, so they receive no transfer. Hence targeting is not a Pareto improvement.

Nongovernmental Organizations

So far the model has assumed that policy authority is vested in a government that has sole responsibility for making transfers to the poor. This monolithic approach parallels the standard model of government transfer activity in public economics. The reality, however, is quite different, and here we consider some implications of allowing NGOs as well as the government to make transfers. Even industrial countries with extensive welfare systems have large voluntary sectors that deliver services to the needy (Kramer 1981).

Although NGOs play many roles, two are paramount, one a direct policy function and the other an indirect policy function. As direct policy actors, NGOs may provide health care, credit, education, or food aid. These functions are directly related to the poverty alleviation process. As indirect policy actors, or influencers of

government policy, NGOs can act as pressure groups, mobilizing citizens to influence the policy process (often referred to as empowerment or encouraging participation) or giving technical expertise. The approach an NGO takes is determined by the government's willingness to cede its monopoly on policymaking. This section analyzes how an NGO's role as a direct policy actor influences the design of poverty alleviation programs, leaving discussion of its indirect role to the next section.

Although NGOs in developing countries have not been studied extensively by economists, there is a large literature on private philanthropy—and government responses to it—in industrial countries. The basic idea is that a government considers whether to socialize an activity that has traditionally been undertaken by private charity. The usual argument is that government financing of the activity may not lead to any increase in provision, since for every dollar the government spends, private charities spend one dollar less (Bernheim 1986; Roberts 1987; Kingma 1989). Although this one-for-one view is largely discredited by both theory and evidence, the notion that the benefit of government intervention should not be measured as though every dollar spent yields an additional dollar's worth of service provision is robust up to the point where private charity is driven to zero.

The situation may be different in developing countries, where indigenous private voluntary activity is presumed to be less important than public provision.[12] In this model the NGO function is largely a substitute for, or at least supplement to, government activities. In this case, then, the concern is how NGOs might crowd out government support.

The idea is easily illustrated. Consider a country in which an NGO and the government have the same preferences and technology for delivering support to the poor.[13] The government and the NGO have similar motivations: finding the cost-minimizing way of helping the poor reach a certain level of consumption. Assume that the country contains a well-defined and homogeneous target group of poor and that information is perfect. The shared objective is to raise the poor's consumption to some level z.

The two entities also share a common participation constraint. The government's strategy is a transfer function denoted $B_G(y)$; that of the NGO is denoted $B_N(y)$. These show how much these actors are willing to give to a particular poor person as a function of their earnings. Suppose that these strategies form a Nash equilibrium (the two sides coordinate their strategies since failure to do so would increase their costs). The choice of a benefit schedule in this context is equivalent to having the government and the NGO choose consumption income pairs (x_i, y_i) $i \in \{G, N\}$ such that the utility of a poor person is

$$x_G + x_N - h\left(\frac{y_G + y_N}{a_P}\right). \tag{8}$$

In this case there is family of Nash equilibria that satisfies the conditions:[14]

$$x_G + x_N = z \text{ and } z - h\left(\frac{y_G + y_N}{a_P}\right) = u_P. \tag{9}$$

The outcome for the poor is the same as it would have been had the NGO not been involved. The difference is that the cost of poverty alleviation is shared between the government and the NGO, so the taxpayers—not the poor—benefit from NGO involvement. Such an outcome would certainly explain government's preference for NGO activity—a preference that is not based on any notion of altruism. It also suggests an important lesson for any study trying to look rigorously at the contribution of NGOs to poverty alleviation: finding that an NGO is helping the poor does not guarantee that the poor are really benefiting in an equilibrium sense. A model of government behavior is required to generate the appropriate counterfactual—would the outcome for the poor have been any different in the absence of NGO activity.

Some version of this logic is likely to hold up in a wide range of settings beyond the simple one described here, provided that government behavior is modeled endogenously. Thus some part of NGO transfers to the poor can become a transfer to taxpayers. The special case of a complete crowding out could be relaxed in a world where the activities of NGOs and governments are imperfectly substitutable. If the NGO takes a more generous view of poverty than the government, the modeling can be amended as follows. Suppose that the government's poverty line, z_G, is less than the NGO's poverty line, z_N. In this case there is a Nash equilibrium in which:

$$x_N = z_N \text{ and } z - h\left(\frac{y_N}{a_P}\right) = u \tag{10}$$

with $z_G = y_G = 0$. In this case the NGO is the only entity making transfers to the target group, and NGO activities affect both the rich and the poor. The rich benefit because they would have made a transfer to the poor in the absence of the NGO activity (albeit at a lower level). The poor benefit because their consumption is higher.[15]

Another common argument is that NGOs have a comparative advantage in making transfers to the poor. One version of this argument is that NGOs have better information about the poor because of local contacts. Another is that NGOs reduce the leakage in delivering benefits that results from corruption in government bureaucracies. In both cases it is considered more efficient for the NGO, rather than the government, to provide services directly to the poor. Taken a step further, this argument suggests that it makes more sense for the government to fund NGOs than to undertake provision itself. In appraising the benefits of such arrangements, however, it is important to remember that many of the cost savings may accrue back to the government (in our framework, the rich) rather than go as transfers to the poor. The exception is the case where an NGO can reach groups that the government cannot. Since the government's transfer is already zero, there is no way that an NGO can crowd it out further. It is not sensible to analyze such cases by looking at what NGOs and governments are doing in practice, however, since our model predicts that where NGOs are active it often is optimal for the government to reduce its support to zero.

Some observers have suggested that there may be a role for government in coordinating the actions of NGOs that are willing to provide support to the poor, particularly if the actions of several NGOs are complementary. For example, two NGOs that wanted to provide vital inputs (such as money and local expertise) to a poverty alleviation program would only do so if the other agreed to participate. The government could play a role that is quite independent of whether it actually finances these activities.

Program Generosity

The analysis so far has taken the generosity of poverty alleviation programs as a given. But after determining the least-cost way of achieving a program's objectives, the government must decide how high to set its targets. Although the decisionmaking process is guided by each country's history and institutions, some decisionmaking procedures reflect general incentives. Equating the marginal cost of poverty alleviation with its marginal benefit is a useful starting point for thinking through these issues.

Policy Choices and Costs

Recent years have seen a resurgence in economic models of government behavior and the implications for policy choice (Mueller 1989). For economists involved in such study, two issues are paramount. The first is predictive; that is, can a set of citizen preferences be used to predict policy outcomes, given a particular set of institutions? In a pioneering effort on this front, Downs (1957) attempted to model the implications of two-party representative democracy. Downs suggested that political parties trying to get the most votes would offer the policy platform preferred by the median voter. This analysis has had a significant influence on how economists think about politics. A vast body of work that has gone beyond this early model, however, has shown that the convenient prediction of the median outcome is far from robust.

The second issue is normative. That is, is there any reason to expect the outcomes attained in political equilibrium to possess desirable normative properties? The idea that a tyrannous majority could exploit minorities in a democracy makes it difficult for some observers to believe that democratic political equilibrium is equitable. Work in the public choice tradition, beginning with Buchanan and Tullock (1962), also has suggested problems with assuming that democratic policymaking is efficient.

In general, a system of government can be thought of as determining *control rights* to policy choice. In dictatorships such rights are usually seized by force; in democracies they are assigned through elections. Suppose that a particular citizen in a society has been given the right to choose policy on behalf of others. Also suppose that initially there are no constraints on the policy choices. Whereas the main complications in designing the optimal cost-minimizing poverty alleviation program stem from heterogeneity among the poor, those involved in determining program

generosity stem from heterogeneity among the rich. Thus policymakers must find a way to aggregate diverse opinion.

In the case of the homogeneous rich the choice of policy can be determined straightforwardly. Once a cost-minimizing program is in place, the rich face the cost schedule $C(z)$ associated with a particular poverty line z. These costs are shared by taxpayers according to the prevailing tax system—here a head tax. If the rich share the same valuation of poverty alleviation introduced as λ_R above, then their utility from a level of poverty z is

$$y_R - \frac{C(z)}{n_R} - \xi(z, \lambda_R). \tag{11}$$

Society's poverty alleviating objective is then chosen so as to maximize equation 11.

To help forge a link with the discussion in the preceding section, it is useful to consider the first-order condition for the choice of z:

$$\frac{1}{n_R}\frac{\partial C(z)}{\partial z} = -\frac{\partial \xi(z, \lambda_R)}{\partial z}. \tag{12}$$

Equation 12 simply says that policymakers choose a poverty line that equates $1/n_R$ times the marginal cost of poverty alleviation with their marginal valuation of poverty alleviation. Because poverty alleviation is a public good financed by a head tax, its perceived marginal cost is only $1/n_R$ times the marginal cost (the individual's tax price). The marginal cost of poverty alleviation ($\partial C/\partial z$) serves as the "price" of poverty alleviation.

Thus the marginal cost of poverty alleviation determines the generosity of the program in this setting. Other things being equal, a program with a larger marginal cost of poverty alleviation will be less generous toward the poor. Using the previous section as a starting point, it is possible to calculate the marginal cost of raising the poverty line. For example, in the first model with full information and homogeneous poor,

$$\frac{\partial C(z)}{\partial z} = n_P\left(1 - \frac{a_P}{h'(y^*_P/a_P)}\right). \tag{13}$$

Hence a larger group of poor will tend to imply less generous treatment.

It is interesting to consider how this finding relates to targeting. Are programs that are better targeted also more generous? In general, the answer is no. This can be shown in the second model laid out on page 122 by supposing that the government switches from a program that gives the same benefit to the deserving and the undeserving to one that offers a lower transfer to the undeserving. Such a program will raise the marginal cost of poverty alleviation because the government has to raise transfers across the board when it increases them to the deserving poor.[16] Moreover, it is more attractive for the undeserving to masquerade as deserving when the poverty line is higher.

Thus greater targeting may lower the average cost of alleviating poverty, but it does not necessarily lower the marginal cost. Hence greater targeting may lower the

transfer that the poor receive by raising the marginal cost of poverty alleviation, once the effects on program generosity are taken into account.[17] This is not to suggest that this is a general result; whether it occurs depends on the specifics of the model. It is worth remembering, however, that the targeting literature's focus on average cost has overlooked a fact of political economy—that marginal cost is also important.

Heterogeneous Rich and Political Economy

Next consider some implications of differences in views among the rich about the appropriate poverty line. To model this, suppose that the rich (n_R) are labeled from the least to most generous, $\lambda_1 \leq \ldots \leq \lambda_{n_R}$. Let λ_m denote the median value of the λ's (assumed unique). Associated with any particular rich person j will be a preferred level of program generosity or poverty line: z^*_j.[18] Thus there is now disagreement over the poverty line. All individuals of type k, for whom $z^*_k > z^*_j$, will regard the scheme selected by individuals of type j as insufficiently generous. In a similar vein there may be some rich individuals of type k, for whom $z^*_k < z^*_j$, who view the provision for the poor selected by individuals of type j as too generous. In the many countries that survey attitudes toward the poor, it is clear that views about the generosity of transfers differ along the lines suggested by such a model.

The political process determines which group's views about generosity prevail. Downs's model of political competition predicts that the outcome preferred by the median voter will prevail. In the case above, citizen j is the median in the distribution of λ's, so that equation 12 holds evaluated at λ_m the median value of λ_j.[19] If this is the case, it will clearly give rise to some level of poverty alleviation denoted by z^*_m.

Now consider the perspective of an outsider to the society looking at the way policy is designed. These outsiders, be they international organizations, NGOs, or just concerned citizens, may have views about policy design that differ from those that prevail in the society in question. How can these outsiders change the way that society treats the poor?

If the political mechanism that determines program generosity is taken as a given, as many people believe it should be, the only reasonable way to change the treatment of the poor would be to change the attitudes of voters. One approach is to try to involve traditionally unregistered or disenfranchised groups in the political process, thereby shifting z^*_m in this simple model. This approach is one possible interpretation of the empowerment mandate of many NGOs. The importance of the media in sensitizing voters to the plight of the poor also can be important here, shifting the distribution of preferences to encourage redistribution toward the poor.

The analysis so far has been confined to targeted programs in which transfers flow from taxpayers to the poor. Another possibility is to use diverging views on distribution among the heterogeneous taxpayers to the benefit of the poor by installing a universal rather than a targeted transfer mechanism. Such programs often appear to be designed for the convenience of the middle class (Goodin and Legrand 1985). Gelbach and Pritchett (1995) present evidence supporting the empirical relevance of this interpretation in developing countries.

The model used so far must be modified to study this possibility since it assumes that taxpayers have identical incomes and that programs are financed by a head tax, precluding redistribution among taxpayers. Suppose instead that programs are financed by an income tax and that taxpayers are divided into two groups, rich and middle class, who share the same views about helping the poor. This is the simplest model that can be used to contrast a universal program, which makes an unconditional transfer to all citizens, with a targeted transfer, which minimizes the cost of attaining a particular consumption level.

We continue to assume that taxpayers dominate the political process. If the middle class is small, then the analysis remains largely unchanged. Since the rich cannot gain personally from a universal program, they will always prefer programs that target the poor. Thus the most interesting possibility arises when the middle class grows large. This is the case, mentioned above, in which universalization of programs is considered most likely.

The appendix presents a formal model of this choice; here I summarize just the results. It is sometimes in the interest of the middle classes to select a universal program. This is the case when the redistributive gain to them is large enough. For the poor, a universal transfer program has one immediate consequence—it raises their utility, even if the rich and middle class care only about their consumption. This outcome results from the unconditional nature of such programs, which do not attempt to raise the poor's consumption subject to voluntary participation.[20]

Another way to change the treatment of the poor is to use direct lobbying activity to influence policies. In all political systems, democratic or otherwise, there are groups that can influence policy. These groups can be internal lobbies (composed of citizens) or international actors (including the World Bank and some NGOs).

The formal approach used here follows that introduced by Grossman and Helpman (1994). Suppose that there are some individuals who act as lobbyists and that they influence policy outcomes by making transfers to a policymaker in exchange for changes in z. Assume that the lobbyists' preferences for aid to the poor are similar to those of the rich described above.[21] The transfer function for lobbyist k—that is, the amount of money that the policymaker receives if the change to policy z is implemented—is $T(z, \lambda_k)$. The set of K individuals with this kind of access to the policymaker is denoted by κ. Now consider how these arrangements affect the policy outcome.

Suppose that the policymaker is individual i. Formally, we solve for a Nash equilibrium set of contribution functions to the policymaker and the policy choice. In the particular equilibrium that we will study here, it can be shown that the optimal choice of z by the policymaker solves the first-order condition

$$\frac{1}{n_R}\frac{\partial C(z)}{\partial z} = -\frac{1}{(K' + 1)}\sum_{j \in \kappa \cup \{i\}} \frac{\partial \xi(z, \lambda_R)}{\partial z} \tag{14}$$

where K' is the number of lobbyists who are both taxpayers and lobbyists.[22] The only difference between equation 14 and equation 12 is on the right side. If all lobbyists are taxpayers, then the right side is the average marginal benefit of alleviating

poverty across the group that can influence policy and the policymaker. In this case the scheme will be more generous toward the poor if the group that influences policymakers is, on average, more generous than the policymaker. If some nontaxpayers are in a position to influence the policymaker, then the marginal benefit will be larger, other things being equal, because these groups perceive a benefit to poverty alleviation but incur no cost.

In this model the policymaker earns rent since some kind of favor (here we assume that it is money) is being exchanged for helping the poor. The winners are the people who wanted policy toward the poor to be more generous; the losers are the people who thought policy was already too generous. Equation 14 is the theoretical underpinning of the idea that pro-poor lobbying can result in more generous programs. In the United States interest group models have been applied most extensively to lobbying for protectionist measures, although Piven and Cloward (1977) suggest that lobbying also can affect treatment of the poor. Whether lobbying has in fact led to more generous support for the poor in developing countries is an area worth investigating.

This model, while highly stylized, can be thought of as underpinning the way in which external agencies work to change policy toward the poor by offering inducements to policymakers. Another way to think of the model is as conditional aid—offering the policymaker a transfer in exchange for changing the policy outcome. This mechanism may seem corrupt because the policymaker is earning rent—in effect, a bribe for changing policy in favor of the poor. But since the policymaker receives only a small fraction of the cost of transfers to the poor, this approach actually may be an inexpensive way for outside agencies to help the poor. The model respects the distribution of property rights, that is, the right of a particular individual to change policy. And it is this right that allows the policymaker to earn rent.[23] For this reason it may seem as though it would always be better for lobbyists to make direct transfers to the poor, rather than use the lobbying process. But this is not necessarily the case. First, as noted above, transfers from an NGO can crowd out government support. More important, the activity here raises z, which only happens with direct transfers from NGOs after the government has withdrawn all support from the program.

This discussion of the political economy of program design is necessarily sketchy and incomplete. But many of the issues raised here are important. The main point is that treatment of the poor depends on social decisionmaking procedures, and the extent to which the interests of the poor are represented in these processes cannot be divorced from how well policy is directed toward aiding the poor. It is important for those who are giving policy advice to understand how this decisionmaking works and to know the constraints on helping the poor in a particular society—this is the essence of a political economy approach.

Conclusion

Most economists studying poverty alleviation policies take a technocratic approach to policymaking, ignoring the context in which policy is made. Political economy

approaches often seem remote from the real concerns about how information and incentives feed into program design. As this article suggests, however, the divide between these two concerns can be bridged.

A positive theory of poverty alleviation programs can begin with an approach that focuses on cost-minimizing programs. This approach acknowledges that policymakers can learn from models of government optimization in which constraints on policy choice are rigorously specified and their implications derived. It also provides a satisfying conceptual basis for targeting. Thus the extensive analytical literature on poverty alleviation policies has much to say that is important and relevant.

The other important element is the institutions that implement policies. Although the discussion presented here of the role of NGOs and governments in determining policy is preliminary, it does sketch out some of the modeling issues. The principal aim is to initiate a wider investigation of policy formation by development economists (see, for example, World Bank 1996).

Perhaps the main limitation of the framework used here is that it ignores dynamic considerations, particularly changes over time in the productive abilities of the poor in response to policies. Studies of this kind, both theoretical and empirical, are still in their infancy. However, the conceptualization of policy problems presented here should prove helpful in developing an understanding of policy design in that context.

Many people in many countries believe that governments do not do enough for the poor. It is possible to view targeting as the solution, but the context in which antipoverty policy is decided is crucial. Giving better advice about incentives and information constraints is important. But it is essential to consider the full set of constraints on policy, and how these can be changed.

Appendix. Tradeoff between Universal and Targeted Programs

Suppose that there are three groups—the poor (P), the middle class (M), and the rich (R). Let the number of poor, middle class, and rich be n_P, n_M, and n_R, with the total population being N. Assume that the middle class is a majority of the population. Any transfer program is financed by a tax on the income of the rich and the middle class. The preferences of the middle class and rich are as described for the rich in the main text: $x_i - \xi(z, \lambda_i)$, where $i \in \{M, R\}$.

In a universal scheme the transfer paid to all groups is denoted by T_U. If y_M is the income of the middle class and y_R is the income of the rich, then this will satisfy the government budget constraint $T_U N = t_U (y_M n_M + y_R n_R)$. The assumption that the middle class is a majority implies that they will get to choose policies. Then the optimal universal program will solve:

$$\text{(A.1)} \quad V_U^M = \max_{t_U} \{(1 - t_U)y_M + t_U(y_M\beta_M + y_R\beta_R) - \xi(t_U(y_M\beta_M + y_R\beta_R) + y_P, \lambda_M)\}.$$

where β_i is the fraction of type i's in the population. Given the simplifying assumptions, it is clear that this will yield the optimal choice of $t_U = 1$.

Under a targeted scheme the government budget constraint is $C(z) = t_T\,(y_M n_M + y_R n_R)$. Hence the middle class's preferred policy choice solves:

$$\text{(A.2)} \qquad V_T^M = \underset{z}{Max}\left\{\left(1 - \frac{C(z)}{(y_M n_M + y_R n_R)}\right)y_M - \xi(z, \lambda_M)\right\}.$$

Thus the model would predict the use of a universal scheme on political economy grounds if $V_U^M > V_T^M$.

Our first observation comes from the fact that a universal scheme, unlike a targeted scheme, does not make the receipt of transfers conditional, so the poor have a higher utility level under a universal scheme. A universal scheme would be the expected choice as inequality among taxpayers increased. To see this, consider a mean-preserving spread in the incomes of the taxpayers—that is, holding $y_M n_M + y_R n_R$ fixed while lowering y_M and raising y_R. Increasing the income spread will lower the utility of the targeted program relative to a universal program. As income among the taxpayers becomes more equal the middle class could be expected to favor a targeted program for similar reasons.

Notes

1. Ernest Aryeetey's article in this volume deals with programs that provide credit. The framework discussed here also would be relevant to the long-run program design issues, about which much less is known both theoretically and empirically.

2. These are taken as given in this analysis. The analysis could be extended to include interventions aimed at changing these rates.

3. For ease of exposition, I do not specify the assumptions needed for everything to be "well-behaved" in what follows.

4. For simplicity, this model ignores the labor supply decision of the rich, assuming that they supply labor inelastically.

5. We assume that this function is decreasing in x_P, which says that the rich prefer the poor to enjoy more consumption, other things being equal, and that it is increasing in λ_R so that a larger value of this parameter represents greater sympathy for the poor.

6. This is defined implicitly by $z - h\,(y^*_P/a_P) = u_P$.

7. These are labels only, and no normative connotations are implied.

8. Formally,

$$b_U + y_U - h\left(\frac{y_U}{a_U}\right) \geq b_D + y_D - h\left(\frac{y_D}{a_U}\right).$$

9. Putting these two cases together, we obtain the following benefit level for the undeserving:

$$b^*_U = \max\left\{0, z - h\left(\frac{y^*_D}{a_U}\right) - u_U\right\}.$$

10. Then the incentive constraint for the undeserving is

$$b_U + y_U - h\left(\frac{y_U}{a_U} + c_U\right) \geq b_D + y_U - h\left(\frac{y_D}{a_U} + c_D\right).$$

11. Models of targeting that emphasize the potential income cost of public sector work are rare. Datt and Ravallion (1994) develop a structural model of household labor allocation that gives credence to the view that there is a loss of home production due to participation in a workfare program. This is among the few applied studies of targeted interventions that takes the modeling of incentive consequences seriously.

12. This is not always the case—consider the large literature on the importance of private transfer activity between family members. There are good reasons to believe that indigenous systems of social support predate state support and have continued despite extensive state intervention in this area.

13. To simplify matters, this discussion assumes that NGOs make cash transfers to the poor.

14. The scheme must also satisfy a pair of constraints:

$$x_i - h(y_i/a_P) \leq u_P \text{ for } i \in \{G, N\}$$

which says that the poor would not wish to deviate to accept a transfer from just one of the groups.

15. Because they are kept at their reservation utility constraint, however, their utility is no higher. This is because we have assumed that the NGO cares only about the poor's consumption rather than their more general well-being. Otherwise the poor could gain in utility as well.

16. This is true if $b^*_U > 0$. Let γ be the number of deserving poor, then we can calculate that in the untargeted program

$$\frac{\partial C(z)}{\partial z} = n_P\left(1 - \frac{a_D}{h'(y^*_D/a_D)}\right)$$

and in the targeted program

$$\frac{\partial C(z)}{\partial z} = n_P\left(\gamma\left(1 - \frac{a_D}{h'(y^*_D/a_D)}\right) + (1-\gamma)\frac{\partial b^*_U}{\partial z}\right).$$

It is straightforward to show that

$$\frac{\partial b^*_U}{\partial z} = \left(1 - \left(\frac{a_D}{h'(y^*_D/a_D)}\right)\left(\frac{h'(y^*_D/a_U)}{a_U}\right)\right) \text{ and } \frac{h'(y^*_D/a_U)}{a_U} < 1.$$

Hence the marginal cost is higher in the targeted case.

17. This result is like the one suggested in Gelbach and Pritchett (1995). The mechanism here is very different, however. In a model with income effects there may be some offsetting effect due to the average cost of poverty alleviation having fallen. Note also that the results here are entirely in terms of consumption levels. A more welfarist model could deliver parallel results for the utility of the poor.

18. Formally, this solves

$$\frac{1}{n_R} \cdot \frac{\partial C(z^*_i)}{\partial z} = -\frac{\partial \xi(z^*_i, \lambda_i)}{\partial z}.$$

19. This model implicitly assumes that the poor have no voice in policy. One could clearly add this possibility to the model. The median voter result here would require the assumption that the cross partial derivative between z and λ_i in $\xi(z, \lambda_i)$ be positive. If individuals differ in more than one characteristic, then the median voter result typically has no force. There is little agreement on how to characterize political equilibrium in such.

20. Although these results suggest that universalization is good for the poor, this is not always the case. If the poor also are forced to share the tax burden, universal programs may be worse for them (see Epple and Romano 1996).

21. However, they need not also be taxpayers in the society. Although they are taxpayers (that is, groups of rich in the society where valuations are heterogeneous) in this case, they can also be external to the society.

22. The derivation of this equation and full details of the assumptions required for it to hold are available from the author on request.

23. In Besley and Coate (1996) the policymaker is allowed to change in response to lobbying, which allows for some interesting possibilities.

References

Atkinson, Anthony. 1987. "Income Maintenance and Social Insurance." In Alan Auerbach and Martin Feldstein, eds., *Handbook of Development Economics,* vol. 2. Amsterdam: North-Holland.

Atkinson, Anthony. 1995. 1995. "On Targeting Social Security: Theory and Western Experience with Family Benefits." In Dominique van der Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence.* Baltimore, Md.: The Johns Hopkins University Press for the World Bank.

Bernheim, B. Douglas. 1986. "On the Voluntary and Involuntary Provision of Public Goods." *American Economic Review* 76: 25–59.

Besley, Timothy, and Stephen Coate. 1992. "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty Alleviation Programs." *American Economic Review* 82 (1): 249–61.

———. 1995. "The Design of Income Maintenance Programs." *Review of Economic Studies* 62(2): 187–222.

———. 1996. "Lobbying and Welfare in a Representative Democracy." London School of Economics, Department of Economics.

Boone, Peter. 1996. "Political and Gender Oppression as a Cause of Poverty." London School of Economics, Department of Economics.

Bruno, Michael, Lyn Squire, and Martin Ravallion. 1995. "Equity and Growth in Developing Countries: Old and New Perspectives." Policy Research Working Paper 1563. World Bank, Washington, D.C.

Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent.* Ann Arbor: University of Michigan Press.

Cornea, Giovanni Andrea, and Frances Stewart. 1995. "Two Errors of Targeting." In Dominique van der Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence.* Baltimore, Md.: The Johns Hopkins University Press for the World Bank..

Dasgupta, Partha. 1994. *An Enquiry into Well-being and Destitution.* Oxford: Oxford University Press.

Datt, Gaurav, and Martin Ravallion. 1994. "Transfer Benefits from Public Works: Evidence from Rural India." *Economic Journal* 104 (6): 1346–69.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Dreze, Jean, and Amartya Sen. 1990. *Hunger and Public Action.* Oxford: Oxford University Press.

Edwards, Michael, and David Hulme. 1995. *Non-Governmental Organizations—Performance and Accountability: Beyond the Magic Bullet.* London: Earthscan Publications.

Epple, Dennis, and Richard Romano. 1996. "Public Provision of Private Goods." *Journal of Political Economy* 104(1): 57–84.

Gelbach, Jonah B., and Lant H. Pritchett. 1995. "Does More for the Poor Mean Less for the Poor: the Politics of Tagging." Policy Research Working Paper 1523. World Bank, Washington, D.C.

Goodin, Robert, and Julian LeGrand, eds. 1987. *Not Only the Poor: The Middle Classes and the Welfare State.* London: Allen and Unwin.

Grossman, Gene, and Elhanan Helpman. 1994. "Protection for Sale." *American Economic Review* 84(4): 833–50.

Kingma, Bruce Robert. 1989. "An Accurate Measurement of the Crowd Out Effect, Income Effect, and Price Effect of Charitable Contributions." *Journal of Political Economy* 97 (5): 1197–1207.

Kramer, Ralph M. 1981. *Voluntary Agencies in the Welfare State.* Berkeley: University of California Press.

Lipton, Michael, and Martin Ravallion. 1995. "Poverty and Policy." In Jere Behrman and T. N. Srinivasan, eds., *Handbook of Development Economics,* vol. 3a. Amsterdam: North-Holland.

Mueller, Dennis C. 1989. *Public Choice II.* Cambridge: Cambridge University Press.

Nichols, Albert, and Richard Zeckhauser. 1982. "Targeting Transfers via Restrictions on Recipients." *American Economic Review* (Papers and Proceedings) 72 (2): 372–77.

Roberts, Russell D. 1984. "A Positive Model of Private Charity and Public Transfers." *Journal of Political Economy* 92 (1): 136–48.

Sen, Amartya. 1995. "The Political Economy of Targeting." In Dominique van der Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence.* Baltimore, Md.: The Johns Hopkins University Press for the World Bank.

World Bank. 1996. *The World Bank Participation Sourcebook.* Washington, D.C.

Zeckhauser, Richard. 1971. "Optimal Mechanisms for Income Transfer." *American Economic Review* 61 (8): 324–34.

Comment on "Political Economy of Alleviating Poverty: Theory and Experience," by Timothy Besley

Arsenio M. Balisacan

Timothy Besley's article should become required reading for students of development economics everywhere. Although the idea of targeting limited resources to achieve poverty alleviation objectives is quite simple, realities and myths have often conspired to muddle it. Some of these myths are unveiled in Besley's article, and here I offer a few more, especially within the context of developing countries.

Targeting

Much of the literature on targeting, including Besley's article, focuses on industrialized countries. Although the theories behind targeting are essentially the same for developing countries as for industrial countries, the practical importance of certain elements of targeting varies significantly between the two groups of countries. Developing countries, for example, have far more severe information problems because of their underdeveloped infrastructure and low levels of human capital development. At the same time the issue of targeting may be much simpler in developing countries because the majority of the poor—and the poorest of the poor—are concentrated, usually in agriculture or in resource-poor regions dependent on farming. Under these circumstances indicator-based targeting could significantly reduce poverty.

Conventional economic analysis of targeting often proceeds from the assumption that poverty alleviation budgets are fixed and that the objective is to design a poverty alleviation scheme that minimizes poverty using these budgets. Besley's article extends that approach, exploring various institutional arrangements that influence poverty alleviation budgets. The model he proposes for thinking about policy choices is conceptually elegant. Its implicit assumption is that policy choices, especially with respect to poverty alleviation budgets, are clear-cut.

Arsenio M. Balisacan is professor of economics at the University of the Philippines at Diliman.

Annual World Bank Conference on Development Economics 1996

In practice, however, things are not so simple. In many developing Asian countries, for example, policymakers view economic growth and industrialization not so much as ends in themselves as the (necessary) means to alleviate poverty. The hard policy choices involve the slicing of the budgetary pie into funds for direct poverty alleviation programs and funds for indirect poverty alleviation programs (such as growth and industrialization). Part of the policy debate concerns whether money spent on direct interventions is more (or less) effective at reducing poverty than money spent on indirect programs. In this debate timing of policy actions and results is critical; it is, in large part, what the political economy of poverty alleviation is all about.

Some analysts seem to believe that efforts to alleviate poverty are a phenomenon that emerged only in the 1970s and 1980s. In fact, efforts to reduce poverty and increase economic growth have been present in many developing Asian countries since World War II. Only the emphases and strategies have changed over the years. During the 1950s and 1960s, rapid economic growth through import-substituting industrial development was the central focus of efforts aimed at improving living standards, at least in such countries as India (Srinivasan 1996) and the Philippines (Balisacan 1996). During the 1970s and 1980s the focus shifted to growth with equity. Then the urgent need to address the macroeconomic difficulties arising from domestic and global shocks during the 1980s moved development priorities away from the direct meeting of basic needs. Now, with these difficulties receding (at least for many developing Asian countries), attention is shifting back to growth with equity—with a new wrinkle.

Nongovernmental Organizations

The 1990s have seen an unprecedented involvement of nongovernmental organizations (NGOs) in the poverty alleviation efforts of many developing Asian countries. This development has been influenced by NGOs' increased access to financial support from developing country governments on the one hand and from bilateral and multilateral donors on the other. Indeed, NGO-government collaboration is the current buzzword in policy discussions on poverty alleviation and environmental protection in developing countries.

It might be too early to draw conclusions on how effective this collaboration is in achieving poverty alleviation objectives. But Besley's analytical results give us an idea about what we can expect. In particular, the proposition that an income transfer from NGOs to the poor can crowd out what would otherwise be transferred by the government is very interesting. In a related vein, evidence of crowding out of interhousehold transfers by public transfers in developing countries is beginning to emerge (Cox and Jimenez 1995). I would not be surprised if similar evidence of crowding out were found in NGO-government collaboration in a number of developing countries.

The implicit assumption in Besley's article is that NGOs provide their own resources—that is, they do not compete with government agencies for funds from

the same pool. In practice, NGOs in developing countries get most of their resources from the public budget or from external sources (such as multilateral aid agencies) that might otherwise have been part of the government's pool of funds for poverty alleviation. If NGOs are better at alleviating poverty than government agencies—that is, if they achieve more poverty reduction than the government does using the same amount of money—then that is welfare improving.

But is that the case? Consider the Philippine Help for the People (Tulong sa Tao) program. This program was intended to address the basic needs of poor families, particularly in the nonagricultural rural sector, by making credit available for employment and income-generating activities. Technical and financial support came from a multilateral development agency, but NGOs served as intermediaries for credit and technical assistance to address directly the needs of the poor. Although the NGOs involved were among the country's most development-oriented, assessment found that the average size and frequency of loans made under the program did not vary across income classes. Moreover, only about a third of program beneficiaries were poor by Philippine standards. Even more surprising, about a third of the beneficiaries had incomes ranging from three to more than twenty times the official poverty line. Thus poverty reduction attributable to the program was very limited. Of the group of program beneficiaries, most of whom were not poor to begin with, only 7–9 percent moved out of poverty after program implementation. It would be hard to argue that this performance was superior to what could have been achieved by sending a helicopter to drop bales of money on the countryside.

This leads to another element that has not entirely escaped Besley's model—that NGOs respond to opportunities not only for influence peddling but for rent seeking as well. The experience with many NGOs, at least in some developing Asian countries, has been quite disappointing so far. The spectacular growth of NGOs in the Philippines—from only a few hundred in the late 1980s to over 20,000 in the early 1990s—has coincided with the emergence of government policies encouraging the sourcing of both government and externally generated funds through NGOs. Politicians and their spouses have established their own NGOs; projects of dubious value have come and gone, all in the name of the poor.

Empirical work on targeting in developing countries—including the role of NGOs—is clearly wanting. Besley provides a useful framework that could guide such work. The World Bank could help expand our knowledge of what works in poverty targeting and what does not.

Economists are accustomed to thinking about market failures and government failures in the context of policy design. Soon we may be adding a third type of failure to our vocabulary: NGO failures. But the fourth type—economists' failures—may not be far behind if we fail to inform public policy about what works and what does not in poverty alleviation.

References

Balisacan, Arsenio M. 1996. "Philippines." In M. G. Quibria, ed., *Rural Poverty in Developing Asia*, vol. 2. Manila: Asian Development Bank.

Cox, Donald, and Emmanuel Jimenez. 1995. "Private Transfers and the Effectiveness of Public Income Redistribution in the Philippines." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence*. Baltimore, Md.: The Johns Hopkins University Press.

Srinivasan, T. N. 1996. "Economic Liberalization and Economic Development: India." Yale University, Economic Growth Center, New Haven, Conn.

Comment on "Political Economy of Alleviating Poverty: Theory and Institutions," by Timothy Besley

Karla Hoff

Any program designed to alleviate poverty must, at least implicitly, answer the questions: What are the sources of inequality in earnings? And how is inequality between individuals generated? In Timothy Besley's article, the answer is differences in intrinsic ability. Working from this view, Besley provides a clear exposition of the principles of transfer policy aimed at increasing the resources of the poor at the least budgetary cost.[1]

This approach seems overly narrow, mainly because it is empirically dubious to rule out imperfections in risk and capital markets as a source of poverty. Once these market imperfections are taken into account, poverty alleviation policies can be seen as having feedback effects on individuals' capacity to earn income. By setting a floor under incomes, poverty alleviation policies increase individuals' ability to make production choices that entail risk, to invest in their children, to avoid fire sales of assets, and to pledge collateral. Redistribution policies can loosen the constraints that keep poor people poor, and this has implications for the design of transfer policy.

Poverty Alleviation Programs As Insurance

There are many examples of the efficiency costs that arise when low-wealth people are active in a high-risk environment without good instruments to smooth their consumption.[2] Households with low wealth, whose consumption levels are more vulnerable to income shocks than those of rich households, make production choices that sacrifice expected profits for greater self-protection. Binswanger and Rosenzweig (1993) quantify this effect. They studied semiarid areas of India, where volatile weather patterns implied high income risk to the people who farmed there, and used the data to determine the effect on average profits as the environment became riskier. They found that an increase in risk, measured as an increase by one standard deviation in the coefficient of variation of the monsoon onset date, induced income-smoothing efforts that reduced average profits by 35 percent for the quarter

Karla Hoff is assistant professor of economics at the University of Maryland at College Park.

Annual World Bank Conference on Development Economics 1996

of households with the lowest current wealth, by 15 percent for households with mean wealth, and hardly at all for the quarter of households with the greatest current wealth.[3]

Rosenzweig and Wolpin (1993) examine the loss in income that is due specifically to the fact that low-wealth individuals adopt capital accumulation and divestment strategies that smooth their consumption. They find that a large share of the savings of Indian farmers is held in bullocks, which are bought and sold in well-developed regional markets. Less than one-third of mid-size farmers own a pump, although a pump would increase farm profits by more than 70 percent on average. Bullocks serve as a buffer stock—a store of value—whereas pumps cannot.[4] After an adverse weather shock, small and mid-size farmers sell bullocks to protect consumption. The depletion of productive assets in turn lowers their future incomes and exacerbates the variance of incomes. Thus a transitory negative income shock can have long-term effects that depress household productivity.[5]

Rosenzweig and Wolpin's study was inspired by the observation that bullocks change hands rapidly across households and villages that use similar technology, a finding that would not be expected in a setting of complete markets. Sheep have been observed to serve a similar purpose as a store of value in rural areas of Indonesia (Nerlove and Soedjana 1996).

The general point is that with imperfect risk markets and constraints on borrowing for consumption needs, poor households trade off increases in expected incomes for reduced risk. This argument suggests that poverty alleviation policies, such as India's popular food-for-work programs, can reduce long-run poverty by increasing individuals' ability to cope with risk. This is an empirically testable proposition. Cain (1981) compares the experiences of a village in Bangladesh and three villages in India after natural disasters (a flood in Bangladesh, drought in India). In India public relief employment helped cushion the effects of the disaster; together with institutional credit, it reduced distress sales of land. In the Bangladesh village, where comparable mechanisms to help the poor had not evolved, many distress sales of land occurred. The result was polarization in ownership of land, which aggravates income inequality and inequality of access to credit.

The main implication of these findings is that policies that are ostensibly short-term poverty alleviation strategies—such as food-for-work programs and food ration coupons in India and many other countries—may actually raise the long-run average earning power of the poor, reduce long-run inequality in asset holdings, and through both channels reduce the need for poverty alleviation policies later on. Poverty alleviation programs, particularly if households are confident that the programs will be available even under the worst of weather extremes, provide insurance.

Poverty alleviation programs also may increase the income-earning power of the next generation of adults. Downward shocks to income in poor households have a measurable effect on the health and body size of children (Glewwe and Jacoby 1995; Alderman and others 1996; and Morduch 1995). Early childhood malnutrition delays enrollment in primary school. Both malnutrition and delayed enrollment tend to reduce the lifetime earnings of adults.

Poverty Alleviation, Collateral, and Tenancy Contracts

Certain kinds of transfers to the poor may even enhance the efficiency of credit markets by providing a collateralizable asset, or may improve the incentives faced by sharecroppers under tenancy contracts. Access to credit depends on the borrower's ability to pledge collateral; transfers can provide a form of collateralizable wealth. In Sri Lanka an experimental program allowed people who participated in a government loan program to pledge their rice ration coupons as collateral; as a result borrowers' repayment rates increased substantially (Sanderatne 1986). In a village in Kerala, India 45 percent of households pledged, in the private informal credit market, the ration cards that gave them access to subsidized "fair price" shops (Platteau and others 1980). Households that did not pledge their coupons had a floor on consumption and some insurance against increases in grain prices. Households that pledged their coupons may have gained greater access to capital. Thus transfers can perform two distinct roles. In the second role collateral is like a catalyst, changing agents' scope for opportunistic behavior and improving the operation of credit markets.

When transfers reduce incentive problems, even marginal changes in the distribution of income may shift outward an economy's feasible production set. Transfers may yield an unambiguous increase in welfare and output even when they are financed by a distortionary tax on labor earnings (Hoff 1994; Hoff and Lyon 1995). Since the distortion in capital markets may be large, it is not surprising that introducing a small distortion in labor-leisure choices that reduces the magnitude of the big distortion may be desirable.

Besley's article highlights the following puzzle: what kind of social welfare function can rationalize a redistribution program that focuses solely on people below a given poverty line? Put another way, if taxpayers want to redistribute income from people who are only slightly above the poverty line to those below it, then why are they not also concerned about redistribution between people in the top quarter of the income distribution, say, and people who are barely above the poverty line? Such a lack of concern cannot be explained within a framework of perfect markets and the standard assumption that the marginal utility of income is decreasing. But they can be explained within a framework in which people of low income and wealth are hamstrung by imperfect risk and capital markets.

Recent work in economic theory develops models that define a kind of "poverty line" not in the usual way—in terms of consumption—but in terms of the minimum resources that people need in order to gain access to markets and to contractual forms that elicit high effort (Aghion and Bolton forthcoming; Hoff and Lyon 1995; Legros and Newman forthcoming; and Hoff 1996). Starting from assumptions about imperfect information between transactors in markets (workers and employers, borrowers and lenders) and about indivisibilities in technology, this work derives wealth thresholds below which people will not be able to participate in credit markets or in certain kinds of labor contracts. Below that threshold, people experience discrete losses in welfare and income. These losses provide a normative basis for setting a reduction in the number of people below the "poverty line" as a social objective.

Empirical studies of agricultural productivity in developing countries have started to confirm some of the predictions of this theory. Tenant farmers, for example, have contracts in the form of the share of output that they retain, with the remaining share being paid to the landlord. Laffont and Matoussi (1995) observe that in rural Tunisia a tenant's share of output takes one of four values: $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$, or 1. From their empirical work Laffont and Matoussi drew two conclusions. First, the higher the working capital of a tenant at the beginning of the period, the closer he or she was to a rent contract. Thus a pure rental contract was observed as long as the tenant was sufficiently wealthy. Second, holding everything else constant (including all observed inputs), switching the average tenant farmer from a sharecropping contract to a nonshare contract would increase output by 50 percent. Laffont and Matoussi interpreted the increase in output to reflect differences in incentives for high effort between the two kinds of contracts.

These estimates would seem to promise huge gains from policies that shift the division of agricultural output in tenants' favor. An empirical test of that idea was offered by the agricultural tenancy reform, called Operation Barga, that began in 1977 in the Indian state of West Bengal.[6] A primary goal of the reform was to change the division of output between landlords and tenants in favor of tenants. As a result of Operation Barga the division of output on many tenancies changed from an even split to a 70–30 split in favor of tenants. By 1982 the reform had reached about half the state's sharecroppers, and over the next decade West Bengal achieved a breakthrough in agricultural growth. Banerjee and Ghatak (1996) estimate that more than a third of West Bengal's growth in agricultural production during 1981–92 was due to the tenancy reform.

Thus the land tenancy reform not only redistributed income to the poor, it also changed the incentives of the poor to be productive by changing the tenancy contracts under which they supplied labor. In this way redistribution substantially increased the income generated by the poor.

Human Capital Spillovers and Political Economy

Another type of market imperfection—human capital spillovers—seems important to targeting theory and has implications for the political feasibility of redistribution as well. Many economists believe that a person's investment in skills benefits not just that person but also other agents indirectly through a production externality. For example, high-skill labor may be more productive when low-skill labor is better trained, and any worker may be more productive when his or her co-workers have better skills. If imperfections in capital markets prevent low-income individuals from borrowing to finance an investment in skills, then human capital spillovers provide self-interested high-income taxpayers with a motive for redistribution.

Following Perotti (1993) and Benabou (1995), consider an economy in which the poor cannot invest in human capital without transfers and in which the redistribution policy is determined by the median taxpayer, who has above-average income. If the poor are only slightly poorer than the median taxpayer, then the median

taxpayer has an incentive to support transfers for education for the sake of the human capital spillover benefit. But if the poor are much poorer than the median taxpayer, then the median taxpayer will not support transfers because the cost of the redistribution will be high relative to the spillover benefit. In this case the political equilibrium will not entail transfers to the poor, and economic growth will be slow. This model illustrates the striking idea that when redistribution promotes growth and taxpayers are self-interested, whether high growth is a politically feasible outcome may depend on the initial distribution of income.

Conclusion

The tools described in Besley's article tell us what economists know about the cost-minimizing way to transfer income to a fixed set of individuals who have intrinsically low income-earning power. But any approach to targeting that does not consider how people's income-earning power is determined is necessarily incomplete. Imperfect markets for insurance and capital can keep poor people poor. Poverty alleviation policies can mitigate these market failures. This view suggests new criteria for defining a poverty line and for designing transfer policies. In economies where human capital spillovers are important and the poor cannot finance investment in human capital, this approach yields novel implications for the political feasibility of redistribution and growth.

Notes

1. Besley calls this the technocratic approach to policymaking because the perspective taken is that of a planner with a well-defined social objective. Besley's basic framework is an extension of James Mirrlees's (1971) analysis of optimal taxation in an economy where people differ in their income-earning power but the government cannot differentiate among them. Besley extends this analysis by introducing work requirements that screen people with low income-earning power from the rest and by taking poverty reduction rather than a general utility measure as the social objective.
2. The most important instruments for consumption smoothing are the insurance and savings and credit instruments provided by markets and nonmarket institutions. An overview of informal credit and risk-sharing institutions in developing countries is provided in Besley (1995).
3. Results of a similar magnitude were obtained when inherited wealth was used instead of current wealth levels to define wealth classes.
4. As evidence, consider that in areas where crops were hurt by adverse weather, the probability that a farmer sold livestock increased 34 percent. The consequences for equilibrium growth of the channeling of investment into a buffer stock rather than into a more productive but illiquid investment are explored in Bencivenga and Smith (1991).
5. Morduch (1994) emphasizes the distinction between transitory and persistent poverty, and the consequences of imperfect market structures for the persistence of poverty.
6. *Barga* is a local term for sharecropping. Sharecropping tenancies account for nearly half the net cropped area in West Bengal.

References

Aghion, Philippe, and Patrick Bolton. Forthcoming. "A Theory of Trickle-Down Growth and Development." *Review of Economic Studies.*

Alderman, Harold, Jere R. Behrman, Victor Lavy, and Rekha Menon. 1996. "Child Nutrition, Child Health, and School Enrollment: A Longitudinal Analysis." World Bank, Policy Research Department, Washington, D.C.

Banerjee, Abhijit V., and Maitreesh Ghatak. 1996. "Empowerment and Efficiency: The Economics of Tenancy Reform." Massachusetts Institute of Technology, Cambridge, Mass.

Benabou, Roland. 1995. "Unequal Societies." New York University, Department of Economics.

Bencivenga, Valerie R., and Bruce D. Smith. 1991. "Financial Intermediation and Endogenous Growth." *Review of Economic Studies* 58 (2): 195–235.

Besley, Timothy. 1995. "Nonmarket Institutions for Credit and Risk Sharing in Low-Income Countries." *Journal of Economic Perspectives* 9 (3): 115–27.

Binswanger, Hans P., and Mark R. Rosenzweig. 1993. "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments." *Economic Journal* 103 (January): 56–78.

Cain, Mead. 1981. "Risk and Insurance: Perspectives on Fertility and Agrarian Change in India and Bangladesh." *Population and Development Review* 7 (3): 435–74.

Glewwe, Paul, and Hanan G. Jacoby. 1995. "An Economic Analysis of Delayed Primary School Enrollment in a Low-Income Country: The Role of Early Childhood Nutrition." *Review of Economics and Statistics* 77 (1): 156–69.

Hoff, Karla. 1994. "The Second Theorem of the Second Best." *Journal of Public Economics* 25: 223–42.

———. 1996. "Market Failures and the Distribution of Wealth: A Perspective from the Economics of Information." *Politics and Society* 24 (4): 411–32.

Hoff, Karla, and Andrew Lyon. 1995. "Non-Leaky Buckets: Optimal Redistributive Taxation and Agency Costs." *Journal of Public Economics* 58 (3): 365–90.

Laffont, Jean-Jacques, and Mohamed Matoussi. 1995. "Moral Hazard, Financial Constraints and Sharecropping in El Oulja." *Review of Economic Studies* 62 : 381–99.

Legros, Patrick, and Andrew Newman. Forthcoming. "Wealth Effects, Distribution, and the Theory of Organization." *Journal of Economic Theory.*

Mirrlees, James. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (4): 175–208.

Morduch, Jonathan. 1994. "Poverty and Vulnerability." *American Economic Review* 84 (2): 221–25.

———. 1995. "Income Smoothing and Consumption Smoothing." *Journal of Economic Perspectives* 9(3): 103–14.

Nerlove, Marc, and Tjeppy D. Soedjana. 1996. "Slametans and Sheep: Savings and Small Ruminants in Semi-Subsistence Agriculture in Indonesia." University of Maryland, Department of Agricultural and Resource Economics, College Park.

Perotti, Roberto. 1993. "Political Equilibrium, Income Distribution, and Growth." *Review of Economic Studies* 60: 755–76.

Platteau, Jean-Philippe, J. Murickan, A. Palatty, and E. Delbar. 1980. "Rural Credit Market in a Backward Area: A Kerala Fishing Village." *Economic and Political Weekly* (October): 1765–80.

Rosenzweig, Mark, and Kenneth Wolpin. 1993. "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Assets in Low-Income Countries: Investments in Bullocks in India." *Journal of Political Economy* 101 (2): 223–44.

Sanderatne, Nimal. 1986. "The Political Economy of Small Farmer Loan Delinquency." *Savings and Development* 10: 343–53.

Floor Discussion of "Political Economy of Alleviating Poverty: Theory and Institutions," by Timothy Besley

Whenever we try to introduce antipoverty programs in Africa, said a participant, the problem is budget. Thus there is always a tradeoff between the amount a country spends on short-term and on long-term poverty alleviation programs. That is an issue that I thought would be reflected in a discussion of least-cost solutions, since presumably they free more resources for long-term poverty alleviation and growth-oriented programs. Is there another approach? How can countries strike a balance between short-term and long-term solutions

Besley responded by saying that he had been asked to focus on short-term poverty reduction; the next session (on Ernest Aryeetey's article) focused on long-term issues. By dividing the agenda in this way, however, both articles failed to address precisely the question the participant had asked. The cost-minimization model could be applied to a dynamic environment, in which individuals invest in their abilities over time. Besley was not sure what the results would be and how the models would work, but he agreed on the need for a model that looked at both short-term and long-term poverty alleviation.

Another participant asked Besley if decentralized programs—in which localized governance allows local residents to participate more fully in poverty alleviation efforts—would be more effective at reaching and serving the poor, since even the best nongovernmental organizations (NGOs) cannot reach the poor in all parts of a country. Besley replied that there is an extensive literature on U.S. experience with the issue, begun in the 1950s, when Tiebout (1956) argued that a decentralized government structure should deliver the services that people who participate in government want. A good way to accurately reflect those preferences would be through as much decentralization as possible, with microunits making the decisions. But there is also a literature suggesting a big tradeoff between local decisionmaking and redistribution.

One complication is migration. In the United States, for example, there had been a significant inflow of poor people into the state of Wisconsin, which for years had

This session was chaired by Callisto E. Madavo, director, East Asia and Pacific Department, at the World Bank.

Annual World Bank Conference on Development Economics 1996

more generous welfare benefits and social services than other states. Migration is also the subject of considerable debate in Europe. As Europe unifies and labor becomes freely mobile across countries, are people going to migrate to countries where social services and safety nets are more generous? Although there is not much evidence of such migration now, except perhaps in higher education, it could well happen in the future. It is also the case that in a decentralized system certain kinds of public goods from which the poor benefit disproportionately simply would not be represented. And if migration is allowed, the amount of redistribution at the local level could become extremely limited. There must be fifteen or twenty countries that have undertaken serious decentralization efforts. These are interesting efforts, but when it comes to poverty alleviation policy, there will always be these tradeoffs among decentralization, migration, and redistribution.

A participant asked Besley if he could discuss distribution not to households but within households. Besley said that the only empirical study of the subject he knew well examined child benefits in the United Kingdom. Under the U.K. scheme mothers receive a cash payment for every child in the household; the agency distributing the money does not release it to fathers. The evidence suggested that this approach has a significant effect on household consumption patterns. How income enters a household, how it is disbursed, and who earns it may have important implications for poverty alleviation efforts.

The same participant asked the panel about the potential for self-targeting in poverty alleviation programs. Besley said he assumed that the participant meant something like rural public works programs in which people self-select to participate, which he had touched on in his article. There seems to be a presumption, Besley said, that the best of all schemes is one in which income is observable, so that if the goal is efficiency administrators are able to get an accurate fix on the resources of the poor. Self-targeting is seen as an alternative to collecting increasingly detailed information about the poor. Although some studies have found that self-targeting programs are able to achieve moderately accurate targeting at relatively low administrative costs, the evidence is limited.

The best work on the subject, according to Besley, has been done by Martin Ravallion and others at the World Bank (see, for example, Datt and Ravallion 1993). They have examined various schemes to see what can be inferred about people's income by their decision to participate in a program—to see whether programs using the test of a work requirement, but with no knowledge of people's income, succeed in selecting the poor. And there is some evidence that that is the case. But there is also an interesting side issue: if a scheme provides work for people, does that crowd out work on the farm or in other home-based productive activities?

Self-targeting becomes less important as countries become more administratively sophisticated, observed Besley. There is very little self-targeting in industrial countries. And yet workfare programs—which require some kind of work in return for social services—are one of the biggest reforms being discussed in the United States and Europe. Workfare can be seen as analogous to the rural public works projects in developing countries. There could be a strong argument for self-targeting even in

very sophisticated schemes in which people's circumstances can actually be monitored. Targeting in developing countries need not be an either-or arrangement, added Arsenio Balisacan (discussant). For example, regional targeting can be combined with self-targeting..And that approach need not be inconsistent with the cost-minimization approach.

Karla Hoff (discussant) noted that very little work had been done on another, related issue: if there really are market imperfections that make poor people poor, then shadow prices for capital must be different for the poor than for the rich. And that suggests one way to try targeting funds to the poor. NGOs in industrial countries have been extremely active in subsidizing limited-equity housing. With limited-equity housing, residents own the structure and improvements, but not the land. People can live in the house for as long they want and pass it on to their heirs, but they cannot cash in any capital gains on the land. They can only cash in capital gains on their improvements, and only to a limited extent. Such a scheme basically gives people money up front and takes it back at the end. Those kinds of subsidies would be attractive to the poor but not to the rich, who can self-finance rather than go out and borrow at a high rate. Supporters of limited-equity housing have known this for a long time, and Hoff thought that this insight could be incorporated into a model of self-targeting.

References

Datt, Gaurav, and Martin Ravallion. 1993. *Income Gains for the Poor from Public Works Employment—Evidence for Two Indian Villages.* LSMS Working Paper 100. Washington, D.C.: World Bank.

Tiebout, Charles M. 1956. "A Pure Theory of Local Government Expenditures." *Journal of Political Economy* 64 (5): 416–24.

Rural Finance in Africa: Institutional Developments and Access for the Poor

Ernest Aryeetey

Because the obstacles to financial market development in Africa are not simply policy induced but have structural and institutional origins that cannot be addressed with macroeconomic policy reform alone, financial market liberalization has not been effective in improving credit delivery. Many directed credit programs also have failed to improve rural credit access. Market integration is needed to loosen the structural and institutional constraints to efficient credit allocation. Integration through a financial systems development approach would allow formal credit institutions with surplus financial resources to link up with semiformal institutions that lack such resources but have a comparative advantage in reaching the poor. Strengthening ties between the formal and informal sectors is critical if credit is to be allocated equitably and efficiently. Policy can be used to enhance the development of linkages between market segments by using the fiscal system and the regulatory and supervisory systems to provide incentives for formal institutions to allocate wholesale credit through informal agents.

The current paradigms of financial market liberalization and directed credit have proved inadequate for addressing the structural and institutional constraints on Africa's financial markets. These constraints arise from the lack of adequate information for lenders and the absence of credible contract enforcement mechanisms. When rural economic agents do not expect to profit from entering credit and insurance markets, they avoid these markets and finance their activities with funds raised from other activities. The result is a much smaller financial system. The risk reduction behavior of lenders remaining in the market leads to fragmented markets, ineffective credit allocation, and a wide credit gap. The constraints on the financial system cannot be removed by macroeconomic and financial sector reforms alone. Nor can innovative schemes that apply the operating principles of the informal sector within more formal structures circumvent the problems of the financial system.

Ernest Aryeetey is senior research fellow at the Institute of Statistical Social and Economic Research, University of Ghana.

Annual World Bank Conference on Development Economics 1996

The extent to which so-called innovative credit retailing schemes have increased access to credit by the rural poor in Africa remains unclear. Such schemes have been implemented at a time when macroeconomic and financial sector reforms have put both formal and informal financial institutions under considerable strain in their attempts to deliver credit cost-effectively. The need for intervention has been justified by market failures that have persisted even after reforms (Jackelen and Rhyne 1991). Indeed, much recent analytical work has underscored the strong impact of market failures (Hoff and Stiglitz 1990) and provided a rationale for some intervention, even if the nature of the intervention has not been specified (Besley 1994).

Yet the World Bank (1989) and many economists have often been unsupportive of intervention in credit markets, believing that liberalization of financial markets is the best way to free up resources for expanding credit markets. Many economists within the Bank have supported the financial market paradigm under which governments that adopt directed credit schemes are urged to attack the conditions—including market imperfections and extreme income inequality—that made directed credit desirable (Caprio, Atiyas, and Hanson 1994). In recent years, however, the Bank's stance on intervention has mellowed considerably, and some project officers now support the directed credit paradigm (see, for example, Holt and Ribe 1991 and Yaron 1992).

In Africa, however, the issue is not about choosing between a financial market paradigm and a directed credit paradigm. It is about ensuring that the institutions that are in place for delivering credit to the rural poor are able to function effectively and efficiently after financial markets have been liberalized. Because the obstacles to financial market development are not simply policy induced but have structural and institutional origins that cannot be addressed with macroeconomic policy reform alone, financial market liberalization has not been effective in improving credit delivery (Nissanke and Aryeetey 1995). In Africa and elsewhere many directed credit programs also have failed to improve rural credit access (Adams 1984). The approach proposed here is one of market integration through financial systems development. Integration would permit institutions with surplus financial resources—including many African banks (Nissanke and Aryeetey 1995; Popiel 1994)—to link up with institutions that lack such resources but that have a comparative advantage in reaching the poor cost-effectively.

Because of the dearth of research on the functioning of rural financial markets in Africa, I draw mainly on the few available secondary sources to discuss empirical findings. I begin with recent analyses of rural households' production and consumption behavior and their treatment of credit relative to risk and insurance and the systems that they have developed to counter the effects of market failure. In view of the high-risk nature of their environment, households' use of credit is minimal. In the fragmented African credit markets lenders' actions to reduce risk and lower transactions costs create gaps in the market. Recent institutional developments in rural finance have not yet resolved the problem of how to reach large numbers of potential borrowers cost-effectively. Interventions are needed that will lead to increased specialization and significant collaboration among lenders to ensure that

institutions with surplus funds can deliver credit to marginal borrowers through lenders that can achieve significant outreach and sustainability.

Conceptual Developments in the Analysis of Rural Financial Markets

Recent analyses of credit market conditions have tended to rely on trade theory to explain supply and demand in financial services (Besley 1994). Credit and insurance are assumed to respond to gains from trade. Savings represents the intertemporal trade of current consumption for future consumption. Credit is received or extended against a future repayment obligation or claim. Insurance can be viewed as trade across states of nature. Thus people demand savings and credit facilities because they expect a gain over time. The conditions under which that gain will occur are influenced by the nature of information (probably asymmetric) possessed by borrowers and lenders and depositors and deposit-takers on intertemporal conditions and by contract enforcement possibilities in financial markets.

The literature on the economic theory of credit markets and saving decisions in economies characterized by incomplete markets and imperfect information is growing. Following the lead of Stiglitz and Weiss (1981), a number of theoretical papers have explored the implications of imperfect information and incomplete markets for contractual agreements in credit markets in low-income environments (see Alderman and Paxson 1992 for a useful bibliography). This body of work provides a new theoretical foundation for policy interventions to correct market failure. By studying the institutional arrangements through which financial transactions take place, this research has shown that credit transactions reflect the economic environment in which they occur. In Africa most of the economic environment is characterized by objective risk, with unpredictable variations in income as a result of weather and other exogenous processes. In the absence of complete insurance markets, credit transactions take on a special role in allowing resources to be transferred in response to income shocks. The costly acquisition and asymmetric distribution of information are essential aspects of this environment.

Trading Credit among the Poor in Incomplete Markets

The demand for credit by poor rural households is dictated by assumptions underlying the rural household model. Rural households derive income from farm activity, nonfarm production activity, and wage labor. If at any given time these activities provide inadequate funds for both consumption and investment, households may seek to borrow. Opportunities for supplementing income with credit are important in determining households' attitudes toward farm and nonfarm production activity and the provision of labor in the midst of failed markets for products and factors.

According to the basic agricultural household model, households attempt to maximize utility in terms of consumption and leisure by combining farm production, nonfarm production, and labor (Reardon, Crawford, and Kelly 1994). They also seek to stabilize income, manage risk, and achieve food security. The house-

hold model is influenced by the economic and institutional environment outside the household, including markets and government policies, which affect output, input, credit, land, and labor markets. A household's objective function is affected by choices of consumption of its own produce or purchased food, for example, as well as by the impact on feasibility and the relative profitability of nonfarm and farm activities.

The behavior of the peasant household in the absence of a credit market has been modeled by de Janvry, Fafchamps, and Sadoulet (1991). Their model of a household producing a cash crop and a food crop with two inputs (labor and all other inputs) can be modified to explain decisions on farm and nonfarm production. The household is assumed to try to maximize its utility from farm and nonfarm production subject to a cash constraint. If all markets (including credit markets) existed, all prices would be exogenous, and households would choose independently between farm and nonfarm activities. If no formal market for credit exists, however, the two sets of decisions become linked by an endogenous set of prices that responds to different changes in the external environment, including market prices and credit conditions. This set of endogenous prices determines the marginal utilities from farm and nonfarm production outcomes, since inputs are assumed to be fungible. Faced with unsatisfactory credit market conditions, a household could maximize its utility by switching inputs from nonfarm activity to farm activity in order to ensure that farm activity continues in anticipation of future market failures. This is a second-best option for the rural household.

Trading Insurance in Incomplete Markets

In addition to diversifying their sources of income, rural households can use partially functioning credit markets to provide insurance against income shocks. Besley (1994) notes that credit and insurance institutions exploit the gains from trade and suggests that "all of the ideas of competitive equilibrium can be extended to economies with uncertainty by imagining that a whole array of contingent commodities is available; these are commodities whose returns are indexed to the state of nature. The existence of markets for a complete set of contingent commodities preserves the link between perfect planning and the market allocation" (p. 36). Despite the absence of a complete set of contingent commodities, several rural institutions for risk sharing probably come close to yielding efficient outcomes. Thus two risk-averse individuals could "potentially strike a deal that dominates autarky" (Besley 1994, p. 36). Besley models efficient risk sharing in rural conditions, focusing on the pooling of resources, as suggested by the anthropological literature. In its simplest form insurance is provided by households' ability to gain access to the production of other households through some communal arrangement, thereby increasing their own utility when faced with an income shock.

The availability of insurance can also be constrained by information. Since insurance is simply a trading of risks, inadequate information about the nature of the risk faced by each individual (adverse selection) and possible changes in private behav-

ior after being insured (moral hazard) mean that most insurance arrangements are only partial.

Interaction between Lenders and Borrowers in Incomplete Rural Credit Markets

The framework provided by Stiglitz and Weiss (1991) on incomplete information equilibrium has been the basis of much empirical analysis of credit market interactions.[1] The role risk plays in the allocation of credit through its effect on transactions costs underlies a good deal of recent analytical work. Much of the risk faced by lenders is captured by principal-agent relationships that can be analyzed within game-theoretic frameworks. Lenders are concerned with being repaid (Aryeetey and Udry 1994); borrowers are concerned about their chances of receiving future loans and so about their individual reputations. These concerns condition the interactions that lead to repeated relationships in a game played under suboptimal conditions.

Increased risk resulting from inadequate information increases the transactions costs of lenders, forcing them to adopt measures to reduce risk. The result will be an incomplete information equilibrium that remains constrained Pareto efficient. Carter (1988) shows that the profitability of bank lending falls with increasing risk, as in the case of small farms that are assumed to be less productive than large farms. After modeling lender behavior where group characteristics of borrowers are fixed and there is no voluntary default, Carter shows that if there were no contractual restrictions on banks "lenders could always realize the same profits on loans to small farms as on loans to large farms by offering different loan terms (based on available information) to the different groups" (p. 90). Under such an arrangement small productive borrowers could face unfavorable loan terms because of the (incorrect) perception that they are only as productive as the average small borrower. If contractual restrictions (such as fixed collateral values and ceilings on interest rates) are placed on lenders, nonprice credit rationing would result in the face of excess demand. In either case the small farm borrower would be discriminated against in the equilibrium position. If borrowers were allowed to respond to contract terms freely, it is likely that the allocation of credit between consumption and production would change with the contract terms and that more creditworthy borrowers would opt out of the market earlier as contract terms worsened. In this situation self-imposed interest rate restrictions and unequal nonprice rationing might result, similar to what happens when the market is characterized by exogenously restricted equilibrium.

Using Groups to Tackle Default Risk in Incomplete Markets

The literature on default risk focuses on two ways of reducing that risk: repeat lending and group lending arrangements (Stiglitz 1989). Repeat lending provides households with an incentive to repay their loans in order to be able to borrow again. Social sanctions that punish defaulters also encourage repayment. This

mechanism is highlighted in a model of a long-term relationship between borrower and lender when imperfect enforcement mechanisms prevail (Aryeetey and Udry 1994).

In group arrangements the group may repay a member's loan in order to maintain future access to credit. This mechanism relies on the assumption that individuals are less likely to default if they know their actions will harm others in their group. The Grameen Bank is an oft-cited example of how group arrangements can limit default (Hossein 1988). Besley and Coate (1995) show that groups that connect people socially are more likely to be efficient than other groups. This could be a crucial factor in explaining the difficulties faced by many innovative schemes and microfinance projects that rely on the group concept.

Economic Activities of the Rural Poor and Their Borrowing Behavior

To what extent does the empirical evidence on the use of credit by rural households confirm the patterns predicted by the household model? The use of credit by rural households appears to be relatively low in Africa, and economic activity appears to be financed mainly by switching incomes among household activities. This is not surprising considering that nonfarm activities account for an important share of farm household income (Reardon, Crawford, and Kelly 1994). In Ghana, for example, Living Standards Measurement Survey (LSMS) data show that in 1992 nonfarm self-employment provided as much as 33 percent of rural household income (farm activity brought in 36 percent and wage labor 20 percent). A survey of microenterprises in rural Ghana found that more than two-thirds of the sample of 1,100 entrepreneurs also owned farms (Riedel and others 1988).

Demand for and Use of Credit

It is difficult to derive reliable estimates of rural demand for credit, which can be conceived in several ways. Potential borrowers who cite lack of credit as a constraint reveal perceived demand. Some portion of borrowers who express a desire for but do not seek credit because of market imperfections and institutional barriers represent potential demand. Lenders, of course, can respond only to revealed demand, that is, potential borrowers who apply for credit at prevailing interest rates. Directed agricultural credit schemes assume that both revealed and potential demand for credit far exceed supply. Many African governments and several credit institutions that rely on a supply-leading approach to finance operate on this assumption. Although the assumption has been criticized (Adams 1971, 1984), the perception that demand far outweighs supply persists. The income shocks that are common in rural Africa make it likely that demand will remain high.

It is not clear from the scanty data, however, that the demand for credit is as high as has generally been believed. According to the third LSMS for Ghana, only 33 percent of the sample classified as "rural poor" (that is, living below the poverty line) had any debt obligations throughout 1993, and only 8 percent had tried to borrow

and been refused credit. The LSMS for Tanzania reveals similar statistics. In both Ghana and Malawi surveys of potential borrowers found that up to 30 percent of the fifty small nonfarm enterprises sampled had never applied for a bank loan (Nissanke and Aryeetey 1995); an even greater share had never tried to obtain loans in Tanzania. That so many small potential borrowers had never actively sought formal credit was attributed to the perception that bank credit was not available to small enterprises. This perception suggests that the absence of supply creates its own lack of demand, expressed as low revealed demand.

LSMS household data for Ghana suggest that borrowing is used to smooth what households view as temporary shocks to consumption. The survey found that poor households borrowed mainly to purchase consumer goods (25.4 percent) and to expand (nonfarm) business (24.2 percent); loans for investment in agriculture (purchases of land and other agricultural inputs) accounted for only 11.2 percent of loans (table 1). That 68.6 percent of loans were made by relatives, friends, and neighbors is indicative of the practice of trading risks within communities (Udry

Table 1. Sources and Purposes of Loans in Ghana
(percent)

Characteristic	*Poor households*	*All households*
Source of loan		
Relatives, friends, neighbors	68.6	67.2
Traders	15.9	16.9
Other[a]	15.5	15.9
Purpose of loan		
Consumer goods	25.4	25.6
Business expansion	24.2	23.8
Housing, education, health	18.6	19.9
Ceremonies	11.6	11.4
Agriculture	11.2	10.5
Other	8.9	9.0
Type of collateral required		
Land, cattle, house	3.3	3.0
Other	4.1	4.2
None	92.6	92.9
Gender of household head who obtained loan		
Male	63.8	65.2
Female	36.2	34.8
Request for loan rejected	8.8	8.3
Reason for rejection		
Insufficient income	41.7	39.6
Insufficient collateral	16.7	15.8
Previous debt problems	6.8	8.8
Inappropriate purpose of loan	8.7	10.4
Other	26.1	25.4

a. Includes the formal financial sector.
Source: Computed from Ghana Living Standards Measurement Survey (GLSS III) data, Ghana Statistical Service.

1994). Only 15.9 percent of credit came from traders, who often lived in other communities. In general, the borrowing characteristics of poor households were not very different from those of other households.

Many surveys of rural nonfarm enterprise finance in Africa suggest that entrepreneurs turn to moneylenders only as a last resort (Nissanke and Aryeetey 1995; Bagachwa 1994; Soyibo 1996b). Although 73 percent of the rural microentrepreneurs studied by Aryeetey (1992) failed to secure bank loans, only 6 percent of them used credit from moneylenders and other informal sources to finance planned investments. Microentrepreneurs who did not use informal credit scaled down their planned investment, used personal savings (accumulated income) to finance part or all of the planned investment, or abandoned the planned investment.

The hypothesis that credit is used to smooth out temporary income shocks in order to finance household consumption is supported by the low revealed demand for credit. In rural Ghana the average loan represents only 1.5 percent of average income and 29 percent of average deposit balances (table 2). Thus households borrow far less than they are capable of mobilizing by making their assets liquid.

Credit As Insurance

Udry (1994) has tested the hypotheses that loans are state-contingent commodities and that the amount households borrow depends on the size of their income and how their income is likely to be affected by the current and expected future environment. His empirical model of northern Nigeria suggests that the size of a loan by a household is a function of individual wealth, age, and other household characteristics. Thus "households borrow more when they suffer an adverse shock, and they lend more when they are faced with a positive shock" (p. 39). Udry tests for full income pooling by determining whether the amount borrowed or lent depends only on aggregate village income or individual income and whether repayment by individuals is affected by the shocks that lenders and borrowers suffer. The hypothesis of full income pooling is rejected, but the notion that loan transactions in rural environments are important for risk sharing is supported.

The Ghanaian household survey suggests that households mainly want to lend to borrowers whose wealth can be used as security for repayment. Households with surplus incomes that seek to lend look for borrowers with nonfarm rather than

Table 2. Average Loan As a Share of Income and Savings in Rural Ghana, 1988

Category	*Amount (U.S. dollars)*	*Share (percent)*
Average loan	45	—
Average income	2,998	1.5
Total savings	1,597	2.8
Total financial savings	258	17.5
Average bank balance	156	28.8

Source: IPC 1988.

Table 3. Correlation between Amount of Loan Received and Selected Variables for the Poor in Ghana, 1991–92

Variable	*Pearson correlation coefficient*	*Probability*
Farm income	–0.04881	0.1074
Nonfarm income	0.09119	0.0026*
Total value of assets	0.18689	0.0001*
Value of house	0.69044	0.0001*
Value of land	–0.05923	0.4219

* Significant at the 0.05 percent level.
Source: Computed from Ghana Living Standard Measurement Survey (GLSS III) data, Ghana Statistical Service.

farm incomes. Correlations are positive and significant between the amount of the loan received and nonfarm income, total value of assets, and value of the house (table 3). Correlations are negative but not significant for farm income and the value of land.

Switching of Incomes between Farm and Nonfarm Household Activities

Rural entrepreneurs often suggest in interviews that they use personal savings to finance working capital. Probing often reveals that they seldom accumulate savings from their incomes over an extended period to use for these expenditures, however (Aryeetey 1992). In reality, "personal savings" often refers to additional income from a secondary activity—trading, farming, commercial transport, real estate, or rent seeking—that is directly expended on whichever activity is earning profits at the time.

In rural West Africa switching profits from one activity to another is a more common way of financing working capital requirements than is borrowing (Aryeetey 1993). Diagana and Kelly (1994) have shown that nonfarm income has been increasingly used to finance farm input purchases in Senegal since the termination of government-supported credit schemes and groundnut seed distribution programs. Hoffman and Heidhues (1993) report that among landless farmers in Benin only those with nonfarm incomes could obtain credit for farm inputs. Collier and Lal (1986) made similar observations for Kenya.

Switching between activities takes place because credit market failures make the cost of a transaction greater than the utility gain that it produces. Under these conditions "either a surrogate institution will emerge to allow the transaction to take place or the transaction simply does not occur" (de Janvry, Fafchamps, and Sadoulet 1991, p. 1401). In most cases the transaction does not occur. Although the income switch is initially regarded as a temporary measure, over time it becomes a structural feature of the farm household model. If the marginal product of the subsidized activity is less than the marginal product of the other activity, the use of income from one economic activity to subsidize another moves the rural household to a lower net iso-income or utility curve.

Credit Market Characteristics in Africa

Why is formal credit not more widely available in Africa? Financial systems in Africa suffer from "weak resource mobilization, high credit losses, high intermediation costs, and excessive political interference" (World Bank 1994, p. 100). The institutions responsible for ensuring that financial transactions take place to support investments are ineffective, unable to generate significant savings, or to satisfy what is perceived to be significant demand for credit in many rural areas.

If the formal sector (used here to mean mainly banks) is unable to lend to small rural borrowers, it is not because of an inadequate lending base (Nissanke and Aryeetey 1995), but because of problems associated with the maturity transformation of bank liabilities; difficulties administering loans (particularly the screening and monitoring of small borrowers); relatively high transactions costs, partly as a result of loan administration procedures and relatively high risks of default; and crowding out by the public sector.

By contrast, in the informal financial sector, which includes savings and credit associations, savings collectors, cooperatives, and moneylenders, the size of the lending base controlled by each unit is critical. Informal financial institutions often depend on deposits mobilized from a specific group, as in the case of a savings and credit association, or on the lender's surplus funds (from profits from the lending business or other businesses).

In a study of policy, structural, and institutional constraints to financial integration in Sub-Saharan African financial markets, Nissanke and Aryeetey (1995) found that reform efforts focused on improving the targeting and operation of formal sector institutions, including improvements to supervision and internal regulations, rather than on relationships between different institutions. They found severe agency problems, significant information asymmetry, and poor legal contract enforcement mechanisms in the financial markets they studied. The study also identified the following characteristics:

- Formal and informal market segments usually lend to distinct types of borrowers.
- Relatively little interaction takes place between the formal and informal market segments.
- Within market segments contractual terms do not differ significantly.
- Few spillovers from formal to informal market segments take place.
- Informal lenders appear to be more concerned about adverse selection than moral hazard.
- Mechanisms for developing substitutes for collateral and enforcing sanctions in localized contracts make contract enforcement easier for informal than for formal lenders.
- Repeat lending is more common among formal than informal lenders.

These characteristics suggest the need for institutions to oversee the efficient control and regulation of the credit market. They also suggest that agency relationships need to be improved, with acceptable contract enforcement mechanisms for all parties.

Nature of Credit Markets: Who Can Borrow and Where?

Rural credit markets in Africa are fragmented because the formal and informal markets serve groups of clients with distinct characteristics and needs, and there is almost no interaction between different types of institutions. Any benefits of specialization stemming from segmentation are outweighed by the negative effects of the weak linkages between the two segments. Almost no funds flow between the two segments, and wide differences in interest rates prevail. With little possibility of substituting funds among different lenders, the fragmented markets have difficulty intermediating between savers and investors. Not being able to allocate financial resources freely, financial institutions cannot always transform and distribute risks and maturities efficiently. As a result the two segments function differently and serve distinct socioeconomic groups.

In rural areas banks are interested in lending mainly to large plantations and processing plants. In the few instances where banks lend to small farms, the loan is usually made through a group arrangement and is often guaranteed by the government or by donors. In Malawi (Chipeta and Mkandawire 1996b) and Tanzania (Bagachwa 1996), where some of the most significant agricultural lending takes place, most loans are made to large cash crop exporters (Malawi) and export parastatals (Tanzania).

Demand for loans from moneylenders, who generally charge the highest rates for credit, comes mainly from potential borrowers with no other options. Moneylenders are the only source of informal credit that does not require borrowers to satisfy specific membership obligations. The short maturity periods and high interest rates charged by moneylenders make this source of credit unattractive for potential borrowers seeking working capital and loans for fixed investments.[2] In rural Nigeria farmers, market women, other traders, nonfarm entrepreneurs, and other self-employed craftspeople borrow from moneylenders. Farmers sometimes borrow from moneylenders during the planting season to maintain their households until the next harvest; they may also borrow to cover the cost of funerals and other social events (Adegboye 1969). Rural Nigerian households usually borrow from their own communities (Udry 1990), but they sometimes travel far to borrow from a moneylender.

Credit from savings and credit associations and cooperatives is used mainly for consumption and sometimes for working capital (Besley and Loury 1992).[3] Small short-term loans are suitable only for activities with quick turnover, such as cereal production. Group membership is an essential aid in screening loan applicants and ensuring that contracts are enforced. Because of expectations of increased risk and increased probability of losing homogeneity as a group expands, many associations have limitations on growth. Homogeneity provides members with a sense of familiarity that engenders mutual trust. Some rotating savings and credit associations have suggested that they were not keen on mixing people whose backgrounds and interests were very different (Bortei-Doku and Aryeetey 1995). This restriction on the size of the group limits the use of these sources of credit to borrowers whose demand for loans is irregular. Savings (*susu*) collectors grant "advances" to trusted

clients and traders at local markets in need of short-term credit. When collectors lend to nondeposit clients, they tend to behave like moneylenders.

Direct and Indirect Linkages between Segments

Direct, or institutional, linkages are manifested by flows of funds between segments. Indirect, or market, linkages are present when activities in one segment are affected by activities in other segments through the influence of the financial market. Both types of linkages are weak in most of the African countries studied (Nissanke and Aryeetey 1995).

There is little evidence of informal lenders obtaining bank loans for their lending businesses. In Tanzania only seven of twenty-one trader-lenders studied had ever applied for bank credit (Bagachwa 1994). In Malawi only 23 percent of informal lenders had applied for and received bank loans, and 85 percent of those loans were made for uses other than the borrowers' moneylending activities (Chipeta and Mkandawire 1996a). The pattern is not much different in Ghana (Aryeetey 1996) or Nigeria (Soyibo 1996b), where only 20 percent of informal lenders applied for and received bank credit. Bank loans to informal lenders specifically for on-lending were made mainly to cooperatives in farming communities.

Indirect interaction between formal and informal lenders is even less common. Bell's (1990) model of competition between institutional and informal lenders allows for such price interaction to occur. However, his model is based on the assumption that "low-cost institutional credit" is available to informal lenders, an assumption that seldom holds for African financial markets. Because of the deep segmentation of the markets, informal lenders rarely respond to changes in formal sector loan rates by altering their own rates (Nissanke and Aryeetey 1995). Formal interest rates have risen substantially since financial sector reforms began in many countries, while informal rates have remained virtually constant, indicating the fragmentation of the market.

The absence of spillovers is characteristic of segmented markets. In Malawi only 35 percent of small businesses that were unsuccessful in receiving bank loans sought informal sources of credit (Chipeta and Mkandawire 1996b). In Ghana, where less than half of a sample of fifty-five mainly rural nonfarm enterprises were successful in obtaining bank credit, more than half of those who failed to secure loans did not attempt to borrow elsewhere (Aryeetey 1996). Spillovers do not occur because, following Bell's (1990) analysis, even though the borrower has an unsatisfied demand for finance, none of the offers made by the informal sector would provide an income benefit greater than that available without borrowing.

Dealing with Incomplete Markets: Loan Screening, Monitoring, and Contract Enforcement

The theoretical developments discussed earlier suggest the importance of penalties in credit markets in developing countries. In the absence of sound formal contract

enforcement mechanisms, both formal and informal lending institutions rely on loan administration practices that emphasize screening loan applicants rather than monitoring the use of loans. This emphasis would suggest that adverse selection rather than moral hazard is the primary problem lenders confront (Nissanke and Aryeetey 1995).

SCREENING METHODS. In screening prospective borrowers, formal lenders tend to focus on project viability, whereas informal lenders rely on the character and family history of the borrower. The difficulty that banks have in obtaining information about borrowers explains why they rely more on repeat borrowing than informal lenders do. In Nigeria most loans to small rural enterprises are made to repeat borrowers (Soyibo 1996a).

Screening in the informal sector relies extensively on personal knowledge of borrowers (Udry 1990). The development of personal ties and the use of borrower proximity in decisionmaking are mechanisms for countering adverse selection and moral hazard (Yotopoulos and Floro 1991). The more rural is the environment, the greater is the need to personalize ties in order to minimize information asymmetry. Familiarity with the borrower often reduces the need to rely on repeat borrowing. In northern Nigeria agricultural lending among relatives, acquaintances, and neighbors is the norm (Udry 1990). Susu collectors in Ghana can make a loan decision in an instant simply by looking at the card on which deposits are entered to confirm that the borrower is a regular depositor. A moneylender's new clients are often introduced by people the lender knows well and who are prepared to give their word as a guarantee of payment.

In savings and credit associations and cooperatives loan screening is done at the time a member is admitted. The emphasis is not necessarily on whether members can pay back loans but rather on their commitment to the group's goals. Since applicants invariably have similar incomes and credit requirements, assessment of the individual's character and reliability is most important.

LOAN MONITORING. It is often suggested that informal lenders have better repayment records than do formal lenders mainly because they constantly monitor the uses to which their loans are put (Yotopoulos and Floro 1991). Such monitoring is often assumed to take the form of regular visits to project sites. Nissanke and Aryeetey (1995), however, found that relatively little monitoring of this type is done by either formal or informal lenders after loans have been given out. According to their study, half the credit unions in Ghana never monitor any loans and 22 percent do so only some of the time; moneylenders and other informal groups seldom visit the project sites of their borrowers. When lending is localized, the need for project visits is reduced. For moneylenders, who are more likely to have clients in other localities, the need for monitoring is reduced because they know the person who introduced the borrower to them.

Many banks fail to monitor projects because they lack the equipment needed for regular project visits. In Ghana most project officers in state-owned banks know that

they are required by internal regulations to visit projects regularly, but they are constrained from doing so by lack of transport (Aryeetey 1996).

LOAN REPAYMENT AND CONTRACT ENFORCEMENT. Nissanke and Aryeetey (1995) found loan repayment rates to be much higher for informal lenders than for formal lenders. These higher repayment rates are not necessarily the result of more aggressive contract enforcement, as Shipton (1991) has suggested. In contrast to Shipton's findings in the Gambia, Nissanke and Aryeetey (1995) found little evidence of litigation by either formal or informal lenders. They observed that despite the relatively large incidence of defaults, none of the sample bank branches in Ghana foreclosed on any loan in 1992. In Tanzania, where default rates were highest and 94 percent of bank loan portfolios were rated as substandard, doubtful, or lost as of the end of 1991 (Bagachwa 1995), few legal cases were brought because the largest defaulters were public entities. In the absence of proper ownership documentation and a well-functioning legal system, confiscation of collateral by formal lenders is almost impossible. To discourage default informal lenders go to the homes of their clients to deliver verbal warnings. In membership arrangements, lenders use the threat of dismissal from the group to encourage repayment.

The stronger portfolios of informal lenders likely reflect their more efficient procedures for calling in loans and the borrowers' knowledge that the informal lender is more likely to act on threats to foreclose on collateral since the cost of foreclosing is likely to be higher for a bank than for a rural moneylender (who can always find a relative to farm foreclosed land until a loan is repaid). Collateral thus has different meanings in the formal and informal sectors, which affects borrowers' attitudes toward repayment. The game-theoretic framework is useful for predicting how a borrower responds to the two sectors.

Transactions Costs

Informal lenders in Africa tend to have much lower transactions costs (administrative costs plus default risk costs) than do formal lenders (Nissanke and Aryeetey 1995). For the formal sector costs vary by type of bank. In Ghana, for example, transactions costs for a small-scale agricultural loan are about 6 percent of the loan amount for a commercial bank and about 10 percent for a development bank. In the informal sector transactions costs are generally much lower (table 4). Informal lenders are able to keep costs low because of the techniques they use to screen prospective borrowers. The fragmented units of the informal sector will continue to lend to their target groups as long as they cannot change client groups without increasing the level of risk (thereby increasing transactions costs).

Credit Gaps in Rural Africa

Because of fragmentation few lending units in Africa meet the needs of small borrowers, many of whom are interested in borrowing less than $1,000 for up to eigh-

Table 4. Mean Transactions Costs of Informal Sector Loans, 1992
(percentage of loan amount)

	Moneylender		*Susu collector*		*Cooperative*		*Credit union*	
Country	*Urban*	*Rural*	*Urban*	*Rural*	*Urban*	*Rural*	*Urban*	*Rural*
Ghana	1.8	2.7	0.9	0.6	0.3	0.3	2.6	4.4
Malawi	0.6	0.6	—	—	0.2	0.2	0.4	0.1
Nigeria	3.2	2.7	0.6	0.6	1.0	0.6	1.9	0.6
Tanzania	1.7	2.6	—	—	0.1	0.1	2.5	3.0

Source: Nissanke and Aryeetey 1995.

teen months at interest rates far below 30 percent a year (Nissanke and Aryeetey 1995). The absence of this kind of formal credit has hurt rural borrowers, especially microenterprises in small towns, many of which have been unable to expand.

The credit gap refers to the unmet credit needs of potential borrowers who are unwilling to borrow at the terms offered by informal lenders and who cannot gain access to formal loans because of a failure to meet eligibility criteria. This credit gap can be filled only by reform of institutional structures.

Recent Institutional Developments in Rural Financial Markets

To counter the effects of credit market failures that result in fragmentation and the exclusion of many potential borrowers, a variety of credit schemes have been introduced in many African countries. In their study of microfinance institutions in West Africa, Webster and Fidler (1996) found such programs to be common in some countries (Burkina Faso, the Gambia, Guinea, Guinea-Bissau, and Mali) and rare in others (Chad, Mauritania, São Tome, and Sierra Leone). A SODECON (1990) study in Kenya, where K-REP, probably the best-known microfinance scheme in Africa, operates, identified as many as forty organizations involved in microfinance projects. Evaluations of innovative credit retailing and other microfinance projects in Africa suggest that they have been less successful there than in other developing regions (Christen, Rhyne, and Vogel 1994).

Objectives and Strategies of Innovative Credit Schemes

Innovative credit retailing schemes are usually community-managed credit and savings schemes that "are established to improve members' access to financial services, build a community self-help group, and help members accumulate savings" (Holt 1991, p. 4). Microfinance programs are more likely to emerge from donor projects and are not necessarily community based. For many innovative schemes credit provision is not the only objective—even for programs for which credit provision is the ultimate objective. Schemes that can be characterized as taking a minimalist approach concentrate only on credit, and all of the project's activities, including training of staff or beneficiaries, are designed to facilitate lending. Schemes following an integrated approach regard training and other forms of

technical assistance to borrowers as integral components of a broader scheme of assistance.

Most of the innovative schemes in Africa have been based on the minimalist approach, although in recent years some schemes have begun to emphasize market principles (Jackelen and Rhyne 1991). Webster and Fidler (1996) observe that many African microfinance arrangements have benefited from experience and best practices in other developing regions. These include issuing short-term loans, starting with small initial loans, focusing on small working capital loans to firms with proven track records, offering specialized services but without targeting, simplifying and localizing services, shortening turnaround time for loan applications, relying on group solidarity or joint liability to motivate repayment, mobilizing savings, and charging full-cost interest rates.

Village banks, for example, tend to extend credit to finance income-generating activities and savings. By establishing joint liability group lending mechanisms, members are expected to overcome collateral requirements. Village banks lend on an unsecured basis, using five-person group guarantees that make each individual responsible for the others, with future access to credit dependent on all members repaying their loans—a principle borrowed from the Grameen Bank. Other microfinance projects in Africa provide credit to individuals and projects. K-REP supports both group borrowing (through Watanos) and individual arrangements with nongovernmental organizations (NGOs) that on-lend. The schemes studied by Webster and Fidler (1996) also include both group and individual arrangements.

Terms of loans by village banks are similar to those of informal lenders. To discourage the rich from seeking village bank credit, loans are small, ranging from $10 to $160 and averaging $60. K-REP has an average loan size of $100 under its Juhudi scheme. Loan maturities are generally short. K-REP expects groups that borrow from it for on-lending to repay their loans within a year. Members of small groups are expected to repay their initial loans within four weeks, with the assurance that they can take a further loan from the revolving fund established by the group. Although interest rates are higher than most formal lending rates in Africa, they tend to be lower than rates charged by moneylenders. The nominal interest rate for K-REP loans in 1994 was 35 percent; the real rate was 6.2 percent. Webster and Fidler (1996) found that effective interest rates on the loans they studied were between 19 percent and 26 percent for five schemes, between 32 percent and 39 percent for three schemes, and 54 percent for one scheme. These rates were comparable to those of some of the best-known schemes in Asia and Latin America (Christen, Rhyne, and Vogel 1994).

Assessment of the Performance of Microfinance Institutions

Holt's (1991) assessment of the achievements of lending by village banks focused on repayment rates, loan sizes, savings levels, program costs, and income from interest. She found that repayment rates in village banking programs were high, averaging 90 percent in the banks studied. Projects that had high repayment rates shared the following characteristics:

- Training programs were provided for project participants.
- Interest rates were not subsidized.
- Formal written membership requirements and screening measures were written into by-laws to ensure discipline among members.
- A savings program accompanied the loan.
- The sociocultural environment strengthened social sanctions, helping to reduce default rates.

The evaluation of credit schemes examined here draws on the framework developed by Yaron (1992), who proposed that innovative schemes' success in moving toward self-sustainability and the extent of their outreach activities be used as measures of success. "Financial self-sustainability is achieved when the return on equity, net of any subsidy received, equals or exceeds the opportunity cost of funds" (Yaron 1994, p. 50). Outreach could be measured in terms of the type of clientele served and the variety of financial services offered, including the value and number of loans extended, the value and number of savings accounts opened, the type of financial services offered, the number of branches and village subbranches, the percentage of the rural population served, the real annual growth of the institution's assets in recent years, and the participation of women as clients.

The achievements of these lending programs need to be assessed relative both to similar programs across regions and to other institutional arrangements within the region. In effect, what needs to be assessed is the compatibility of the innovative schemes with the practices and attitudes in African countries as reflected in informal systems. Various evaluations suggest that while innovative financing and other microfinance projects are performing well in making credit available, local environments often constrain their ability to reduce costs below current levels.

SUSTAINABILITY. A subsidy dependence index applied to the K-REP program revealed a high level of dependence on subsidies (Kiiru, Pederson, and Nzioka 1995). Such an index could be used to measure the percentage increase in the average on-lending interest rate required to compensate the lender for eliminating subsidies. Webster and Fidler (1996) used two sets of indicators to assess operational efficiency, one measuring overall institutional efficiency and the other measuring the productivity of staff members. Ratios of operating costs (mainly expenses associated with head offices, staff salaries, and expenses associated with expatriate staff) to average loan portfolio were found to be very high in most of the nine schemes studied in detail. According to Webster and Fidler, operating costs were high because of the "high-cost nature of the African environment" (p. 31) and the fact that volunteers and in-kind donations, such as business premises, were not used. Staff salaries were high in several cases, weakening the sustainability of the programs. Expenses for expatriates represented 145 percent of the average outstanding loan portfolio in the Credit Mutuel scheme in Senegal, 73 percent in Credit Mutuel in Guinea, and 57 percent in Credit Rural in Guinea.

To assess financial viability, Webster and Fidler looked at the ratio of total revenue to total expenses, where revenue comprised interest income, fees, and interest

on investments, and expenses comprised operating expenses (all administrative costs, depreciation of fixed assets, and losses from loan defaults) and the cost of loan funds. All but two of the institutions in their sample failed to cover more than half of their operating costs. The high level of dependence on subsidies suggests that interest rates need to rise. Although it is assumed that the poor can pay higher interest rates than they are currently paying (Webster and Fidler 1996), the risk structure of the scheme could be significantly affected.

Webster and Fidler (1996) measured institutional durability by examining the balance between social and financial objectives, issues of governance (including decentralization and participation), staff incentive and training systems, and factors in the environment that increase the institution's chances of survival in the future. They found that in all cases lending decisions are handled locally. Jackelen and Rhyne (1991) suggest that "extremely rapid and decentralized approvals . . . [have] been found to be one of the key elements of success of these programs as borrowers find it in their self-interest to repay loans in order to obtain additional credit" (p. 12). Rapid loan processing is also characteristic of all informal sector loans. Studies of West Africa reveal that microfinance programs have invested in staff training, which is expected to improve their sustainability (Webster and Fidler 1996).

OUTREACH. The number of borrowers reached by microfinance programs in Africa is far lower than those reached by successful institutions in Asia and Latin America (Christen, Rhyne, and Vogel 1994; Webster and Fidler 1996); the largest microfinance institutions in Africa reach only as many people as some of the smaller institutions in Asia and Latin America. K-REP had only 6,943 borrowers in 1994, up from 1,245 in 1991 (Kiiru, Pederson, and Nzioka 1995). The West African microfinance programs studied by Webster and Fidler were only marginally larger, with most having fewer than 10,000 borrowers. They noted that even though many of the institutions had managed to develop large rolls of members, they had not converted them into large numbers of borrowers and savers.

Webster and Fidler (1996) attribute the small scale of these institutions in part to the sparse population in many of the rural areas served in the Sahel. In some areas significant portions of the adult population are members of the programs. In Mali, for example, in villages in which Caisse Villageoises du Dogon maintained field offices, about 44 percent of the adult population were members. In such circumstances "capturing large shares of the market will not result in significant scale of operations, especially where populations are small" (p. 43). Webster and Fidler noted other institutional constraints to expansion of microfinance programs in West Africa, including designs for regional operations that cannot be expanded to cover other regions.

Average loan sizes indicate that African microfinance programs are smaller than those in other regions, suggesting that they are more likely to reach the target group of the poorest people. Further evidence that the programs are intended primarily for the poor is the high level of illiteracy among borrowers and the large number of women that borrow through these programs.[4]

Group Arrangements and Innovative Lending Schemes

Innovative lending schemes often extend the concept of the group to groups whose commitment to savings and credit may be weak. The level of commitment among a group of people joining a donor-sponsored credit arrangement is likely to be weaker than it is in a revolving savings and credit association made up of people who came together on their own (Besley, Coate, and Loury 1993). There are other differences as well. So long as members continue to be interested, a revolving savings association can become a permanent feature of the community, whereas an external innovative project is viewed as a temporary arrangement that will be available only as long as a donor supports the project. And saving with and borrowing from a revolving savings association that belongs to its members is different from borrowing from an external institution. Default within a revolving savings association deprives other members of their savings and means of livelihood, whereas default within an innovative scheme merely denies participants access to "what did not belong to them anyway."[5]

Consequently, while an innovative scheme is likely to have a large enough number of potential and willing participants, attempts to expand rapidly are likely to fail because of the large number of uncommitted group members. A likely result of rapid expansion will be high default rates as the forces of peer pressure and social stigma are weakened. In rural Africa borrowing and debt are considered private matters (Adegboye 1969; Chipeta and Mkandawire 1991). Borrowers are under social pressure to repay their loans so that friends and relatives are not made aware of their borrowing. When credit becomes an open and communal affair, the shame of being found out is no longer a burden for the borrower. Similarly, peer pressure is more intense on a participant in a revolving savings and credit association than in an externally financed project. To avoid these problems projects often limit the number of groups that participate (Bortei-Doku 1988).

The basis on which group lending is advocated and practiced by innovative schemes and banks in Africa needs further review. Such a review must consider how communities view financial arrangements within groups.

Improving Access to Rural Finance in Africa

How should credit be made available to rural groups? Using the policy framework of Rashid and Townsend (1993), I briefly examine the merits of targeting credit to distinct rural groups.

Should Credit Be Targeted?

Few empirical tests of the general equilibrium framework for complete markets in Africa have been performed. Udry (1990) concluded that perfectly competitive markets do not exist in northern Nigeria; Deaton (1992) reached similar conclusions about Côte d'Ivoire. Africa has almost no full risk-sharing institutions. Nissanke and

Aryeetey (1995) show that in several countries some individuals, households, and regions remain isolated from markets and from mechanisms for borrowing and lending or insuring against risk, suggesting that considerable scope exists for intervention in rural African financial markets.

Informal lenders throughout Africa already appear to target the poor (including women), albeit not always successfully; attempts by better-financed innovative schemes to target the same poor people have been no more successful. The issue therefore is no longer simply one of targeting or not targeting, but of how to equip institutions that can reach the poor at least cost (that is, informal lenders) to extend their reach.

How Should Targeted Credit Be Delivered?

Which credit and insurance delivery mechanisms might allow financial institutions to achieve a second-best solution? Which financial institutions are most accessible to the rural poor, and how do they affect rural production? What credit conditions are most attractive to rural borrowers, and what institutions are best suited to providing them?

Most of the approaches suggested by Rashid and Townsend (1993), including the acquisition of more information by banks and the strengthening of group schemes, have been tried unsuccessfully by both formal and innovative schemes.

I believe that there is limited scope for enhancing the allocation of credit equitably and efficiently without establishing a closer relationship between the formal and informal sectors. Strengthening ties between the two groups would enable banks to acquire more information on would-be borrowers and innovative schemes to enforce contracts with larger numbers of borrowers as outreach and sustainablility are enhanced. I agree, however, with Besley (1994) that strengthened ties between the two sectors will not be a panacea for financial development. The answer lies in using known and broadly accepted indigenous institutions for retailing credit.

Aryeetey and Steel (1995) provide a framework for linking informal savings collectors in Ghana to the largest commercial bank (the collectors form a registered association with which the bank deals) that could lead to an increase in the proportion of susu depositors that gain access to credit facilities from their susu collectors. Aryeetey and Steel predict that the ratio of loans to deposits could rise from 9 percent to 60 percent if the scheme were implemented. An overdraft facility from the bank would cover about 50 percent of the shortfall in anticipated lending if susu collectors were to increase the average number of depositors from 420 to 600, the number that they believe they could conveniently handle. The scheme was based on the observation that increased lending by collectors often led to larger numbers of depositors.

Government Policies for Enhancing Linkages

If banks have not linked up with informal finance it is because of distrust, inadequate knowledge about informal agents, and prejudice, all of which create a risky

environment for banks (Nissanke and Aryeetey 1995). Policy should be designed to overcome these obstacles. A financial systems development approach should be adopted that builds institutions that serve identified segments. Policy can be used to enhance the development of linkages between the various segments, including the informal sector and such semiformal lenders as NGOs, by using the fiscal system and the regulatory and supervisory systems to provide incentives for formal institutions to wholesale credit through informal agents.

Tax relief could be offered to banks that allocate credit through informal and semiformal agents, with the lost revenues recovered by imposing higher taxes on banks that do not channel credit through the informal sector. Some banks, such as merchant banks, will have no need to use informal agents to allocate credit and can therefore serve as the actual financiers of the subsidy.

Regulatory and supervisory systems could be of considerable importance in providing incentives to banks. If banks perceived that risk could be reduced by dealing with credible semiformal and informal agents, they would be motivated to use them. Effective regulation and supervision of informal institutions would be problematic in some cases, however. Governments would have to follow what Popiel (1994) calls a "proactive approach," consisting of "providing a legal, regulatory, and prudential framework that fosters, and when possible, accelerates financial market development. This framework supports the setting up of mechanisms, institutions and instruments that promote and facilitate this development as the economy grows and market functions expand" (p. 92). Regulation should steer away from restrictive laws and focus on removing obstacles to financial market development. Restrictions on what assets banks may hold could be modified to encourage banks to invest in semiformal financial institutions. As Popiel suggests, "commercial bills and bankers' acceptances based on cooperative or 'mutualistic' guarantees should be developed to establish a link between semiformal and formal financial institutions" (p. 93).

All these considerations could be met through a three-tier approach to bank lending to marginal borrowers that provides an appropriate framework for regulation and supervision. Banks would lend to credible semiformal agents, such as savings and loan companies, finance houses, and credit unions, that would then link up with informal lenders such as susu collectors, cooperatives, and savings and credit associations. Rural borrowers would receive loans directly from informal agents. Where banks have good relations with easily identifiable informal agents, there is no reason the chain should not be shortened by allowing banks to deal directly with the informal institution. The semiformal institutions would in general be the agencies for regulating the smaller informal units since they can usually identify the operators more easily than external bodies can.

Is There a Role for Donors?

Donors can be expected to continue to support international NGOs in the allocation of credit, particularly in countries where nonbank financial institutions are not very well developed. Such NGOs should be used only temporarily, however, while

semiformal institutions and nonbank financial institutions are developed. Donor-supported NGOs need to be encouraged to work through indigenous informal credit institutions as much as possible, playing the role outlined above for semiformal institutions. Donors are in a good position to support the development of strong nonbank financial institutions with financial and technical assistance. Their support would be most efficiently and effectively allocated through private nonbank financial institutions with strong links to the informal financial sector in rural areas.

Who Should Be Targeted?

Rashid and Townsend (1993) suggest compelling reasons for targeting the poor and women. Given that access to various productive assets is directly related to social structures, which vary extensively in Africa, targeting these groups in specific regions of counties would appear to make sense. Even in areas where informal lending units operate, women may be excluded because of prevailing social norms.

Credit should also be targeted to men and women whose borrowing needs are too large for the informal sector but whose creditworthiness makes them ineligible for bank loans. In rural areas this credit gap encompasses farmers who wish to diversify more extensively, whether in crop production or nonfarm activity. In small towns the credit gap encompasses the microentrepreneurs who provide services to farming communities. These potential borrowers could be reached directly by innovative schemes operating out of small towns in various districts.

Notes

1. Based on empirical evidence in Africa, Aryeetey and Udry (1994) developed a modified version of the Stiglitz-Weiss (1981) model of credit allocation when adverse selection is a problem. They show that credit rationing does not need to occur under competitive equilibrium with adverse selection.
2. Loans by moneylenders have average maturities of three months, and few loans are extended for periods longer than six months. The practice of rolling over short-term loans is widespread in Africa, however.
3. In Malawi cooperative loans are most often used for farm working capital (fertilizer, laborers; Chipeta and Mkandawire 1991).
4. In 1994, 1,699 of 2,677 borrowers in K-REP's Juhudi scheme, which strives to meet the demand for rural and urban microenterprise activities, were women (Kiiru, Pederson, and Nzioka 1995).
5. A game-theoretic framework can be used to predict the outcomes for borrowers confronted with the two situations.

References

Adams, D. W. 1971. "Agricultural Credit in Latin America: A Critical Review of External Funding Policy." *American Journal of Agricultural Economics* 53 (2): 163–72.

———. 1984. "Are the Arguments for Cheap Agricultural Credit Sound?" In D. W. Adams, D. H. Graham, and J. D. von Pischke, eds., *Undermining Rural Development with Cheap Credit.* Boulder, Colo.: Westview Press.

Adegboye, R. O. 1969. "Procuring Loans by Pledging Cocoa Trees." Reprinted in J. D. von Pischke, D. W. Adams, and G. Donald, eds. 1983. *Rural Financial Markets in Developing Countries.* Baltimore, Md.: The Johns Hopkins University Press.

Alderman, Harold, and Christine Paxson. 1992. "Do the Poor Insure? A Synthesis of the Literature on Risk Sharing Institutions in Developing Countries." World Bank, Policy Research Department, Poverty and Human Resources Division. Washington, D.C.

Aryeetey, Ernest. 1992. "The Relationship between Formal and Informal Segments of the Financial Sector in Ghana." African Economic Research Consortium Research Paper 10. Nairobi.

———. 1993. "How Important Are Formal and Informal External Finance to Micro-Enterprise Development in Ghana?" In C. Cuevas and M. Benoit-Cattin, eds., *Finance and Rural Development in West Africa*. Montpelier, France: International Cooperation Center of Agronomic Research for Development (CIRAD).

———. 1994. "Financial Integration and Development in Sub-Saharan Africa: A Study of Informal Finance in Ghana." Overseas Development Institute Working Paper 78. London.

———. 1996. "The Formal Financial Sector in Ghana after the Reforms." Overseas Development Institute Working Paper 86. London.

Aryeetey, Ernest, and W. F. Steel. 1995 "Savings Collectors and Financial Intermediation in Ghana." *Savings and Development* 19 (2): 191–211.

Aryeetey, Ernest, and Christopher Udry. 1994. "The Characteristics of Informal Financial Markets in Sub-Saharan Africa." Paper presented at the plenary session of the Research Workshop of the African Economic Research Consortium, Nairobi.

Bagachwa, M. S. D. 1994. "Financial Integration and Development in Sub-Saharan Africa: A Study of Informal Finance in Tanzania." Overseas Development Institute Working Paper 79. London.

———. 1996. "Financial Integration and Development in Sub-Saharan Africa: A Study of Formal Finance in Tanzania." Overseas Development Institute Working Paper 87. London.

Bell, Clive. 1990. "Interactions between Institutional and Informal Credit Agencies in Rural India." *The World Bank Economic Review* 4 (3): 295–327.

Besley, Timothy. 1994. "Savings, Credit, and Insurance." In H. Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics* 3. Amsterdam: North-Holland.

Besley, Timothy, and Stephen Coate. 1995. "Group Lending, Repayment Incentives, and Social Collateral." *Journal of Development Economics* 46 (February): 1–18.

Besley, Timothy, Stephen Coate, and Glenn Loury. 1993. "The Economics of Rotating Savings and Credit Associations." *American Economic Review* 83 (4): 792–810.

Bortei-Doku, E. 1988. "Mid-Term Evaluation of the People's Participation Project." Food and Agriculture Organization Africa Regional Office, Accra.

Bortei-Doku, E., and Ernest Aryeetey. 1995. "Mobilizing Cash for Business: Women in Rotating Susu Clubs in Ghana." In S. Ardener and S. Burman, eds., *Money-Go-Rounds: The Importance of Rotating Savings and Credit Associations for Women*. Oxford: Berg Publishers.

Caprio, Gerard Jr., Izak Atiyas, and James Hanson, eds. 1994. *Financial Reform: Theory and Experience*. Cambridge: Cambridge University Press.

Carter, Michael. 1988. "Equilibrium Credit Rationing of Small Farm Agriculture." *Journal of Development Economics* 28 (February): 83–103.

Chipeta, C., and M. L. C. Mkandawire. 1991. "The Informal Financial Sector in Malawi: Scope, Size and Role." African Economic Research Consortium Research Paper 4. Nairobi.

———. 1996a. "Financial Integration and Development in Sub-Saharan Africa: Case Study of Informal Financial Sector in Malawi." Overseas Development Institute Working Paper 85. London.

———. 1996b. "Financial Integration and Development in Sub-Saharan Africa: The Formal and Semi-Formal Financial Sectors in Malawi." Overseas Development Institute Working Paper 89. London.

Christen R. P., E. Rhyne, and R. C. Vogel. 1994. "Maximizing the Outreach of Microenterprise Finance: The Emerging Lessons of Successful Programs." Consulting Assistance for Economic Reform Paper. IMCC, Washington, D.C.

Collier, Paul, and Deepak Lal. 1986. *Labour and Poverty in Kenya 1900–1980*. Oxford: Clarendon Press.

Deaton, Angus S. 1992. "Saving and Income Smoothing in Côte d'Ivoire." *Journal of African Economies* 1 (1): 1–24.

de Janvry, Alain, Marcel Fafchamps, and Elisabeth Sadoulet. 1991. "Peasant Household Behavior with Missing Markets: Some Paradoxes Explained." *Economic Journal* 101: 1400–17.

Diagana, B. N., and V. A. Kelly. 1994. "Noncropping Income and Cropping Productivity in the Senegalese Peanut Basin." Michigan State University, Department of Agricultural Economics, East Lansing.

Hoff, Karla, and Joseph Stiglitz. 1990. "Imperfect Information and Rural Credit Markets—Puzzles and Policy Perspectives." *The World Bank Economic Review* 4 (3): 235–50.

Hoffman, B., and F. Heidhues. 1993. "Credit Access in the Informal Sector: A Case Study of Two Villages in the Province Oueme in Benin." In C. Cuevas and M. Benoit-Cattin, eds., *Finance and Rural Development in West Africa.* Montpelier, France: International Cooperation Center of Agronomic Research for Development (CIRAD).

Holt, Sharon. 1991. "Village Banking: A Cross-Country Study of a Community-Based Lending Methodology." World Bank, Washington, D.C.

Holt, Sharon, and Helena Ribe. 1991. *Developing Financial Institutions for the Poor and Reducing Barriers to Access for Women.* World Bank Discussion Paper 117. Washington, D.C.

Hossein, M. 1988. *Credit for Alleviation of Rural Poverty: The Grameen Bank in Bangladesh.* Research Report 66. International Food Policy Research Institute, Washington, D.C.

IPC (Integriertes Projekt Centrum). 1988. "Rural Finance in Ghana." A research study for the Bank of Ghana. Frankfurt.

Jackelen, Henry, and Elisabeth Rhyne. 1991. "Towards a More Market-Oriented Approach to Credit and Savings for the Poor." *Small Enterprise Development* 2 (4): 4–20.

Kiiru, W. K., G. D. Pederson, and C. Nzioka. 1995. "Kenya Rural Enterprise Program: A Case Study of Sustainable Micro-Finance Scheme." World Bank, Environmentally Sustainable Development, Washington, D.C.

Nissanke, Machiko, and Ernest Aryeetey. 1995. "Financial Integration and Development in Sub-Saharan Africa." Report prepared for the World Bank, Africa Technical Department, Washington, D.C.

Popiel, P. A. 1994. *Financial Systems in Sub-Saharan Africa: A Comparative Study.* World Bank Discussion Paper 260. Washington, D.C.

Rashid, Mansoora, and Robert M. Townsend. 1993. "Targeting Credit and Insurance: Efficiency, Mechanism Design, and Program Evaluation." Education and Social Policy Department Discussion Paper 47. World Bank, Washington, D.C.

Reardon, Thomas, Eric Crawford, and Valerie Kelly. 1994. "Links between Nonfarm Income and Farm Investment in African Households: Adding the Capital Market Perspective." *American Journal of Agricultural Economics* 76 (December): 1172–76.

Riedel, J., Ernest Aryeetey, O. Gyekye, P. W. K. Yankson, F. Owusu, E. T. Quartey, K. F. Schaedler, and H. Schmitz. 1988. *Small-Scale Manufacturing and Repair Activities in Techiman/Ghana.* Institute for Economic Research, Department for Development Studies, Munich.

Shipton, Parker. 1991. "Time and Money in the Western Sahel: A Clash of Cultures in Gambian Rural Finance." In Michael Roemer and Christine Jones, eds., *Markets in Developing Countries: Parallel, Fragmented, and Black.* San Francisco: International Center for Economic Growth and Harvard Institute for International Development.

SODECON (Social Development Consultants Ltd.). 1990. *Financial Intermediation for Microenterprises in Bangladesh, Kenya, Tanzania, and Zambia.* Helsinki.

Soyibo, Adedoyin. 1996a. "Financial Linkage and Development in Sub-Saharan Africa: A Study of Formal Finance in Nigeria." Overseas Development Institute Working Paper 88. London.

———. 1996b. "Financial Linkage and Development in Sub-Saharan Africa: A Study of Informal Finance in Nigeria." Overseas Development Institute Working Paper 90. London.

Stiglitz, Joseph. 1989. "Rational Peasants, Efficient Institutions, and a Theory of Rural Organization: Methodological Remarks for Development Economics." In Pranab Bardhan, ed., *The Economic Theory of Agrarian Institutions.* Oxford: Clarendon Press.

Stiglitz, Joseph, and Andrew Weiss. 1981. "Credit Rationing in Markets with Incomplete Information." *American Economic Review* 71 (3): 393–410.

Udry, Christopher. 1990. "Credit Markets in Northern Nigeria: Credit as Insurance in a Rural Economy." *The World Bank Economic Review* 4 (3):251–269.

———. 1994. "Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria." *Review of Economic Studies* 61 (3): 495–526.

Webster, Leila, and Peter Fidler, eds. 1996. *The Informal Sector and Microfinance Institutions in West Africa*. A Regional and Sectoral Study. Washington, D.C.: World Bank.

World Bank. 1989. *World Development Report 1989: Financial Systems and Development*. New York: Oxford University Press.

———. 1994. *Adjustment in Africa: Reforms, Results, and the Road Ahead*. A Policy Research Report. New York: Oxford University Press.

Yaron, Jacob. 1992. *Successful Rural Financial Institutions*. World Bank Discussion Paper 150. Washington, D.C.

———. 1994. "What Makes Rural Finance Institutions Successful?" *The World Bank Research Observer* 9 (1): 49–70.

Yotopoulos, Pan, and Sagrario Floro. 1991. "Transaction Costs and Quantity Rationing in the Informal Credit Markets: Philippine Agriculture." In Michael Roemer and Christine Jones, eds., *Markets in Developing Countries: Parallel, Fragmented, and Black*. San Francisco: International Center for Economic Growth and Harvard Institute for International Development.

Comment on "Rural Finance in Africa: Institutional Developments and Access for the Poor," by Ernest Aryeetey

Jean-Philippe Platteau

Characterizing formal and informal financial markets in Africa is not an easy task given the paucity of hard data and the lack of research. Within these limitations, Ernest Aryeetey has nevertheless managed to produce an essentially correct characterization of African financial markets. He rightly emphasizes the extreme fragmentation of these markets, including the almost complete lack of linkages between formal and informal lenders, and identifies several striking features of most of the loans granted by informal lenders, namely their small size, short maturities, and use for consumption purposes.

In attempting to understand the fragmented nature of African financial markets, Aryeetey draws heavily on the recent economic literature that accounts for market imperfections largely in terms of informational problems. More can nevertheless be said about which factors specific to Africa—so-called structural factors—tend to make these problems especially acute in this region. Aryeetey notes that loans by village banks are so small as to discourage rich people from applying and that their maturities are short to bring the disciplining device of repeat borrowing into play. In this comment I would like to complement Aryeetey's diagnosis and, in some instances, to stressing points that he has made in a somewhat incidental manner in his article.

Imperfections in Other Rural Markets

As has been convincingly argued by Binswanger and others, agrarian institutions cannot be properly explained unless both information problems and critical material determinants are simultaneously and jointly taken into account (Binswanger and Rosenzweig 1986). In particular, when population density is low (and land abundant) and high transport costs give rise to spatial isolation, technology is simple, and seasonality is pronounced, regular labor, output, and credit markets cannot develop to any significant extent (Binswanger and McIntire 1987). Credit markets do not

Jean-Philippe Platteau is professor of economics at the University of Namur, Belgium.

Annual World Bank Conference on Development Economics 1996

develop because of both demand and supply constraints: demand is low because technology is simple and the hiring of labor is rare, and supply is limited because of the absence of good collateral (land has no sales value). Moreover, the restricted scope of output markets makes it impossible to use marketing-credit links as a substitute for collateral, and limitations in the labor market impede the emergence of labor-credit links for the same purpose. Constraints on credit supply and demand in land-abundant environments also imply that loan amounts are small and that villagers borrow mainly for consumption, with the consequence that loan maturities are short (Binswanger and McIntire 1987). The seasonality of demand for credit and the covariance of risks in the localized settings of small isolated communities imply the absence of financial intermediation (lenders lend out of equity).

Clearly, there is a tight interdependence among the aforementioned market failures in land-abundant economies; as a result it is hard to see how credit market imperfections can be remedied in a self-sustainable way (that is, without continuous support by external agencies) if imperfections in other rural markets are not corrected. As long as these other market imperfections exist, as they do in many Sub-Saharan countries, most credit activities will remain confined to narrow networks of relatives, neighbors, and friends.

Unfortunately, when more sophisticated agricultural techniques such as irrigation become available and can play a useful role—because of their ability to enhance land yields and to stabilize agricultural incomes—the implied technical shift may be impossible in the absence of heavy subsidies from outside, at least as long as population density remains low. The problem is exacerbated when communication infrastructure is deficient, which it tends to be precisely because of the high cost of providing such infrastructure when the population is scattered.

Let me elaborate a bit. In areas suitable for irrigation, tubewells and other mechanical pumping devices are essential for securing complete control of water. When the population is scattered, however, the unit costs of servicing complex mechanical equipment are high. Given roughly five to six times higher rural population density per arable hectare in the Punjab than in the arable Sahel, a service center with a given radius of activity would require five to six times as many adoptors per square kilometer in the Sahel to be economically viable (Delgado and Ranade 1987). Because of long distances between customers—think, for example, of the dispersed location pattern of small-scale irrigation perimeters along the Senegal or the Niger river—the market for mechanical servicing may fail to emerge. Alternatively, natural monopolies may become established, with farmers being charged excessively high prices (Platteau 1995).

The same problem arises in connection with credit markets, where it is likely to be even more serious given the crucial importance of private information problems. Thus information collection and loan monitoring and recovery are all the more difficult and costly when distances between the financial intermediary and the potential borrowers are large. Without external intervention, irrigators will be unable to obtain the loans required to purchase their mechanical equipment since no sizable lender will be ready to engage in credit transactions. There is apparently nothing

wrong with this logic since the business for which credit is not forthcoming is unprofitable for the same reason that lending is not profitable. This said, the African continent is full of development projects that subsidize this kind of operation, thereby facilitating the emergence of economic activities that are not sustainable in the medium and perhaps even the long run. Unless externalities of some sort can be invoked to justify such interventions, the financing of unsustainable projects is harmful insofar as it increases Africa's dependence on external aid and obscures the hard logic of economic efficiency.

In contrast, when population density rises, rural output and input markets tend to develop on their own, facilitating the emergence of credit markets through both demand and supply effects. On the one hand demand for credit increases because there are expanding market opportunities in both agricultural and nonagricultural production, technology becomes more complex, and the need for working capital grows (partly because the hiring of labor becomes increasingly widespread). On the other hand supply constraints tend to vanish, or at least to decline, because land acquires a sales value and may therefore become acceptable as collateral for willing lenders and because transactions costs diminish with the rising density of customers.

The collateral effect should not be overestimated, as it has been in much of the literature: as Aryeetey notes, foreclosure on mortgaged land may prove difficult when titles do not exist and legal systems are malfunctioning. Even when titles are established, foreclosure often turns out to be difficult for reasons pertaining to the political economy of a country or the region (Platteau 1996a). Conversely, as I have found in Rwanda (André and Platteau 1996) and Burkina Faso, land mortgage and foreclosure may take place in the absence of formal titles, within personalized networks such as rotating credit and savings associations. When population density is sufficiently high, however, informal credit markets tend to develop because of the above-mentioned credit-marketing and credit-labor nexus. There is ample evidence for this effect in the densely populated fishing communities of West Africa, where fish merchants have emerged as an important source of medium-term credit for financing working capital expenditures and possibly of long-term credit for meeting fixed capital needs. Some of these funds may even reach crew laborers as current consumption loans extended by boat owners.

The lesson from this discussion is probably that, if the main objective is achieving long-term sustainability, institutional innovations in the credit sphere should first be attempted in areas that are most conducive to financially viable credit business, that is, in areas where population density is comparatively high, where alternative technologies are available, and where market opportunities are expanding. The number of such areas is rapidly increasing in Africa, and the effectiveness of microcredit programs will be more reliably tested there than in areas where the structural conditions for efficient financial intermediation are not met.

Though Aryeetey considers that credit constraints arise from the supply rather than the demand side of the market, the discussion thus far has shown that both demand and supply limitations can account for the low activity of credit markets in underpopulated regions of Africa. More puzzling is the fact that even in densely

populated countries credit may be underdeveloped because of a lack of demand. In Rwanda, one of the African countries that has been most successful in developing rural networks of people's savings and credit organizations (known as *banques populaires*), 9 percent of depositors borrow money from their local banks, and few loans are taken for agricultural production despite remarkable efforts to disseminate information about investment opportunities through linkages between banques populaires and rural development projects (Gentil and Fournier 1993). The situation is similar in neighboring Burundi. How should we interpret this striking finding? Does it mean that farmers perceive the investment opportunities as unprofitable or too risky, or are they reluctant to take loans for other reasons (such as social stigma)? Clearly, we need to know more about the attitude of African farmers toward savings, investment, and credit before we can devise appropriate strategies.

Sharing Norms

In addition to material determinants, sociocultural features of African societies, particularly the pervasive presence of sharing norms, impinge on the effectiveness of financial markets. These norms encourage rich individuals to share their prosperity with less fortunate members of their social group (particularly the extended family or the clan). As long as incomes are subject to short-term variations and these variations are poorly correlated across households, gifts tend to be reciprocated in the not-too-distant future. In this case sharing norms help sustain a system of voluntary transfer payments, or state-contingent loans, that provides mutual social insurance against noncovariate risks, thereby serving as a substitute for the missing insurance markets. Moreover, as economic theory shows, the implicit contracts involved are self-enforcing because reputational effects are brought to bear as soon as repeated interactions occur among the participants (see, for example, Coate and Ravallion 1993). It is therefore not surprising that among relatives and close personal acquaintances loans or transfers (the distinction is not clear-cut) of small amounts and short duration (used mainly for consumption) are common in Africa.

When prosperity—perceived in Africa as resulting largely from luck—appears to be sustained over a longer period of time, one should not expect that "lucky" individuals will spontaneously let other members benefit from their situation. It is here that social norms of sharing, usually backed by beliefs in witchcraft, prove necessary to motivate rich and dynamic individuals to share their incomes. (Persisting "luck" is often attributed to the manipulation of supernatural forces, which may jeopardize the survival of the social group; Platteau 1996b.)

The implication of this feature of African societies for informal loan transactions is straightforward. An individual who becomes rich enough to lend money to people in need of credit is particularly vulnerable to local social pressures for income redistribution, and a failure to comply with these pressures is likely to arouse envy and hatred. Consequently, the business of granting medium- or long-term loans to fellow villagers—which the rich individual theoretically can afford to do out of

accumulated equity funds—is highly risky: poorer borrowers will not feel morally obliged to repay the loans because borrowers are likely to believe that nonrepayment is a legitimate way to redistribute income between rich and poor people.

Under such circumstances rich individuals will refrain from establishing a lending business; their reluctance to do so partly explains why there are few long-term credit transactions in the sphere of personalized relations. Of course, rich individuals could move outside their native area and settle in places where they will not be subject to strong pressures for redistribution. In this case, however, the informational advantages of operating in a familiar milieu are lost. A serious market failure problem may therefore arise even where population density is high and geographic isolation is low.

The problem is not insuperable, however. Collective savings and credit associations such as the *tontines* avoid the problem by organizing their saving within a cooperative framework that ensures that all participating members will improve their position. Tontines thus provide a socially accepted alibi for protecting people's savings from all sorts of social pressures (Henry, Tchente, and Guillerme-Dieumegard 1991). This is a hopeful sign since efforts to build up financial organizations that rest on local savings and credit cooperatives are still common in Africa. Yet Aryeetey rightly points out that innovative credit schemes such as group lending and cooperative credit associations have not lived up to expectations. Why this is so is a complex question that is hard to answer definitively, given the limited amount of detailed information available.

Letting Indigenous Institutions Grow

Understanding the causes underlying the variations in performance of different types of innovative credit schemes in Africa is an urgent task. One factor that will probably prove important is the degree of external intervention. There is evidence that under certain conditions externally provided rewards can reduce individuals' intrinsic incentives to act and contribute and impair their capacity for self-reliance. This crowding-out effect is certainly at play in many African countries, where chronic dependence on external assistance has instilled in people the notion that solutions to their problems will come from the outside.

This sense of dependency is inherent in the African distinction between "hot" and "cold" money. Hot money is obtained from one's own efforts and should therefore be well taken care of. Cold money is perceived as a windfall, a gift from rich agencies with boundless funds at their disposal. These rich agencies are generally located abroad and their resources are channeled through aid development projects or local governments. As a result the discipline of repeated interactions does not come into play.

Innumerable examples of the failure of credit schemes can be attributed to the fact that from the start beneficiaries never thought that they would have to repay the loans or that they believed that they were entitled to the loans, which they saw as gifts. High delinquency rates are systematically observed when the money comes

from the state, either through public agencies or so-called people's organizations (such as credit cooperatives) established solely to channel state funds.

Following disappointing experiences with state channels and organizations, external aid agencies, both official and private, have tended to jump into local initiatives involving local participation and to offer financial and technical support. This is a serious mistake. Local organizations must be allowed to develop at their own pace, using their own resources and efforts, even though this may mean that development will be slow. External support is needed, of course, but too much support provided too quickly will undermine intrinsic motivation and retard rather than accelerate development. Everyone familiar with Africa is impressed by the dynamism of its emerging forms of collective organization: deep and rapid changes are occurring in this part of the world, but they usually take place below the surface, so that standard indicators fail to note them. It is crucially important that these changes not be perverted by hasty interventions by external agencies eager to show that success is possible in Africa.

References

André, Catherine, and Jean-Philippe Platteau. 1996. "Land Tenure under Unbearable Stress: Rwanda Caught in the Malthusian Trap." *Cahiers de recherche de la Faculté de sciences économiques et sociales* 164. University of Namur, Belgium.

Binswanger, Hans, and Mark Rosenzweig. 1986. "Behavioral and Material Determinants of Production Relations in Agriculture." *Journal of Development Studies* 22 (3): 503–39.

Binswanger, Hans, and John McIntire. 1987. "Behavioral and Material Determinants of Production Relations in Land-Abundant Tropical Agriculture." *Economic Development and Cultural Change* 36 (1): 73–99.

Coate, Stephen, and Martin Ravallion. 1993. "Reciprocity without Commitment: Characterization and Performance of Informal Insurance Arrangements." *Journal of Development Economics* 40 (1): 1–24.

Delgado, Christopher L., and C. G. Ranade. 1987. "Technological Change and Agricultural Labor Use." In John W. Mellor, Christopher L. Delgado, and Malcolm J. Blackie, eds., *Accelerating Food Production in Sub-Saharan Africa.* Baltimore, Md.: The Johns Hopkins University Press for the International Food Policy Research Institute.

Gentil, Dominique, and Yves Fournier. 1993. *Les paysans peuvent-ils devenir banquiers? Epargne et crédit en Afrique.* Paris: Syros.

Henry, Alain, Guy-Honoré Tchente, and Philippe Guillerme-Dieumegard. 1991. *Tontines et banques au Cameroun—Les principes de la Société des amis.* Paris: Karthala.

Platteau, Jean-Philippe. 1995. "The Food Crisis in Africa: A Comparative Structural Analysis." In Jean Dreze, Amartya Sen, and Athar Hussain eds., *The Political Economy of Hunger: Selected Essays.* Oxford: Clarendon Press.

———. 1996a. "The Evolutionary Theory of Land Rights as Applied to Sub-Saharan Africa: A Critical Assessment." *Development and Change* 27 (1): 29–86.

———. 1996b. "Traditional Sharing Norms as an Obstacle to Economic Growth." *Cahiers de recherche de la faculté des sciences économiques et sociales* 173. University of Namur, Belgium.

Comment on "Rural Finance in Africa: Institutional Developments and Access for the Poor," by Ernest Aryeetey

Dominique van de Walle

In his review of the theoretical literature on credit and the empirical evidence on credit markets and credit schemes in Africa, Ernest Aryeetey argues that informal rural finance systems perform better than formal systems because they have adapted to the high-risk environment. He suggests that the formal sector should learn from informal institutions. It struck me that government interventions aimed at providing safety nets and reducing vulnerability also might benefit from a study of the particulars of successful informal credit systems in Africa. However, I want to limit my observations here to innovative credit schemes targeted to the poor and to ground my remarks firmly in the session's theme of reducing poverty through targeted programs and rural credit.

Coming to this literature from a poverty perspective rather than as an expert on credit, I have long been puzzled. The rhetoric of the policy discussion about targeted credit schemes for the poor is in obvious tension with the prevailing analytics. And it is the analytics that need to catch up.

In the absence of market failures it is not at all obvious that subsidizing credit for poverty reduction objectives would ever be considered. Credit is not an inferior good—or even one with a low but positive income elasticity of demand—and we know that lump sum transfers would do the job of redistribution more efficiently in a world of perfect markets. But credit (and other) markets plainly do not work perfectly—at least by first-best standards. The existence of market failures means that there are potential efficiency gains from intervention. What links credit market failure to poverty reduction is that worthwhile welfare-raising investments are not being made in the economy and that it is likely to be the poor rather than the nonpoor who suffer most. The costs of credit market failure will probably be highest for the poorest, who are least able to self-insure, self-finance, and provide collateral. But as with all interventions aimed at reducing poverty, the real policy issue is whether the marginal impact on poverty from this type of policy is higher than it would be under the

Dominique van de Walle is an economist in the Public Economics Division of the Policy Research Department at the World Bank. The author is grateful to Karla Hoff and Martin Ravallion for useful comments.

Annual World Bank Conference on Development Economics 1996

best alternative use of the same budget. That is a difficult question. My comments concern how we might go about answering it.

Assessing Targeted Credit Schemes

It is often claimed that the goal of directed rural credit schemes is poverty reduction. Indeed, credit for the rural poor and women is championed by some as the single best way to attack poverty. For example, the World Bank now has a Consultative Group to Assist the Poorest that is devoted to supporting targeted credit schemes. Yet—despite the emphasis on poverty in the rhetoric of these schemes—the overwhelming focus of the policy analytics is on self-sustainability as the benchmark for success; it is all about loan repayment rates, loan sizes, numbers of members and borrowers, savings levels, and financial performance. This emphasis is to some extent grounded in the literature in economics on credit markets under imperfect information (Hoff and Stiglitz 1990; Besley 1994; Yaron 1994). The proponents of innovative targeted credit schemes—who by and large seem to be concerned with pro-poor distributional impacts—appear to have sought legitimacy in a paradigm based largely on somewhat narrow efficiency criteria.

In keeping with the literature, Aryeetey uses the performance criteria proposed by Yaron (1994)—self-sufficiency, as measured by the subsidy dependence index, and the extent of outreach—to review the African experience with targeted credit for the poor. Best practice, as evidenced in meeting these goals, has been distilled from "successful" innovative credit schemes and transplanted to Africa. Aryeetey implicitly questions this practice by presenting evidence that donor-driven targeted credit schemes often fail to offer services not already available through the informal sector in many African countries and that the schemes are not always compatible with local institutional environments. This is an important observation.

I would question the basis on which innovative targeted credit programs are advocated and practiced on other grounds as well. Once credit schemes are put in the broader context of poverty reduction policy options, the optimal degree of sustainability will not be "self-sufficiency" if it can be established that a dollar spent on targeted credit has a greater impact on poverty than a dollar spent on alternative policies. Put another way, advocating self-sufficiency implicitly assumes that these are not cost-effective poverty reduction instruments relative to alternatives. By stating that the objective of these credit schemes should be self-sustainability one is ipso facto saying that subsidies should instead be going to other programs that governments subsidize in the name of the poor—notably basic health care, basic education, and cash transfers. That may or may not be correct—we do not know. But it is doubly curious since proponents of self-sufficiency for innovative targeted credit schemes often also insist that targeted credit is an effective poverty reduction scheme.[1]

To add to the confusion, it seems that most of these schemes are in fact being subsidized by donors. Thus the rhetoric of self-sufficiency is also in some tension with the practice. The aim of these schemes is apparently to reach self-sufficiency within some period. The question is how long that period—and hence the total subsidy—should be.

What do we need to know to resolve this issue? As soon as the discussion turns to spending public money on these schemes, it has entered the domain of public economics. The question then is whether, when these schemes are put side by side with other programs, the welfare gains associated with these schemes justify the public outlay. There has been surprisingly little discussion of the benefits of targeted credit schemes—the impacts on the productivity, incomes, or living standards of borrowers. (Exceptions include Rashid and Townsend 1994; Pitt and Khandker 1996; and McKernan 1996.) Given their goal of poverty reduction through income and employment generation, surely these are the relevant measures of performance. Severe methodological problems have marred past attempts at such evaluations; the inherent difficulties in evaluating impacts may have discouraged others from adopting this approach.

More recently, several studies have credibly tested the welfare impacts of these schemes. A few careful studies have attempted to deal with the self-selection and endogeneity of program placement issues that arise in estimating program impacts. Pitt and Khandker (1996) look at the household welfare impacts of three targeted credit schemes in Bangladesh, paying particular attention to the gender-specific impacts. McKernan (1996) tries to isolate the impacts of the credit component from the noncredit "social development" components of the same Bangladeshi programs. This kind of rigorous analysis is sorely needed to better inform policy. However, because these analyses do not introduce the costs of those impacts or compare the net benefits with those of other programs, they cannot be used to determine the best combination of programs.

Clearly, we need comprehensive estimates of distribution-weighted net benefits or (equivalently) measures of poverty reduction for a given cost. And we need comparable estimates for different schemes so that the most cost-effective measures for reducing poverty can be chosen (van de Walle 1995; Ravallion and Datt 1995). Ideally, the measures would include all the costs (including participation costs, disincentive effects, and political economy costs) and all the benefits (including dynamic effects and benefits to facets of nonincome poverty). In practice this would require a more complete understanding of the structure of an economy than is within our reach. But we should not fall prey to ignoring the omissions when it comes to making policy prescriptions.

This kind of analysis could form the basis of a sensible framework for determining what the optimal subsidy to targeted credit should be. If findings indicate that targeted credit schemes are not cost-effective in reducing poverty, then self-sustainability, not poverty, should be the measure used to gauge the success of these schemes. If targeted credit schemes are found to be cost-effective in reducing poverty, these schemes should be subsidized and allocations to other public spending categories cut until it ceases to be true that these schemes are preferable at the margin.

It is sometimes said that subsidies will destroy incentives for escaping poverty. However, there is no reason targeted credit schemes cannot be subsidized in the aggregate while preserving incentives through design features such as setting interest rates above the rate obtainable by the rich but below the level obtainable by the poor (which may be infinite), using group-based lending and other forms of peer

group monitoring, extending small loans, requiring participation in social development programs, and so on. The experience of microenterprise credit has provided some valuable lessons on good design features.

How Well-Targeted Are Targeted Credit Schemes?

One factor that determines whether targeted credit schemes are more cost-effective than other policies is their targeting performance. As in much of the rest of the literature, Aryeetey's article takes it on faith that directed and innovative credit schemes reach and help the poor. Evidence that women and the illiterate feature prominently among participants in African schemes is given as proof that the schemes are pro-poor. The literature examining the many Bangladeshi schemes also takes it on faith that the target groups of women and the functionally landless represent the poor in Bangladesh (Khandker, Khan, and Khalily 1995; Pitt and Khandker 1996; and McKernan 1996). But there are reasons to question this assumption. It has been found that what constitutes poverty by cultural or social standards is often far from identical to what constitutes poverty as defined by objective criteria (for an example in India, see Lanjouw and Stern 1991). Land is a symbol of wealth in rural Bangladesh, and landlessness of poverty. One study that found a strong negative correlation between landholding and poverty in Bangladesh still found that perfect targeting to the landless entailed sizable leakage to the nonpoor and imperfect coverage of the poor: some of the poor owned land, while some of the rich (including teachers, doctors, and shopkeepers) did not (Ravallion and Sen 1994). Differences in land quality make the relationship even more complicated.

Do targeted credit schemes reach the poor? Design features undoubtedly help promote targeting to the poor. Given the loan sizes and the fact that programs such as the Grameen Bank impose participation costs, the rich will surely have better alternatives. But the poorest of the poor are probably not being reached extensively by microcredit schemes (Sen 1995; Mosley 1996). Common sense suggests that the poorest segments of the population are unlikely to have the appropriate skills or health status or access to the kinds of institutional and physical infrastructure necessary to make productive investments profitable. And they are more likely to require consumption loans.[2] My own work assessing social safety net and antipoverty programs in Vietnam confirms this. Credit for income-generating activities has been embraced as the primary poverty alleviation measure for the able-bodied. But households that are too poor to use a loan productively or that might use the loan for consumption purposes are not extended credit.[3]

The results of at least one careful evaluation that allows for self-selection into the schemes finds that "high-profit households are more likely to participate in Grameen Bank" (McKernan 1996, p. 31). Schemes that rely on peer monitoring and intimate local knowledge are often able to screen out bad credit risks. This does not on its own imply that governments should not pursue targeted credit programs; it does suggest that we should also be looking at complementary interventions aimed at the poorest of the poor.

Notes

1. The self-sufficiency measures developed by Yaron (1994) and others are intended to help measure the cost of subsidized credit. Yaron does not claim that self-sufficiency is the goal.

2. Many successful group lending programs have tied insurance—or consumption loans—to credit for directly productive investments. The Grameen Bank requires individuals to contribute to an insurance fund. It also builds schools and promotes training programs. The objective is to use the group insurance fund to increase households' ability to take on risky productivity activities, to increase human capital, and to strengthen the community's commitment to the Grameen Bank (thereby lowering default rates by individuals who expect to "graduate" from eligibility for direct loans).

3. The view that credit should be used only for directly productive purposes has been thrown into doubt by recent research indicating that liquidity constraints and (hence) exposure to uninsured risk often force poor people to take costly actions that reduce their productivity (Morduch 1995).

References

Besley, Timothy. 1994. "How Do Market Failures Justify Interventions in Rural Credit Markets?" *World Bank Research Observer* 9 (1): 27–47.

Hoff, Karla, and Joseph Stiglitz. 1990. "Introduction: Imperfect Information and Rural Credit Markets—Puzzles and Policy Perspectives." *The World Bank Economic Review* 4 (3): 235–50.

Khandker, Shahidur R., Zahed Khan, and Baqui Khalily. 1995. *Sustainability of a Government Targeted Credit Program: Evidence from Bangladesh.* World Bank Discussion Paper 316. Washington, D.C.

Lanjouw, Peter, and Nicholas Stern. 1991. "Poverty in Palanpur." *The World Bank Economic Review* 5 (1): 23–55.

McKernan, Signe-Mary. 1996. "The Impact of Micro-Credit Programs on Self-Employment Profits: Do Non-Credit Program Aspects Matter?" Brown University, Department of Economics, Providence, R.I.

Morduch, Jonathan. 1995. "Income Smoothing and Consumption Smoothing." *Journal of Economic Perspectives* 9 (3): 103–14.

Mosley, Paul. 1996. "Growth Versus Equity?" In David Hulme and Paul Mosley, eds., *Finance Against Poverty.* London: Routledge.

Pitt, Mark M., and Shahidur Khandker. 1996. *Household and Intrahousehold Impacts of the Grameen Bank and Similar Targeted Credit Programs in Bangladesh.* World Bank Discussion Paper 320. Washington, D.C.

Rashid, Mansoora, and Robert M. Townsend. 1994. "Targeting Credit and Insurance: Efficiency, Mechanism Design and Program Evaluation." Education and Social Policy Department Discussion Paper 47. World Bank, Washington, D.C.

Ravallion, Martin, and Gaurav Datt. 1995. "Is Targeting through a Work Requirement Efficient? Some Evidence from Rural India." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence.* Baltimore, Md.: The Johns Hopkins University Press.

Ravallion, Martin, and Binayek Sen. 1994. "Impacts of Land-Based Targeting on Rural Poverty: Further Results for Bangladesh." *World Development* 22 (6): 823–38.

Sen, Binayek. 1995. "Income-Earning Environment of the Poor: Aspects of Dislocation, Adjustment, and Mobility." In Hossain Zillur Rahman and Mahabub Hossain, eds., *Rethinking Rural Poverty: Bangladesh as a Case Study.* New Delhi: Sage Publications.

van de Walle, Dominique. 1995. "Incidence and Targeting: An Overview of Implications for Research and Policy." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence.* Baltimore, Md.: The Johns Hopkins University Press.

van de Walle, Dominique, and Kimberly Nead, eds. 1995. *Public Spending and the Poor: Theory and Evidence.* Baltimore, Md.: The Johns Hopkins University Press.

Yaron, Jacob. 1994. "What Makes Rural Finance Institutions Successful?" *The World Bank Research Observer* 9 (1): 49–70.

Floor Discussion of "Rural Finance in Africa: Institutional Developments and Access for the Poor," by Ernest Aryeetey

A participant from the World Bank's International Trade Division asked Jean-Philippe Platteau (discussant) about the relation between low population density and the unavailability of credit and other services. If there is an inverse relation at work—as population density increases, so does demand for credit and, therefore, the supply of credit and other services—perhaps less land per capita is not as bad as we have thought it was, since it means that credit and other services are forthcoming. But, he asked, is there something else at play in Africa? For example, is the quality of the land poorer than in, say, South Asia, where people would gladly give up higher density for more land. Is there something specific about the areas of low population density in Africa that sets Africa apart from other areas of low population density, such as in North and South America, where things have ended up very differently? In other words, does limited demand explain the differences in access to credit between Africa and other regions, or are other factors, such as the quality of land, also at work?

Platteau agreed that the quality of land is important. In areas where land is fertile, population density is high. But land is not abundant in all of Africa: in countries like Burundi, Kenya, Malawi, and Rwanda that is certainly not the case. Platteau's point was that innovative credit schemes, such as those Aryeetey proposes, are most likely to succeed in areas where conditions allow for the emergence of informal credit markets. He was indirectly suggesting an area for further research, because it is not enough to say that credit markets do not work in Africa. If research were able to determine what made things work or not work, better interventions could be designed.

A private consultant asked why the panelists thought it was so difficult to implement credit schemes in low-density areas of Africa. After all, Africa has had some success with savings schemes—in sparsely populated Congo, for example, mobile savings banks have been successful—and savings schemes can or should be linked to credit schemes. And in Central America, where he had been a credit officer, it was

This session was chaired by Jean-Louis Sarbib, vice president, Africa Region, at the World Bank.

Annual World Bank Conference on Development Economics 1996

easy to persuade commercial banks to lend to the poor. There is no reason to be so pessimistic about getting credit to the poor; it is just a matter of trying and experimenting.

Aryeetey agreed that there had been successful schemes in Africa, but often at considerable cost—and cost is a critical issue in any effort to increase access to credit. If increasing access requires subsidies, who will pay for them? Platteau added that the participant was right to point out that people often were eager to deposit their savings somewhere to keep them safe, but, he added, it is not all that easy to shift from savings to credit schemes. In densely populated areas, such as the fishing communities along Senegal's coast, schemes had been devised for fishermen to make collective deposits in a bank. But in three years none of that that money had been used to make loans, and it is not clear why. If one were to ask the fishermen what they needed, they would say credit. Yet a credit facility is available that they do not use. Explaining the credit needs of the poor is a topic requiring more research.

A gender specialist from the World Bank asked Dominique van de Walle (discussant) about the household welfare effects and net benefits of targeted credit schemes and alternatives to targeted credit. Even though, as van de Walle mentioned, only a few studies are available, had they drawn any conclusions on the benefits of investing in women rather than in men? Van de Walle said that the papers suggested that credit schemes in Bangladesh had improved women's income and status within the household, though she was not certain of their costs. The social development element of credit from the Grameen Bank had huge administrative costs, but it could be that much of the bank's success comes from the education, training, and family planning information it provides. One study that separated the credit from the social development impact found the social development component to be extremely important. Because the studies are relatively recent, she was less comfortable drawing conclusions about the schemes' impacts on children. But one of the findings seems to be that some girls end up staying home from school because their mothers have set up productive enterprises that benefit from having additional help.

A participant from the World Bank's Agriculture and Natural Resources Department questioned what he viewed as van de Walle's implicit criticism of outreach and self-sustainability indicators as appropriate measures of performance. If the goal is to reduce poverty and increase incomes, he asked, how else should we measure performance? Impact indicators are notoriously difficult to assess. It would be impossible, for example, to measure the change in net lifetime earnings resulting from an education intervention. So we look at output indicators, which tend to be more measurable, assuming that if we achieve a certain output, we will have achieved a certain impact. Outreach and self-sustainability are output indicators associated with an ultimate impact, and it would be difficult to make a case for measuring an impact as a readily available alternative to such output indicators.

Van de Walle responded that she was not saying that impact measurements are readily available, but that efforts should be made to focus on impacts. A tremendous amount of money goes into credit schemes in Bangladesh—something like $500 million a year for the Grameen Bank alone. And yet it is not clear what effect the money

is having on poverty. There is no way of knowing the counterfactual, of course, and many things have happened in Bangladesh. Still, poverty has increased. If the objective is to reduce poverty, it is important to consider whether alternative approaches might have worked better. The literature on these schemes does not emphasize poverty reduction; it emphasizes loan repayment rates. But loan repayment rates do not tell us anything about what borrowers do with their loans. Aryeetey agreed with many of van de Walle's views and admitted that his background tended to make him view the issues as a financial systems analyst. When he wrote the article he had concluded that it was impossible, with the data available, to examine the effects of many schemes in Africa, but he agreed that it is fertile ground for research.

The president of a private communications service asked the panelists whether they saw the private sector as an asset or a threat in the provision of credit. The private sector has fostered the development of service industries in many parts of the world, including the computer industry in India and fiber optics in Africa and Latin America. The private sector is able to generate revenues and resources that could be used to foster independence, but there are always conflicts with the public sector, in the form of major regulations and other limitations that hinder productivity.

The private sector will not start providing credit to the poor because it wants to reduce poverty, responded Aryeetey. It will do business with the poor only if it can turn a profit, but not otherwise. Governments may see credit schemes as a possible solution to poverty problems and thus may try to establish a relationship with the private sector. But many investors are convinced that lending to the poor is not profitable.

Someone had said that one way to get financial systems to function properly would be to improve efficiency, so that banks had a reason to operate in rural areas. Aryeetey said that he had once shared that view but that he had changed his mind. He had seen banks in different countries after reform, after they had tried to strengthen their rural operations, and with the infrastructures they had it was almost impossible for banks to transform themselves in three to four years. Aryeetey believed that trying to force institutions to do something they were not prepared to do was the wrong approach, especially when informal institutions with a lot of experience in rural areas were capable of linking up with formal institutions that had resources, and he would encourage them to continue operating there.

In closing the session, Jean-Louis Sarbib (chair) said that he was glad to hear from the experts but that were no easy answers to many of the issues that had been raised. As somebody involved in the Bank's operations, he regularly confronted the problem of, say, getting credit to farmers in the Sahel who want to buy wheelbarrows to improve the quality of their soil or seed to plant for the next harvest. Sarbib urged the research community to address these issues with the people in operations in mind. Decisionmakers ask the people in operations very real questions, to which they often can give only imperfect answers. In many cases people in operations do not even know what to measure: The sustainability of the system? The efficiency of subsidies? Or something else?

Legal Systems and Economic Development

The Rule of State Law and the Rule-of-Law State: Economic Analysis of the Legal Foundations of Development

Robert D. Cooter

In a rule-of-law state the law is consistent with social norms that embody citizens' sense of justice, and the law is obeyed out of respect. Under such a system private citizens supplement official enforcement of the law, which is critical because officials lack the information and motivation to enforce the law effectively on their own. In a rule of state law, by constrast, people disobey the law or they obey it out of fear of punishment. The rule-of-law state facilitates economic development; the rule of state law impedes it. The policy implication for the World Bank and other development institutions is that technical legal assistance should focus not on establishing new codes and regulations but on developing intermediate institutions and a community of judges, lawyers, and scholars that can shape law so that it conforms to reality.

The aim is a rule-of-law state, not the rule of state law.

—Vasili Vlasihin (1992)

A modern economy needs effective laws to promote cooperation among people. Yet states enact many laws that few people obey. People tend to disobey, or obey out of fear, laws that are not consistent with social norms and to obey laws that reflect social norms.

In a modern economy, with its many specialized business communities, social norms arise to coordinate the interactions of people (Ullmann-Margalit 1977; Taylor 1982, 1987; and Rubin 1993). Such communities may form around a new technology, such as computer software; a body of knowledge, such as accounting; or a particular product, such as credit cards. I refer to the norms of these specialized business communities as "the new law merchant" (Cooter 1994).

Robert D. Cooter is Herman F. Selvin Professor of Law at the University of California at Berkeley. This article draws on two related papers (Cooter 1994 and 1996). The author is grateful to Dhammika Dharmapala, Susan Rose-Ackerman, and Michael Trebilcock for comments. The classic argument for decentralized law is found in Leoni (1991) and Hayek (1976).

Annual World Bank Conference on Development Economics 1996

As an economy develops and the law becomes more complex, citizens need more information about the law in order to obey it. Alignment of state law with social norms enables citizens to economize on costly legal counsel by using morality as a guide to legality. A large body of evidence indicates that ordinary citizens and many businesses actually know little about the legal consequences of their acts. For example, business people often sign contracts without understanding their legal implications, take risks without understanding their legal liability, and invest in property improvements without fully understanding their ownership rights (Macaulay 1963; Ellickson 1991). This behavior is rational when state law reflects social norms. When that is the case, most people perceive the law as just, and many people obey the law out of respect, thereby creating a rule-of-law state.[1]

As an economy develops and the law becomes more complex, officials, too, need more information to enact and enforce laws. When laws reflect social norms, private citizens take risks or use their resources to help officials enforce state law (by filing complaints with the police or testifying against alleged offenders, for example). Citizens also enforce social norms by imposing informal punishments, such as gossip, rebukes, and shunning. By contrast, when state laws are inconsistent with social norms, many people perceive the law as unjust or irrelevant and disobey it or obey it out of fear of punishment, thereby creating a rule of state law. Private citizens are reluctant to help officials enforce laws that are perceived as unjust or irrelevant. When state law conflicts with social norms in such areas as contracts, property, and torts, citizens and businesses often disobey the law out of ignorance or willfulness and follow morality instead. Businesses usually ignore contractual remedies for breach, for example (Bernstein 1992; Macaulay 1963). Neighbors often ignore property law in trespass disputes (Ellickson 1991). Insurance companies sometimes allocate liabilities for claims without regard to underlying tort law (Ross 1970).

And as an economy develops and the law becomes more complex, the information and motivation constraints on laws tighten. Aligning state law with social norms loosens the information constraints on government in a growing economy. The rule-of-law state thus promotes economic development. Conversely, if law is misaligned with social norms, the constraints on information created by a growing economy tighten. The rule of state law thus impedes economic growth.

Theory of Games, Norms, and the Rule of Law

To explore the relationships between state law, effective law, and economic development, I use a model of social norms that explains how the internalization of norms strengthens people's willingness to punish violators informally. The costs of punishing a violator (expending one's resources or exposing oneself to the risk of retaliation) diminish as more people participate in punishing wrongdoers; that is, the technology of informal enforcement is characterized by increasing returns to scale and multiple equilibria. Given the existence of multiple equilibria, can a society move from an inferior equilibrium, in which few people obey the law, to a supe-

rior equilibrium, in which most people obey it? I show how state institutions and intermediate institutions can tip the balance in the direction of widespread obedience.

In the tradition of economics this article analyzes "markets" for social norms and praises decentralized law. Unlike traditional economics, however, my model assumes that some participants act morally rather than from narrow self-interest. Narrowly self-interested actors violate social norms or conform adventitiously depending on which behavior increases their wealth and power. In contrast, moral actors obey social norms out of respect, even at the cost of modest losses in wealth or power. I assume that some actors are moral.

Social Norms in Business

The agency game depicted in figure 1 is a paradigm for cooperation in business. The first player to move, the principal, decides whether to make an investment of amount "1." If the principal decides not to invest, the game ends and the players receive nothing. If the principal decides to invest, the second player, the agent, decides whether to appropriate or cooperate. Appropriation is merely redistributive: the agent appropriates the principal's investment of 1, and the sum of the payoffs in the upper right quadrant of figure 1 is 0. Cooperation by both players is productive: the principal recovers the investment, and the sum of the payoffs in the upper left quadrant of the figure is 1.

If the agency game is played only once, the agent's best move is to appropriate. Knowing this, however, the principal concludes that the best move is not to invest. The one-shot game of investment thus has a single solution, which is unproductive. An enforceable contract can overcome this inefficiency by changing the agent's incentives. The costless recovery of expected damages, for example, provides the principal with an incentive to invest regardless of the probability of breach of con-

Figure 1. Agency Game

		Agent: Cooperate	Agent: Appropriate
Principal	Invest	0.5 / 0.5	1.0 / −1.0
	Don't invest	0 / 0	0 / 0

tract by the agent. Similarly, the costless collection of expected damages from the agent provides the agent with a strong incentive to perform.

Although enforcement of contracts typically requires coercion by a third party, such as the state, the problem can also be solved without recourse to third-party enforcement. Investment in a business network usually occurs among people who maintain enduring relationships. When players commit to an enduring relationship, the problem of cooperation is usually solvable, provided that players can observe each other's moves and that they do not discount the future too heavily (Axelrod 1984). (The exceptions to this generalization—noted in Harrison and Hirshleifer 1989, Hirshleifer and Coll 1988, and Anderson and Hill 1975—need not concern us here.) Assume that the agency game depicted in figure 1 is repeated indefinitely, transforming a one-shot game into a repeated game. In any round of the repeated game in which the principal invests, the agent enjoys an immediate advantage through appropriation. A successful strategy for preventing such opportunistic behavior, called tit-for-tat, is for the principal to respond in the next round by refusing to invest and to begin investing again in a subsequent round (Axelrod 1984). The experience of immediate punishment usually suffices to stop opportunistic behavior by the agent and restore cooperation. Theory and experimental evidence indicate that tit-for-tat comes very close to maximizing a player's payoff under a variety of circumstances.[2]

Enduring relationships can be based on kinship, friendship, ethnicity, or religion, to name a few bases of commitment. Relationships substitute for state-enforced law among criminals, under communism, within tribes, and in much international trade. These groups often develop compelling ideologies to stabilize relationships. The ideology of kinship, for example, can extend biologically based relationships, as shown in the theory of cross-cutting ties of kinship developed in anthropology by Levi-Strauss and others (see Harris 1968). Rather than analyzing the anthropology of business, however, I turn to the tentative relationships that form the fabric of most modern business transactions.

Contracts solve the problem of cooperation through law, and enduring relationships solve the problem of cooperation through repetition. In tentative relationships the problem of cooperation is solved by social norms. To model tentative relationships, assume that there are an indefinitely large number of players who form pairs to play each round of the agency game. After each round some of these partnerships are maintained and others are terminated. Partnerships are terminated if the principal dissolves the partnership after the agent appropriates or the partners separate amicably after a change in business conditions makes the relationship unproductive. Appropriators play only once with any particular principal, whereas cooperators form stable relationships (Schussler 1993; Dawes and Orbell 1993).

The equilibrium concept for this kind of game draws on evolutionary theory.[3] Think of players as vehicles for competing behaviors and ask which of these behaviors will survive in competition with the others, assuming selection favors the behavior associated with the higher payoff. Thus the proportion of players using a particular strategy increases as long as that strategy produces above-average payoffs

(and that the proportion of players using a strategy decreases as long as that strategy produces below-average payoffs). Since competition tends to eliminate all below-average strategies, every strategy surviving in equilibrium earns the same rate of return.

A social norm that everyone obeys does not exist in practice; a social norm that no one obeys does not exist by definition. In practice some players cooperate and others appropriate. In such an internal equilibrium both strategies earn the same expected payoff, as required to survive.

It is easy to see why both strategies might earn the same expected payoff. When a partnership dissolves, the players must search for new partners, which requires the expenditure of resources and time. Because appropriating agents form unstable relationships and repeatedly search for partners, they expect a high payoff occasionally. In contrast, cooperating agents form stable relationships (and therefore seldom search for partners) and expect a modest payoff often.

In goods markets a stable internal equilibrium usually exists when an increase in the quantity of production causes the marginal cost of production to increase. Similarly, in the agency game a stable internal equilibrium usually exists when an increase in the proportion of appropriators causes the expected payoff from appropriation to decrease.[4]

Internalization of Social Norms

Interactions such as those captured in the agency game usually generate social norms that impose obligations on members of a community (such as the rules of conventional morality). Every individual accepts some social norms and rejects others. A person who accepts a norm "internalizes" it by making a personal commitment to be governed by the norm. Because social norms are general, a person who internalizes a norm feels that others ought to abide by it as well and so may criticize or punish people who violate it. Such informal punishment deters some potential violators.

When a significant proportion of people in a community internalize a norm, the norm directs behavior internally by respect and externally by fear. A social norm exists, according to the positive theory of law, when it achieves a minimum level of effectiveness in directing behavior within a community (Dworkin 1977). (The many refinements and criticisms of the positive theory of norms need not concern us here; see Fuller 1958, 1964 and Hart 1958.)

Internalizing a norm can tip the actor's motivational balance. Economic models often view motivation as a calculus of psychological benefits and costs.[5] From this perspective internalization attaches a guilt penalty to violating a norm, which can change the sign of the net psychological benefits (Casson 1991). Suppose, for example, that after internalizing the norm against appropriation, the agent in figure 1 incurs a cost of 0.7 from violating the norm. The net payoff to the agent from appropriating the investment would fall to 0.3 (figure 2). Note that under this assumption cooperation is the dominant strategy for both players.

Figure 2. Agency Game with Internalization of Social Norm against Appropriation

		Agent: Cooperate	Agent: Appropriate
Principal	Invest	0.5, 0.5	0.3, −1.0
	Don't invest	0, 0	0, 0

Figure 2 depicts an agent who maximizes net benefits. Some moral theorists object to this characterization of motivation. A Kantian decisionmaker, for example, does not aim to maximize anything; thus figure 2 cannot depict Kantian reasoning. Given the lack of agreement on motivation, I will not rely on a particular moral theory. Instead, I rely on the fact that internalizing a norm makes a person more willing to sacrifice to uphold it.

Emergence of Social Norms

To see how social norms arise, consider the principal in figure 1. If the principal believes that the agent is a cooperator, the principal will invest; if the principal believes that the agent is an appropriator, the principal will not invest. Every agent has an incentive to send signals that induce principals to invest since the principal must invest if the agent is to receive a positive payoff. Thus every agent will signal "cooperation," whether the real strategy is cooperation or appropriation.

Recall that cooperation is productive and appropriation is unproductive. People who want to increase the community's wealth will praise cooperation. If everyone in a business community has an incentive to signal conformity with a practice and if conforming to the practice increases the community's wealth, no one in the community will argue publicly against the practice. These two conditions—uniform signaling by individuals and wealth creation for the group—create a consensus that community members ought to conform to a particular practice. The consensus may cause some members of the community to internalize the norm and to inculcate it in young people, so that a new norm emerges in the community.[6] A social norm thus typically emerges when increasing the community's wealth is consistent with individual incentives for signaling.[7]

Changes in the Evolutionary Equilibrium of Social Norms

How will the emergence of a social norm in a business community affect behavior? Typically, the evolutionary equilibrium changes because individuals punish violators of the norm, not because they obey the norm themselves.

People who internalize a norm will cooperate even if the objective payoff (the payoff in terms of wealth and power) for cooperating is slightly lower than the payoff for appropriating. I call this behavior principled conformity to the norm. In contrast, people who do not internalize the norm will cooperate only if the objective payoff for cooperation is at least as high as the payoff for appropriation. I call this behavior adventitious conformity to the norm.

When the objective payoff for cooperation is at least as high as the payoff for appropriation, obeying the norm creates no "strain of commitment" (the term is from Rawls 1971), and an outside observer cannot distinguish between an individual who cooperates adventitiously and one who cooperates from principle. In contrast, a higher objective payoff for appropriation than for cooperation strains commitment and induces only principled conformers to cooperate.

In an evolutionary equilibrium all strategies that persist yield the same objective payoff. In an internal equilibrium some players pursue the strategy of cooperation and others pursue the strategy of appropriation. Equilibrium is reached by adjusting the number of people who conform adventitiously to the norm; the presence of players who conform out of principle does not affect the equilibrium.

To grasp the point, assume that the rate of return for players in the agency game equalizes when eighty players cooperate and twenty players appropriate, and that sixty players cooperate out of principle and twenty players cooperate adventitiously. Now assume that one of the appropriators is persuaded to cooperate. That change in behavior causes a disequilibrium in which eighty-one players cooperate and nineteen players appropriate. Equilibrium will be restored by a change in the strategy of one of the adventitious cooperators from cooperation to appropriation. When equilibrium is restored, eighty players once again cooperate and twenty players appropriate. Thus the internalization of the norm by one more player changes the identity of the cooperators, but not their number.

Marginal players change their strategy when objective payoffs change by a small amount, whereas players who are not on the margin ("inframarginal") persist in their strategy when objective payoffs change by a small amount. In an internal evolutionary equilibrium adventitious conformity is marginal and principled conformity is inframarginal.

Dynamics of Punishment

Agents who have internalized a social norm will bear modest costs to enforce it for the benefit of others through informal sanctions such as gossip, rebukes, and shunning.[8] The internalization of norms thus promotes the sanctioning of wrongdoers by disseminating information about their past acts.

An increase in enforcement by one agent changes the evolutionary equilibrium toward more cooperation. Assume an initial situation in which agents enjoy the same expected payoff from cooperating and appropriating. Now assume that some agents begin using some of their resources to publicize the identities of appropriating agents. The risk of publicity reduces the expected payoff from appropriating, which in turn causes the expected payoff from cooperating to exceed the expected payoff from appropriating. To restore equilibrium, some agents must switch from appropriating to cooperating until the expected payoffs equalize. The willingness of some agents to censure appropriators thus increases aggregate cooperation in equilibrium.

An individual agent bears the cost of censuring appropriators, the gains from which are shared by many players in a competitive situation. Enforcing agents are thus principled, and no adventitious agent enforces.

Does informal enforcement of norms actually take place? Anthropological evidence indicates that enforcement of norms occurs without the support of the state and even in the face of state opposition. In order to sway opinion within close ethnic groups, for example, squatters in Papua New Guinea sometimes hire lawyers to draft real estate "contracts" for buying and selling land that belongs to someone else (Cooter 1991a).

Modeling punishment helps to explain some peculiar features of norms. To begin, I assume that the cost of enforcement includes some risk of harm to the individual who enforces a norm and some expenditure of personal resources. A person who externalizes a norm will not pay the expected cost without receiving an offsetting benefit; a person who internalizes a norm will pay more to enforce it than the benefit received in return.

Even people who have internalized a norm will pay only up to a certain amount to enforce it. As the cost of enforcement rises, fewer people are willing to pay to enforce the norm. Thus a graph of the enforcement function would slope downward to indicate that the expected cost of enforcing, c, must decline in order to attract more enforcers (figure 3).[9]

Next I consider the actual cost of enforcement. Informal punishments (gossip, rebukes, shunning) may require effort, time, or other expenditure, and an informal enforcer may face confrontation or revenge. The risk of confrontation and revenge—and thus the cost of enforcement—tends to fall as the proportion of people willing to punish increases (figure 4). As the proportion of enforcers, E, rises toward the maximum possible value (1), the expected cost of enforcement falls to its minimum value, $\underline{c}$. Conversely, as the proportion of enforcers falls to 0, the expected cost of enforcement rises to its maximum level, $\bar{c}$.

Equilibrium is reached where figure 3, the willingness to pay for enforcement (demand curve), intersects figure 4, the cost of enforcement (supply curve). Intersection of the curves graphed in figures 3 and 4 indicates an equilibrium in the number of enforcers and the cost of enforcement.[10] In other words, if the actual number of enforcers equals the number required to sustain the current cost of enforcement, the cost of enforcement remains constant.

Figure 3. What Enforcers Are Willing to Pay

Cost of an act of enforcement

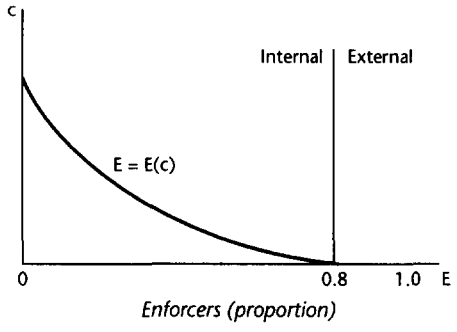
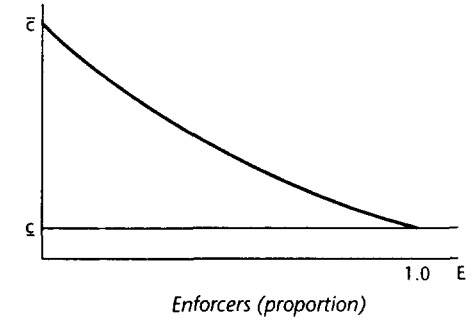


Figure 4. What Enforcement Actually Costs

Cost of an act of enforcement



Graphing several different equilibria helps to explain some observed cases. Consider first a situation in which some proportion of the people who internalize the norm enforce it in others (figure 5). An equilibrium occurs where the two curves intersect (at E^*,c^*). The dynamic behavior of the system is easily explained. If the actual number of enforcers exceeds the number required to sustain the current cost of enforcement, the cost of enforcement will fall. In figure 5 this situation is shown at points where c is more than c^* and E is less than E^*. The arrows in figure 5 indicate the direction of change. Conversely, if the actual number of enforcers falls short of the number required to sustain the current cost of enforcement, the cost of enforcement will rise. In figure 5 this situation is shown at points where c is less than c^* and E is more than E^*, as indicated by the directional arrows. Notice that the directional arrows in figure 5 point toward the intersection of the two curves (at E^*,c^*), indicating that this equilibrium is stable.[11] As explained below, stable equilibria characterize the rule of law state.

Now consider the case in which either everyone who internalizes the norm enforces it or no one enforces it (Taylor 1987; Casson 1991). This situation results in an unstable internal equilibrium (figure 6). Assume that the system begins at the equilibrium (E^*,c^*) and then that a small disturbance moves it downward to the right, where E is greater than E^* and c is less than c^*. Now the actual number of enforcers exceeds the number required to sustain the current cost of enforcement, so the cost of enforcement falls, as indicated by the directional arrow pointing away from the equilibrium (E^*,c^*). Consequently, the system moves even farther from the equilibrium, until the point $0.8,\underline{c}$ is reached, where everyone who internalizes the norm enforces it.

Now consider a move in the opposite direction. Assume that the system begins in equilibrium (at E^*,c^*) and then that a small disturbance moves it upward to the left, where E is less than E^* and c is greater than c^*. Now the actual number of enforcers falls short of the number required to sustain the current cost of enforcement, so the cost of enforcement rises, as indicated by the directional arrow pointing away from E^*,c^*. Consequently, the system moves even farther from the equilibrium, until the point $0, \bar{c}$ is reached, where no one enforces the norm. As explained below, unsta-

Figure 5. Stable Equilibrium

Cost of an act of enforcement

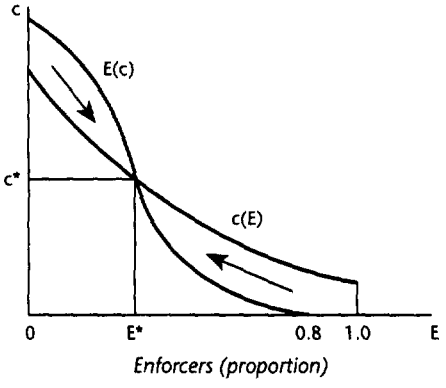
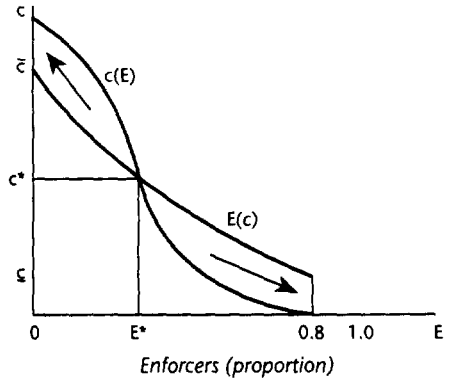


Figure 6. Unstable Equilibrium

Cost of an act of enforcement



ble equilibria characterize the rule of state law. The directional arrows indicate that any slight movement away from the unstable equilibrium (at E^*,c^*) will send the system either to the lower corner, denoted by $0.8,\underline{c}$, or to the upper corner, denoted by $0,\bar{c}$. In the lower corner everyone who internalizes the norm enforces it; in the upper corner no one enforces it.

The chaos that followed the breakup of the Soviet Union can be explained in terms of the model. The Soviet Union exemplified the rule of state law, as depicted by the unstable equilibrium in figure 6. Because state law did not correspond to morality, spontaneous support for law by citizens was weak. Citizens of the former Soviet Union, who were accustomed to a low level of spontaneous support for law, must have expected this tradition to continue after the Soviet government collapsed. These expectations created a self-fulfilling prophecy and caused the system to equilibrate at a level at which private support for state law was low, thereby making state law ineffective.

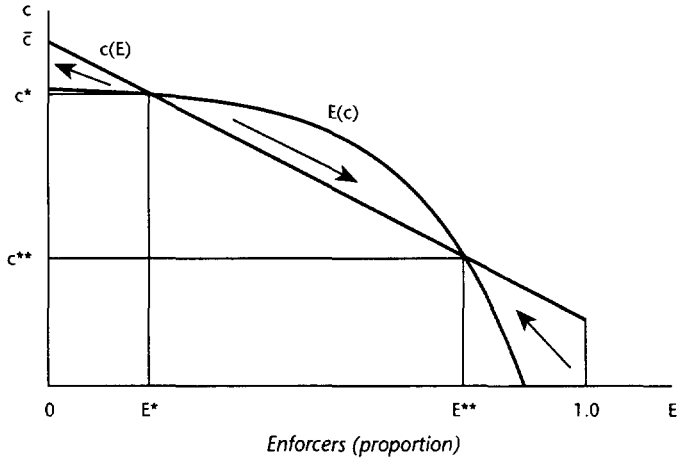
In formal terms, stability occurs when the enforcer's curve, $E(c)$, intersects the cost curve, $c(E)$, from above, as in figure 5; instability occurs when the enforcer's curve intersects the cost curve from below, as in figure 6. Instability is more likely when a small increase in the number of enforcers causes a large decrease in the cost of enforcement, sharply increasing the slope of $c(E)$, and when a small decrease in the cost of enforcement causes a large increase in the number of enforcers, flattening the slope of $E(c)$.

Stability and instability combine in many ways, depending on the shapes of the curves. In figure 7, for example, most people enforce the norm, unless enforcement falls below the tipping value E^*,c^*, in which case few people enforce it. If the system has "tipped out," a policy that causes the system to "tip in" causes a dramatic increase in enforcement up to the level E^{**},c^{**}.

The actual values of E and c at any point in time may be subject to random shocks, which can have a dramatic effect. Assume that the system begins at the point on the vertical axis $0,\bar{c}$, where the cost of enforcement exceeds the amount enforcers

Figure 7. Most Conform or Few Conform

Cost of an act of enforcement



are willing to pay. Now assume that a random shock causes the cost of enforcement to fall below the tipping value E^*,c^*. The system will follow the directional arrows in figure 7 and move to the new equilibrium E^{**},c^{**}, where the cost of enforcement is low and many people enforce the law.

The Rule of Law

The preceding analysis of the costs of informal enforcement provides the foundation for understanding how state law can respond to social norms. According to Locke (1961 [1690]), the state can provide more certain and secure enforcement of social norms. State enforcement is more certain than private enforcement because a written law provides a canonical formulation of the underlying obligation and, in an ideal situation, courts apply the rule with impartiality. State enforcement is more secure than informal enforcement because of the state's monopoly on the official use of force.

Private enforcement and state enforcement typically complement each other. The cooperation of citizens with officials increases the effectiveness of state enforcement and lowers its costs; the backing of state officials increases the effectiveness of private enforcement and lowers its risks. Thus the enactment of a social norm into law and its enforcement by the state shifts the private cost curve $c(E)$ down in figures 5, 6, and 7.

In the stable equilibrium depicted in figure 5, a downward shift in the cost curve pushes the equilibrium to a higher level of private enforcement, as public enforcement "pulls in" more private enforcement. In the unstable situation depicted in figure 6, a downward shift in the cost curve increases the probability that random shocks will cause the system to reach the corner solution characterized by a high level of enforcement rather than the corner solution characterized by a low level of enforcement.

In the complex situation depicted in figure 7, a downward shift in the cost curve has both effects: the stable internal equilibrium shifts to a higher level of private

enforcement and the probability increases that random shocks will cause the system to settle at a high level of enforcement. If most citizens believe that most citizens will enforce the social norm, the system will move to the stable internal equilibrium with a high level of enforcement. But if most citizens believe that few citizens will enforce the social norm, the system will move to the unstable corner equilibrium with a low level of enforcement. Figure 7 thus depicts a self-fulfilling prophecy. In such a social system state enactment can sometimes tip society into conformity with the law merely by causing citizens to believe that most of them will enforce it.

For example, the City of Berkeley, California recently enacted an ordinance requiring owners to clean up after their dogs. Enactment of the law formalized vague social norms. After the law's passage people became more aggressive toward dog owners who ignored the law, suggesting that many people find it easier to enforce a law than a social norm. Enacting the law apparently caused the system to tip from an equilibrium in which there was little private enforcement to a new equilibrium in which many people are willing to enforce the law privately. A similar phenomenon can be observed in the prohibition of smoking in public buildings. Although officials almost never enforce the ban on smoking, signs prohibiting smoking lower the perceived risk to nonsmokers who complain to smokers. (For an interesting collection of studies on public policy toward smoking, see Rabin and Sugarman 1993.)

Consider now what happens when the state imposes laws that contradict social norms. For example, assume that the state punishes people who privately enforce a social norm, which causes the cost curve $c(E)$ to shift up. In figure 5 an upward shift in the cost curve moves the equilibrium to a lower level of private enforcement. In the unstable situation depicted in figure 6 an upward shift in the cost curve increases the probability that random shocks will cause the system to reach the corner solution characterized by a low level of enforcement. In the complex situation depicted in figure 7 both effects operate. In general, the state can undermine obedience to social norms by raising private enforcement costs.

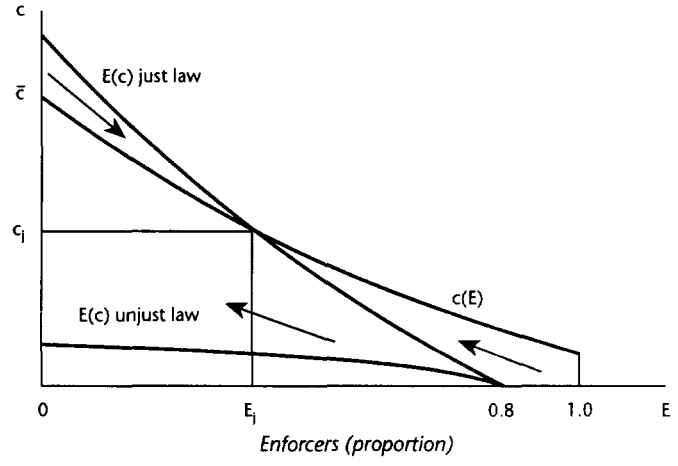
Corruption

Laws that are consistent with social norms appear to be relevant and just to many citizens, who help the state enforce these laws. State enforcement of law, including detecting and punishing officials who accept bribes in exchange for breaking the law, is more effective when supplemented by informal enforcement, as shown in figure 8. Recall that the willingness of private citizens to bear high enforcement costs makes the $E(c)$ curve relatively steep. A relatively steep $E(c)$ curve like the one labeled "just law" in figure 8 stabilizes the system by reaching an equilibrium at E_j, c_j. Private enforcement supplements public enforcement, and the system settles into an equilibrium with effective deterrence of bribes.

Conversely, laws that are inconsistent with social norms are perceived as irrelevant or unjust by many citizens, who are reluctant to help the state detect and punish officials who corrupt such laws. This situation is depicted in figure 8 by the curve labeled "unjust law." Recall that the unwillingness of private citizens to bear high enforce-

Figure 8. Corruption

Cost of an act of enforcement



ment costs makes the $E(c)$ curve low and relatively flat. A low, relatively flat $E(c)$ curve destabilizes the system and creates a corner solution at 0, $\bar{c}$ in which private enforcement is minimal. With minimal private enforcement, corruption flourishes because state officials lack the information required for effective public enforcement.

Suppressing corruption requires help from private citizens, which in turn depends on the responsiveness of the law to social norms. A bribe is the price paid illegally to an official for a service, such as overlooking a regulation or awarding a contract.[12] The willingness of citizens to suppress the corruption of a state regulation depends on whether a social norm will support regulation. I have argued that social norms typically arise in a business community when the public good aligns with individual incentives for signaling. If a regulation reduces external costs, the community would benefit from a social norm supporting the regulation. If most individuals have an incentive to signal conformity with such a norm, the norm will probably emerge. For example, if sound accounting practices could benefit government contractors as a group, and if each contractor with the government wants to proclaim fidelity to sound accounting practices, the business community may develop a norm requiring sound accounting practices. If the social norm emerges, then a law aligned with it will appear to be relevant and just to most people. Consequently, citizens will cooperate in enforcing the law and punishing officials who corrupt the law by accepting bribes. Private enforcement will equilibrate at a high level, as depicted in figure 8, which will make public enforcement cheap and effective.

Now consider the situation in which the state prohibits an activity that creates external benefits. The community will not develop a social norm to support such a regulation. Regulations that restrict competition for the benefit of politically favored groups, for example, will seem unjust to many people, so few people will run risks to help officials enforce such laws. Private enforcement will, consequently, equilibrate at a low level, as depicted in figure 8 for the unjust law. The absence of pri-

vate enforcement will make public enforcement expensive and ineffective, and many officials will take bribes in exchange for not enforcing the prohibition.

The suppression of corruption requires that private citizens believe correctly that a large proportion of them will bear modest costs to punish corruption. The proliferation of regulations inhibits this process by sharply increasing the risk to whistle blowers. The risk increases sharply because prolix regulations require everyone to break the law from time to time. When everyone breaks the law, anyone who reports someone else is vulnerable to counter charges. Thus the proliferation of regulations sharply increases the likelihood that the system will settle in an equilibrium in which few private citizens punish corruption.

Even if private enforcement costs are low, regulations that inhibit competition are unlikely to command much respect. For example, pious legislators from India to Italy have burdened the employment contract with compulsory benefits. In many circumstances not complying with such rules benefits employers and employees without harming anyone else, and black market and gray market jobs that circumvent these regulations flourish. Suppressing these markets is especially difficult because so many people see nothing harmful in them.

Every legal obligation provides a potential source of bribes to officials who enforce it. The more highly regulated is the economy, the greater is the scope for bribes. Thus economies that are just emerging from central planning provide many opportunities for officials to extract bribes. Ideally, inefficient regulations should be repealed, leaving only efficient regulations in place. Public officials, however, show little enthusiasm for repealing laws that provide them with illegal income.

Broad taxes raise more revenue at less cost to the economy than do narrow taxes. In effect, bribes are very narrow taxes assessed against a particular beneficiary of a state service. Consequently, public officials as a whole can obtain higher incomes by honestly administering general taxes than by extracting bribes for particular services. Although officials as a group would benefit from replacing bribes with regular salaries paid out of general taxes, each individual official may benefit from taking bribes. This collective action problem explains why most officials who take bribes in corrupt countries receive low incomes.

Sources of Law for Development

This theory of social norms can be applied to the sources of law for economic development. First I discuss the sources of legal modernization in the West, and then I move on to the problem of the emergence of law from chaos and the stabilization of democracy.

Legal Modernization in the West

Legal modernization in the West began in the eighteenth century, when developments in science and industry greatly accelerated the pace of social change. These social changes prompted rapid changes in law. The law had to recognize the corpo-

rate form of organization, clarify the meaning of "property" for industrial organizations, extend contracts to new financial instruments, develop patent protection for new types of inventions, extend accident law to dangers posed by new technology, allocate losses from bankruptcies in new types of organizations, and develop regulations to protect the environment from new pollutants. I describe four processes by which legal modernization occurred in the West.

COMMON LAW. English common law has a history dating to the medieval period. Throughout most of its history English common law developed as a close-knit community of professional lawyers and judges ruled on cases. According to the legal historian Brian Simpson (1973, 1975), the common law has been more of an institution than a collection of rules. Not until the eighteenth century did Blackstone first organize the leading cases of the common law into modern legal categories.

The common law had to evolve quickly in the eighteenth century to keep pace with changes in business. The use of notes and bills of exchange, which circulated among eighteenth-century merchants as a means of payment and credit, raised new questions about risk allocation. Such questions became acute with the rapid expansion of commerce in the eighteenth century. Lord Mansfield, one of the architects of eighteenth-century common law, recognized that he did not fully understand how businesses used financial instruments. He conducted careful inquiries, scrutinized businesses, and tried to identify and enforce the best financial practices.[13] His elegant solutions continued to be taught in courses on commercial law long after the relevant financial instruments ceased circulating.

In common law systems, intensive litigation alerts judges to the need to change the law. Judges respond to a proliferation of novel disputes by making new law (Eisenberg 1988). Not only do disputes cause precedents, but new precedents may cause disputes (Priest 1987; Cooter 1987).

According to an old principle in jurisprudence, judges must discover common law in social norms rather than invent law in light of their own preferences.[14] To the extent that judges follow this strategy, the common law reflects social norms. Since the eighteenth century new institutions have developed in common law countries to advance this process. Organizations conduct studies to scrutinize current law and issue reports recommending changes to it. In the United Kingdom law commissions perform these tasks. In the United States these tasks are handled by the American Laws Institute (ALI) and the National Commission on Uniform State Laws (NCUSL), which try to keep the law current with developments in business communities (Ogus 1995; Schwartz and Scott 1995).

Although judges' decisions seldom cite economic efficiency, analysis of common law rules reveals that common law is economically efficient (Posner 1979). Since the law and economics first demonstrated this fact, scholars have tried to uncover the "hidden hand" that directs the common law toward efficiency (Rubin 1977; Cooter and Kornhauser 1980). The most compelling explanation is that social norms evolve toward efficiency and that common law promotes efficiency by reflecting social norms.

CIVIL LAW. As early as the eighteenth century legal scholars in England were debating whether common law was appropriate to modern life or an archaic compendium of obsolete practices (Posner 1979). In continental Europe common law was identified with the losing side in the revolutions that brought Napoleon and his followers to power. The victorious revolutionaries, who regarded judges with suspicion for having upheld the old regime, wanted to root out "medieval" practices and replace them with "rational" ones (Dawson 1972). In their view law derives its authority from the popular will as expressed through legislators, not from social norms "uncovered" by judges. The popular will was identified with rationality, whereas social norms were identified with habits. Commissions were appointed to draft codes to supersede the common law. Scholars on the commissions examined pre-Revolutionary law with a critical eye, retaining some parts of it and rejecting the rest (David and Brierley 1985; Dawson 1972; Merryman 1985; and Zweigert and Kotz 1992). Legislators enacted the codes into law.

In theory, judges in civil systems interpret law but do not make it. Codifiers in civil law countries thus appear to have more influence and judges less influence than in common law countries. Whether the apparent difference between the two systems is real or illusory is difficult to resolve because neither system exists as a pure type. In common law countries restatements and codes have legal authority; in civil law countries judges are influenced by social norms. (For alternative views, see Kirchner 1991; Mattei and Pardolesi 1991; and Cooter and Gordley 1991.).

Interpreting some codes, however, looks a lot like uncovering social norms. Both common law and civil codes rely heavily on broad principles that apply in many different circumstances. Ideally, these principles are derived from practices, and the practices give specific content to the principles. The common law of torts, for example, typically holds injurers liable for accidents caused by their negligence; this general principle receives specific content from the actual standards by which particular communities evaluate accidents. When judges apply the negligence principle, they often find the specific standard applicable to the case by identifying the best practices in the relevant community. Civil law judges can proceed on similar lines when interpreting general principles in a code. By relying on judges to use specific practices to interpret general principles, both systems of law empower judges to enforce social norms.

Some codes are based directly on social norms. In drafting the Uniform Commercial Code (UCC) of the United States, which applies to bankruptcy, financial instruments, and contracts between merchants, Professor Karl Llewellyn and his staff tried to identify and articulate the best commercial practices in contemporary business communities, much as Lord Mansfield did when he modernized British commercial law (Hillinger 1985). After Llewellyn's committees completed their work, the UCC was presented to the state legislatures, which enacted it into law. Judges continue to make commercial law in the United States by interpreting the UCC.

Before Napoleon came to power, countries in continental Europe shared a kind of common law called the *ius commune*. Scholars and judges developed the *ius*

commune by interpreting Roman law in light of local traditions and contemporary realities. Some European legal scholars today hope to unify European private law by building on the *ius commune* (Zimmermann 1990).

THE REGULATORY STATE. Some economic and social problems require public policies, not just social norms. For example, a modern economy cannot spontaneously generate effective norms restraining environmental degradation by competitive industries; social norms cannot ensure adequate expenditures on public goods, health programs, or basic scientific research. In the twentieth century countries in Europe and North America addressed these problems by enacting new laws, thereby creating the regulatory state.

A community of scholars, lawyers, and judges typically produces the common law and much of the civil law. In contrast, politicians and administrators have greater influence on regulations. Regulations can take the form of statutes enacted by the legislature, orders issued by the executive, or rules promulgated by ministries. Ministries usually follow procedures prescribed in legislation. Thus when U.S. agencies create new regulations, they must follow procedures stipulated in the legislation conveying authority to them; in the absence of such stipulations, they must follow procedures prescribed in the Administrative Procedures Act.

Because the common law reflects social norms, it can articulate general principles whose specific content comes from community practices. In contrast, regulations often lack a foundation in social norms. Regulators would thus create uncertainty by promulgating general principles rather than detailed instructions. Accordingly, regulations tend to be more specific and detailed than is common law or codes.

Legal principles dictate that legislation trumps the common law whenever the two conflict. When public law crowds out private law, the state takes control of resource allocation.

COMMUNISM. The most complete legal reforms in the twentieth century were carried out by communist revolutionaries, who replaced the law, politics, and economics of the old regimes with central planning. Central planning is a way of making law and goods that combines state planning of the economy and political dictatorship. In communist countries the state repealed or emasculated private law on employment relations, land ownership, antitrust, consumer product liability, and worker safety. Once the legal impediments were removed, officials ruled by decree. Under central planning, government officials formulated the state's goals for the production of commodities, embodied the goals in production targets, and ordered people to meet them. Central planning produced remarkably similar results in vastly different countries, such as Cuba, the Democratic People's Republic of Korea, and Poland, consistently failing to produce consumer goods in abundant quantity or of high quality.

Some formerly communist countries are trying to reinstate lost legal traditions. With funding and prodding from international and foreign agencies, national and foreign scholars are drafting new laws for the formerly communist countries. In Eastern Europe it seems that each country is trying to recreate private law based on

its precommunist codes and on the law of the country that is funding its legal reform (Cadwell 1995; Stephan 1995).

CONSENSUS OR MAJORITY RULE. Four sources of economic law in the West have been described: common law, civil law, the regulatory state, and communism. Ideally, common law responds to social norms, as required by the rule of law. In contrast, the regulatory state imposes rules that depart from social norms. Communism takes the regulatory state to its logical extreme by adopting central planning. Civil law codes can either enforce social norms or impose obligations that are inconsistent with social norms.

A social norm exists when it is effective in a community; it is effective when a sufficient number of community members enforce it. The emergence of a social norm usually requires consensus in the community that people ought to obey the norm. This consensus provides the justification for coercion when the state enforces social norms. In contrast, the existence of a regulation depends on its enactment or promulgation. In a democracy the legislature enacts regulations and the executive branch promulgates them. Democratic elections typically follow the principle of majority rule, which provides the justification for coercion when the state enforces regulations.

Arrow's impossibility theorem and its extensions created skepticism about the ability of majorities to produce coherent laws (Arrow 1963; McKelvey 1979). Without coherence the law cannot be efficient. Empirical research has shown that in democratic states the politics of redistribution often dominates the economics of production (Stigler 1972). Rent seeking is so endemic in democracies that one school of thought holds that companies in regulated industries typically "capture" their regulator and use its powers to restrict competition (Stigler 1975; Kolko 1967). Recent attempts by scholars to rehabilitate majority rule have had limited success (Wittman 1989; Waldron 1993).

Some conservative law and economics scholars believe that there is a stronger case for selectively enforcing social norms than for making law by majorities. The case for law as a selective enforcer of social norms is strongest when applied to business norms that evolve in a system of open competition. The case is weakest when applied to business norms that evolve under noncompetitive conditions; to public goods, such as pollution; to issues of critical morality, such as the suppression of women; or to "evolutionary traps," such as blood feuds. In these situations social norms may be oppressive or missing. Some progressive law and economics scholars remain optimistic about majority rule and regulation. The most optimistic of these progressive theories of democracy rest on such ideas as civic virtue and political discourse, which few economists analyze and some dismiss.

Usurpation, Chaos, and the Emergence of Stable Property Rights

Having discussed sources of law in stable states, I now consider the emergence and stabilization of democracy. Democracy is a system of competition for control of the

state, which has a monopoly on the official use of force. In many countries the state's monopoly on official force is abused, and the government periodically breaks the law or suspends the constitution. Such an assertion of illegal powers by state officials can be deterred, however, and stable power can be created out of chaos.

Many citizens who believe in democracy will monitor officials and censure violations of the constitution, even at some cost to themselves. Recall that the willingness of private citizens to bear high enforcement costs makes the $E(c)$ curve relatively steep, which tends to make the system stable. Another factor, however, destabilizes constitutional monitoring by citizens. The tyrant who successfully usurps power punishes resistance harshly. Thus citizens' willingness to resist depends on their confidence of success, since citizens who resist unsuccessfully pay a heavy price. Recall that when a small decline in the proportion of enforcers causes a sharp increase in enforcement costs, the $c(E)$ curve is relatively steep, which tends to make the system unstable.

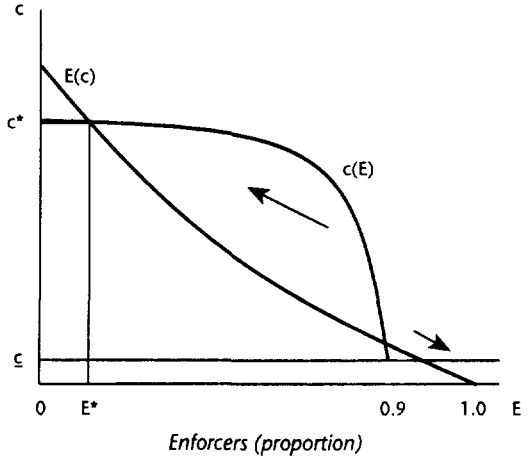
Figure 9 captures the phenomenon of resistance to tyrants. Either 90 percent of private citizens resist tyranny and resistance succeeds, in which case they each pay the minimum price $\underline{c}$ and the system equilibrates at 0.9, $\underline{c}$, or few private citizens resist tyranny and resistance fails, in which case resistors pay the heavy price, c^*, and the system equilibrates at E^*, c^*.[15]

To overcome this instability, private citizens need to resist tyranny collectively. Two such sources of protection—the separation of powers in the constitution and the fragmentation of political parties—are examined here.

Antitrust theorists sometimes quip that the only interesting numbers are one, two, three, and four, suggesting that four or more competing firms achieve much the same results as perfect competition. The U.S. Constitution works much like antitrust law in that it divides government into four institutions—the executive branch, the House of Representatives, the Senate, and the Supreme Court—that

Figure 9. Usurpation

Cost of an act of enforcement



compete for state power. The threat of usurpation comes from the president, who directs the armed forces and the bureaucracy. The armed forces and the bureaucracy, however, would be reluctant to obey presidential commands that the Supreme Court declared illegal and that the two houses of Congress opposed; in order to usurp power, the president would probably need cooperation from the other three powers.

Constitutions that separate powers, however, do not prohibit a single political party from controlling more than one branch of government. A single dominant party may unify powers separated by the constitution. Empirical research comparing different countries shows that a dominant disciplined party in government reduces the power of the courts, whereas competition for office among political parties preserves the separation of powers (Cooter and Ginsburg 1996; Ramseyer 1994).

According to Duverger's law (Cooter forthcoming), two-party competition tends to emerge from single-district elections for the legislature that follow a winner-take-all rule, while proportional representation tends to result in multiparty systems. In some countries proportional representation results in stable, competitive government by coalition; elsewhere it causes legislative paralysis and, sometimes, government by executive decree.

ENCOMPASSING INTEREST. How do stable property rights emerge from chaos? Olson (1993) considers a situation in which roving bandits ravage a country. Eventually, one bandit dominates the others and expels them from the territory. The stationary bandit asserts an exclusive right to exploit the territory for economic gain, sometimes draping this power in legitimacy through self-declaration of divine kingship. Olson shows that rather than expropriate property, the rational stationary bandit will impose general taxes, because general taxes will raise far more revenue than can be obtained by expropriating the property of individuals. A rational stationary bandit will abandon expropriation, guarantee property rights, and set taxes at the level that maximizes state revenue.

Olson explains the difference between roving bandits and a stationary bandit in terms of the more "encompassing interest of the stationary bandit." Roving bandits maximize their wealth at a particular point in time; their interests do not encompass the future wealth of the territory. The stationary bandit strives to maximize long-run wealth; the bandit's interests thus encompass the future tax revenues yielded by the territory. The interests of a secure monarch in a family dynasty correspond to those of a stationary bandit. The interests of dictators, who are not usually able to pass on power to their children and who are often deposed, correspond to those of roving bandits.

Olson extends the argument to the transition from banditry to democracy. The stationary bandit sets the tax rate to maximize the state's revenue. In contrast, a democratic government ideally sets the tax rate to maximize the nation's wealth. Because the tax rate that maximizes the nation's wealth is lower than the tax rate that maximizes the state's revenue (Cooter and Helpman 1974), a change from

banditry to democracy results in lower taxes. In general, the more encompassing the state's interest, the closer it comes to maximizing the nation's wealth, which implies lower taxes. In practice, the government's interest does not encompass the wealth of the nation as a whole because special interests enjoy disproportionate influence in democracies. Nevertheless, Olson believes that a democratic government has a more encompassing interest than a stationary bandit.

INTERMEDIATE INSTITUTIONS. How does limited government arise from absolutism? The stationary bandit may initiate the change to increase state revenue. A stationary bandit can increase the state's income by guaranteeing property rights against appropriation and imposing general taxes. But an absolute power cannot guarantee anything against itself. To maximize state revenue, state power must be limited by guaranteeing property rights.

Besides taxes, the state obtains revenue by borrowing. North (1995) has observed that the borrowing ability of English kings increased greatly in the seventeenth century, when the rising power of Parliament provided some guarantee of repayment. This difference in borrowing ability tipped the balance in wars against Continental absolutists. In general, negotiating loans or bargaining sucessfully requires credibility, which presupposes limits on future action. Constitutional division of powers limits future action. Even without the constitutional division of power, however, political power may need to bargain with economic powers. As industries differentiate and grow, private centers of wealth emerge, whose support the government may need. Limited government may emerge—as it has in the rapidly industrializing economies of East Asia—because economic development requires the state to bargain with many distinct industries.

Private centers of wealth do not rely on spontaneous cooperation to protect their interests. Instead, they organize into intermediate institutions that establish rules by which members govern themselves. Do such institutions inhibit competition or facilitate exchange? Olson (1993) claims that as a society ages and the number of intermediate institutions increases, such institutions seek political rents with increasing effectiveness, until sclerosis clogs the arteries of economic competition. In contrast, Bernstein (1992, 1995) found that associations of commodity dealers (diamonds, grain, textiles) lower the transactions costs of wholesale buying and selling, thereby facilitating trade and cooperation by enforcing social norms.

In practice, an organization that seeks to maximize the wealth of its members will behave both monopolistically and efficiently. By fixing prices, establishing territories, and withholding information from the public, the organization will seek to create monopoly power for its members in dealing with nonmembers. The organization will also seek to minimize agency costs that members incur in dealing with each other. To minimize agency costs, the organization must establish efficient property rights and contracts among members. Thus a general principle of motivation for an intermediate organization can be stated as: monopoly for outsiders, efficiency for insiders. In the worst situation organizations form tight cartels to extract monopoly prices in local markets. In the best situation competition deprives an organization of

monopoly power over outsiders, thus limiting the role of the organization to promoting efficiency among insiders. Free trade makes business organizations virtuous in spite of themselves.

Conclusion

Social norms arise in a group when individuals benefit from representing themselves as conforming to a practice that benefits other members. In an environment of open competition business norms tend to evolve toward efficiency, and competition causes changes to occur rapidly. As an economy develops, constraints of information and motivation tighten on officials who make and enforce state law. To loosen these constraints, state law must be decentralized and modernized by reinforcing business practices rather than dictating what businesses do. Bottom-up lawmaking requires a community of judges, lawyers, and scholars to identify social norms, state them authoritatively, and enforce them selectively.

Sustaining a social norm requires that it be enforced by private individuals. Increasing returns to scale of private enforcement can destabilize the system, causing a society to "tip in" to a high level of private support for state law or "tip out" to a low level of private support. A high level of private support is more likely when the law is consistent with social norms, so that most citizens obey the law out of respect. A low level of private support is more likely when law contradicts morality or seems irrelevant. Laws that convey monopoly profits to politically favored groups by restricting competition seldom command respect.

This analysis can be applied to corruption. To suppress corruption, private citizens must believe correctly that many of them will bear modest costs to help punish corrupt officials. The cost to a private citizen of helping to punish corrupt officials decreases as the number of citizens who enforce the law increases. Consequently, a society may tip in to a high level of enforcement, or tip out to a low level of enforcement. As a group state officials could benefit by eliminating corruption and replacing their loss in bribes with higher salaries paid from general taxes. The faction with the most encompassing interest may oppose the corruption of more parochial officials.

The development of industry in many different sectors creates multiple centers of economic power with which the government may need to negotiate. To increase the credibility of its promises, government may limit its power by separating powers in the constitution and respecting property rights. Economic development thus provides government with an incentive to limit its powers.

The model presented here has implications for international organizations such as the World Bank that help countries develop legal infrastructure.[16] The rule of law cannot come from top-down planning because central officials lack both sufficient information and the right motivation. Technical assistance should not focus primarily on establishing new codes and regulations. Instead technical assistance should help establish intermediate institutions and a community of judges, lawyers, and scholars that can shape the law so that it conforms to reality.

Notes

1. For evidence on the relationship between obedience and perceived justice, see Tyler (1990).

2. Fudenberg and Maskin (1986) have proved that in any game in which players maximize the discounted sum of single-period utilities, the discount rate is not too high, and the players can observe the history of moves in the game, any pair of payoffs that Pareto dominate the minimax can arise as average equilibrium payoffs of the repeated game. Thus repetition of the game makes a Pareto improvement possible. This theorem still leaves unexplained why the probability of a Pareto efficient solution is as high as empirical studies suggest it to be.

3. For an excellent review of these developments, see Bannerjee and Weibull (1993). For a discussion of the relationship between law and evolutionary theory, see Elliott (1985) and Elliott, Ackerman, and Millian (1985). For a pioneering article on evolutionary models of law, see Hirshleifer (1987); for a pioneering book on evolutionary models of economics, see Nelson and Winter (1982).

4. As the proportion of appropriating agents increases, more partnerships dissolve more often. Some of the principals released from these relationships look for new partners. Consequently, the release of principals from existing partnerships tends to lower the expected cost of a successful search by an agent for a partner. Another force works in the opposite direction, however. As the number of appropriators increases, investment becomes less profitable and some principals withdraw from the industry. Withdrawal of principals from the industry increases the expected cost of a successful search by an agent for a partner. On balance, the presence of more appropriators causes search costs to increase when the withdrawal effect dominates the release effect. Thus a stable equilibrium usually exists when the withdrawal effect dominates the release effect. For a similar formal model, see Schussler (1993).

5. Antiutilitarian philosophers typically reject the theory that conforming to a principle of morality involves weighing alternative reasons and balancing them. See, for example, the account of "exclusionary reasons" in Raz (1986) and Kant (1948 [1785]).

6. Arguing along similar lines, Pettit (1990) claims that norms will be "resilient" when nearly everyone approves of those that benefit others and disapproves of those that harm others.

7. For a similar account of the emergence of norms, expressed in the language of philosophy, see Gibbard (1990).

8. In discussing the problem of sanctioning wrongdoers through gossip, Pettit (1990) writes, "People do not have to identify violators intentionally; they just have to be around in sufficient numbers to make it likely that violators will be noticed. And equally, people do not have to discipline violators intentionally, going out of their way for example to rebuke them or report them to others; they just have to disapprove of them—or at least be assumed to disapprove of them—whether that attitude ever issues in intentional activity" (p. 739). Pettit's argument is based on the assumption that people are moved by a concern that others not think badly of them. For a more pessimistic assessment of informal sanction, see Heckathorn (1989). For a discussion of how overenforcement might arise from the interdependence of enforcement actions by private property owners, see de Meza and Gould (1992). For theories of ostracism, see Gruter and Masters (1986).

9. Here is a strict definition of terms for the enforcement function $E(c)$: the density function $f(s)$ over willingness to pay to enforce the social norm

$E = 1 - \int_0^c f(s)$.

10. To be precise, an equilibrium is a pair of values E^*, c^* such that $E^*=E(c^*)$ and $c^*=c(E^*)$.

11. If $E(c)$ cuts $c(E)$ from below, the equilibrium is stable. If $E(c)$ cuts $c(E)$ from above, the equilibrium is unstable.

12. Either the official act is illegal, or the official act is legal and accepting payment for it is illegal. Bardhan (1996) provides an excellent review of the economic literature on corruption.

13. The process of assimilating bills of exchange and negotiable instruments into common law is well documented. The traditional theory is developed in Holden (1955). Holden is criticized in Baker (1979). A revised view, which stresses that Mansfield immersed himself in the minutiae of business practice in order to extract the best principles, is found in Rogers (forthcoming). I benefited from discussions on this point with Professors Dan Coquillette, James Gordley, and Jim Rogers.

14. For an exposition of this old view of lawmaking, see Davies (1986). Rubin (1995) cites Chodorow (1972), Fichtenau (1991), and Lewis (1954) in tracing this line of thought to a belief in medieval Europe that law is both divine and natural. This older view is reflected in the jurisprudence of Dworkin (1977), who asserts that courts should find rights and not make policy. The view that judges consult their own values to fill gaps in the law is developed by Hart (1961).

15. Weingast (1996) uses a two-person prisoners' dilemma game to reach similar conclusions about the instability of resistance to tyranny.

16. For a list of loans and lessons drawn from them, see World Bank Legal Department (1995).

References

Anderson, Terry, and P. J. Hill. 1975. "The Evolution of Property Rights: A Study of the American West." *Journal of Law and Economics* 18 (1): 163–80.

Arrow, Kenneth J. 1963. *Social Choice and Individual Values,* 2nd ed. New York: Wiley.

Axelrod, Robert. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Baker, John. 1979. "The Law Merchant and the Common Law before 1700." *Cambridge Law Journal* 38: 295.

Bannerjee, Abhijit, and Jorgen W. Weibull. 1993. "Evolution and Rationality: Some Recent Game-Theoretic Results." Research Papers in Economics. University of Stockholm, Department of Economics.

Bardhan, Pranab. 1996. "The Economics of Corruption in Less Developed Countries: A Review of Issues." Department of Economics Working Paper and Center on International Development Economics Research, University of California, Berkeley.

Bernstein, Lisa. 1992. "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry." *Journal of Legal Studies* 21: 115–57.

———. 1995. "The Newest Law Merchant: Private Commercial Law in the United States." Paper presented at the annual meeting of the American Law and Economics Association, May, Berkeley, Calif.

Besley, Timothy, and John Mclaren. 1993. "Taxes and Bribery." *Economic Journal* 103: 119–41.

Cadwell, Charles. 1995. "Implementing Legal Reform in Transition Economies." Paper presented at the annual meeting of the American Law and Economics Association, May, Berkeley, Calif.

Casson, Mark. 1991. *The Economics of Business Culture: Game Theory, Transaction Costs, and Economic Performance.* Oxford: Clarendon Press.

Chodorow, Stanley. 1972. *Christian Political Theory and Church Politics in the Mid-Twelfth Century.* The Ecclesiology of Gratian's Decretum. Berkeley, Calif: University of California Press.

Cooter, Robert. 1987. "Why Litigants Disagree: A Comment on George Priest's 'Measuring Legal Change.'" *Journal of Law, Economics, and Organization* 3 (85): 227–34.

———. 1991. "Collective Choice Theory: A Review." John M. Olin Working Paper in Law and Economics 91-4. School of Law, University of California at Berkeley.

———. 1994. "Structural Adjudication and the New Law Merchant: A Model of Decentralized Law." *International Review of Law and Economics* 14: 215–31.

———. 1996. "Decentralized Law for a Complex Economy: The Structural Approach to Adjudicating the New Law Merchant." *University of Pennsylvania Law Review* 144: 1643–96.

———. Forthcoming. *Foundations of Constitutional Law and Economics.* Princeton, N.J.: Princeton University Press.

Cooter, Robert, and Tom Ginsburg. 1996. "Comparative Judicial Discretion: An Empirical Test of Economic Models." *International Review of Law and Economics* 16: 295–314.

Cooter, Robert, and James Gordley. 1991. "Economic Analysis in Civil Law Countries: Past, Present, Future—Introduction to a Symposium." *International Review of Law and Economics* 11: 261.

Cooter, Robert, and Elhanan Helpman. 1974. "Optimal Income Taxation for Transfer Payments under Different Social Welfare Functions." *Quarterly Journal of Economics* 88: 656–70.

Cooter, Robert, and Lewis Kornhauser. 1980. "Can Litigation Improve the Law without the Help of Judges?" *Journal of Legal Studies* 9: 139.

David, Rene, and John E. C. Brierley. 1985. *Major Legal Systems in the World Today.* London: Stevens and Sons.

Davies, Sir John. 1986. "Le Primer Report des Casea et Matters en Ley Resolues et Adiudges en les Courts del Roy en Ireland." In D. Wootton, ed., *Divine Right and Democracy: An Anthology of Political Writing in Stuart England.* London: Penguin.

Dawes, Robyn, and John Orbell. 1993. "Social Welfare, Cooperators' Advantages, and the Option of Not Playing the Game." Paper presented at the fourth international conference on social justice research, July 1–3, Trier, Germany.

Dawson, Philip. 1972. *Provincial Magistrates and Revolutionary Politics in France, 1789–1795.* Cambridge, Mass.: Harvard University Press.

de Meza, David, and J. R. Gould. 1992. "The Social Efficiency of Private Decisions to Enforce Property Rights." *Journal of Political Economy* 100: 561–80.

Dworkin, Ronald. 1977. "The Model of Rules I." In Ronald Dworkin, ed., *Taking Rights Seriously.* London: Ductsworth.

Eisenberg, Melvin A. 1988. *The Nature of the Common Law.* Cambridge, Mass: Harvard University Press.

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes.* Cambridge, Mass: Harvard University Press.

Elliott, E. Donald. 1985. "The Evolutionary Tradition in Jurisprudence." *Columbia Law Review* 85: 38.

Elliott, E. Donald, Bruce A. Ackerman, and John C. Millian. 1985. "Toward a Theory of Statutory Evolution: The Federalization of Environmental Law." *Journal of Law, Economics, and Organization* 1: 313–40.

Fichtenau, Heinrich. 1991. *Living in the Tenth Century.* Chicago: University of Chicago Press.

Fudenberg, Drew, and Eric Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* 54: 533–54.

Fuller, Lon L. 1958. "Positivism and Fidelity to Law—A Reply to Professor Hart." *Harvard Law Review* 71: 630.

———. 1964. "The Morality That Makes Law Possible." In Lon L. Fuller, ed., *The Morality of Law.* New Haven, Conn.: Yale University Press.

Gibbard, Allan. 1990. "Norms, Discussion, and Ritual: Evolutionary Puzzles." *Ethics* 100: 787–802.

Gruter, Margaret, and Roger D. Masters, eds. 1986. *Ostracism: A Social and Biological Phenomenon.* New York: Elsevier.

Harris, Marvin. 1968. *The Rise of Anthropological Theory: A History of Theories of Culture.* New York: Crowell.

Harrison, Glenn W., and Jack Hirshleifer. 1989. "An Experimental Evaluation of Weakest Link/Best Shot Models of Public Goods." *Journal of Political Economy* 97: 201–25.

Hart, H. L. A. 1958. "Positivism and the Separation of Law and Morals." *Harvard Law Review* 71: 593.

———. 1961. *The Concept of Law.* Oxford: Clarendon Press.

Heckathorn, Douglas D. 1989. "Collective Action and the Second-Order Free-Rider Problem." *Rationality and Society* 1: 78–101.

Hillinger, Ingrid Michelsen. 1985. "The Article 2 Merchant Rules: Karl Llewellyn's Attempt to Achieve the Good, the True, and the Beautiful in Commercial Law." *Georgetown Law Journal* 73: 1141–84.

Hirshleifer, Jack. 1987. "Evolutionary Models in Law and Economics." In Jack Hirshleifer, ed., *Economic Behavior in Adversity.* Brighton, Sussex: Wheatsheaf Books.

Hirshleifer, Jack, and Juan Carlos Martinez Coll. 1988. "What Strategies Can Support the Evolutionary Emergence of Cooperation?" *Journal of Conflict Resolution* 32: 367–98.

Holden, James M. 1955. *Early History of Negotiable Instruments.* London: University of London Press.

Kant, Immanuel. 1785. "Foundations of the Metaphysics of Morals (Grundlegung zur Metaphysik der Sitten)." In Herbert J. Paton, trans., *The Moral Law.* London: Hutchinson University Library, 1948.

Kirchner, Christian. 1991. "The Difficult Reception of Law and Economics in Germany." *International Review of Law and Economics* 11: 277.

Kolko, Gabriel. 1967. *The Triumph of Conservatism: A Reinterpretation.* Chicago: Quadrangle Press.

Lewis, Ewart. 1954. *Medieval Political Ideas.* New York: Knopf.

Locke, John. 1961 [1690]. *The Second Treatise of Civil Government.* Edited by Thomas I. Cook. New York: Hafner.

Lyons, Nona. 1983. "Two Perspectives: On Self, Relationships, and Morality." *Harvard Law Review* 53: 125–45.

Macaulay, Stewart. 1963. "Non-Contractual Relations in Business: A Preliminary Study." *American Sociological Review* 28: 55.

Mattei, Ugo, and Roberto Pardolesi. 1991. "Law and Economics in Civil Law Countries: A Comparative Approach." *International Review of Law and Economics* 11: 265.

McKelvey, R. D. 1979. "General Conditions for Global Intransitivities in Formal Voting Models." *Econometrica* 47: 1085–112.

Merryman, John Henry. 1985. *The Civil Law Tradition: An Introduction to the Legal Systems of Western Europe and Latin America.* Stanford, Calif.: Stanford University Press.

Mookherjee, Dilip, and Ivan Png. 1995. "Corruptible Law Enforcers." *Economic Journal* 105: 145.

Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change.* Cambridge, Mass.: Harvard University Press.

North, Douglass. 1995. "Some Fundamental Puzzles in Economic History/Development." Paper presented at the University of California at Berkeley Economics Department seminar on development, October, Berkeley.

Ogus, Anthony. 1995. "Economics and Law Reform: Thirty Years of Law Commission Endeavour." *Law Quarterly Review* 111: 407–20.

Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *American Political Science Review* 87: 567–76.

Pettit, Philip. 1990. "Virtus Normativa: Rational Choice Perspectives." *Ethics* 100: 725–55.

Posner, Richard. 1979. "Utilitarianism, Economics, and Legal Theory." *Journal of Legal Studies* 8: 103.

Priest, George L. 1987. "Measuring Legal Change." *Journal of Law, Economics, and Organization* 3 (85): 193–225.

Rabin, Robert L., and Steven D. Sugarman, eds. 1993. *Smoking Policy: Law, Politics, and Culture.* New York: Oxford University Press.

Ramseyer, J. Mark. 1994. "The Puzzling (In)Dependence of Courts: A Comparative Approach." *Journal of Legal Studies* 23: 721–47.

Rawls, John. 1971. *A Theory of Justice.* Cambridge, Mass.: Harvard University Press.

Raz, Joseph. 1986. *The Morality of Freedom.* Oxford: Clarendon Press.

Rogers, Jim. Forthcoming. *The Early History of the Law of Bills and Notes: A Study of the Origins of Anglo-American Commercial Law.* Cambridge: Cambridge University Press.

Ross, H. Laurence. 1970. *Settled Out of Court: The Social Process of Insurance Claims Adjustment.* Chicago: Aldine.

Rubin, Edward. 1995. "Congress as a Bureaucratic Supervisor." Paper presented at a University of California at Berkeley seminar on law, economics, and organizations, May, Berkeley.

Rubin, Edward, and Robert Cooter. 1994. *The Payment System: Cases, Materials, and Issues,* 2nd ed. St. Paul, Minn.: West Publishing Company.

Rubin, Paul. 1977. "Why Is the Common Law Efficient?" *Journal of Legal Studies* 6: 51–64.

———. 1993. "Growing a Legal System, with Special Reference to the Post-Communist Economies." Working Paper 63. Center for Institutional Reform and the Informal Sector, College Park, Maryland.

Schussler, Rudolf. 1993. "Anonymous Exchange Cooperation." Paper presented at the fourth international conference on social justice research, July 1–3, Trier, Germany.

Schwartz, Alan, and Robert E. Scott. 1995. "The Political Economy of Private Legislatures." *University of Pennsylvania Law Review* 143: 595.

Simpson, A. W. B. 1973. "The Common Law and Legal Theory." In A. W. B. Simpson, ed., *Oxford Essays in Jurisprudence.* Oxford: Clarendon Press.

———. 1975. *A History of Common Law of Contract.* Oxford: Clarendon Press.

Stephan, Paul B., III. 1995. "Toward a Positive Theory of Privatization—Lessons from Soviet-Type Economies." Paper presented at a conference on economic analysis of international law sponsored by the George Mason University Law and Economics Center, May, Fairfax, Virginia.

Stigler, George. 1972. "The Government of the Economy." In R. W. C. and R. S. Eckaus, eds., *Contemporary Issues in Economics.* Boston: Little, Brown & Co.

———. 1975. *The Citizen and the State.* Chicago: University of Chicago Press.

Taylor, Michael. 1982. *Community, Anarchy, and Liberty.* Cambridge: Cambridge University Press.

———. 1987. *The Possibility of Cooperation.* Cambridge: Cambridge University Press.

Tyler, Tom. 1990. *Why People Obey the Law.* New Haven, Conn.: Yale University Press.

Ullmann-Margalit, Edna. 1977. *The Emergence of Norms.* Oxford: Clarendon Press.

Vlasikhin, Vasili. 1992. Comments made at an informal meeting of the Gruter Institute in the Canadian-American Institute, Moscow.

Waldron, Jeremy. 1993. "The Dignity of Legislation." Paper presented at a Columbia workshop on law, Columbia University School of Law, New York.

Weingast, Barry. 1996. "The Political Foundations of Democracy and the Rule of Law." Paper presented at a University of California at Berkeley seminar on law, economics, and organizations, Berkeley.

Wittman, Donald. 1989. "Why Democracies Produce Efficient Results." *Journal of Political Economy* 97: 1395–416.

World Bank Legal Department. 1995. "The World Bank and Legal Technical Assistance: Initial Lessons." Policy Research Working Paper 1414. World Bank, Washington, D.C.

Zimmermann, Reinhard. 1990. *The Law of Obligations: Roman Foundations of the Civilian Tradition.* Cape Town: Juke and Co., Ltd.

Zweigert, Konrad, and Heinz Kotz. 1992. *Introduction to Comparative Law.* Oxford: Clarendon Press.

Comment on "The Rule of State Law and the Rule-of-Law State: Economic Analysis of the Legal Foundations of Development," by Robert D. Cooter

Cheryl W. Gray

In this ambitious article Robert Cooter develops a game-theoretic model of private enforcement and uses it to address many wide-ranging topics, including low-level corruption and the suppression of high-level tyranny. Each section is interesting and provocative on its own; as a whole the article gives a flavor of the rich literature and almost endless possibilities in the application of economics to law.

The main underlying message of the article is that the only enforceable laws—and therefore the only laws that can result in a rule-of-law state—are those that closely coincide with existing social norms. This is true, in Cooter's view, for three reasons. First, because of the high costs of official enforcement, laws must be enforced primarily by the public on a day-to-day basis, and the public readily enforces and abides by only the laws that conform to its social norms. Second, laws that conform to social norms are economically efficient, whereas those that do not are by and large inefficient. Because the body of judge-made common law is developed "from below" based on social norms, it is both efficient and enforceable, while most top-down government regulatory law is neither efficient nor enforceable. Third, not only good commercial law but also the rule of law generally—that is, a limited state—develops automatically as market forces create multiple centers of power. Thus tyrants will be tamed and corruption stemmed through public enforcement if laws follow social norms and markets are allowed to work. The main policy conclusion for the World Bank is that it should not try to impose legal models from above but should instead concentrate on helping judges develop laws that conform to social norms.

My reaction to the article is mixed. On the one hand the fundamental premise—that laws that are consistent with social norms are most apt to be followed—seems both reasonable and powerful. Adopting laws that the public does not accept as legitimate and that therefore sit on the books unenforced can be futile, if not counterproductive. The fact that governments do this so often, particularly in developing

Cheryl W. Gray is principal economist in the Finance and Private Sector Development Division of the Policy Research Department at the World Bank.

Annual World Bank Conference on Development Economics 1996

countries or economies in transition from central planning (with their armies of foreign technical advisers), shows that the message is still worth emphasizing.

On the other hand the article pushes this fundamental premise too far. It is overly optimistic in averring that market practices, codified through judge-made law, will eventually overturn dictators and rid countries of corruption. History may have followed that path in some countries, but it seems to me that there are just as many examples of countries that have never been ruled by law. Similarly, it is unrealistic to expect that judge-made law can design and enforce all of the complex legal rules and institutional enforcement mechanisms needed in a modern market economy, such as rules governing corporations and securities, environmental safeguards, or guidelines for public education and health. Cooter does recognize this last point, albeit as a kind of afterthought rather than as a core constraint on judge-made law.

Nudging Change

Even in areas in which judges could feasibly make law, I question whether it would be optimal. The article's conclusions seem to leave little if any room for rapid, government-led, policy-induced change in norms and behavior in societies. Yet many existing social norms may hinder economic development or, even worse, support intolerably oppressive practices. In the economic sphere, for example, legacies and norms of hierarchy and patronage may give managers (and sometimes the mafia) extreme power over firms and inhibit the growth of effective external mechanisms of corporate governance. Or habits engendered by years of enterprise subsidies and soft budget constraints (particularly in transition economies) may lead firms to view debt repayment as optional, thereby undermining the development of credit markets. With regard to intolerable social norms, one need only consider female circumcision, wife immolation, or, as I once saw, the practice of chopping off young girls' fingers when an important man in the community dies. These are well-entrenched norms, but that does not make them worth supporting or embedding into formal government-enforced law.

In sum, there is everywhere a profound tension between what is doable and what is desirable. In some settings—particularly in developing and transition economies—rapid change is urgently needed to foster economic growth and improve the living standards of hundreds of millions of people. It is not enough to wait decades until social norms slowly evolve. Law must largely reflect accepted practices—that is, what is readily doable—but I believe it can also be an instrument to nudge change in the direction of what is desirable.

We who work in these settings see it happening often. In Russia, for example, the government tried through aggressive regulatory activity in 1993 and 1994 to force companies to establish independent registers of shareholders and to include outside representatives on boards of directors. In Hungary an aggressive bankruptcy law adopted in 1991 tried to change managerial attitudes by forcing into bankruptcy firms that did not pay their bills on time. In Poland a 1993 banking law tried to encourage banks to become more active in collecting bad debts, thereby forcing

restructuring in problem firms. In none of these cases was the law a complete success overnight. Indeed, many firms disregarded the laws or somehow turned them to their own advantage. The laws were nonetheless important instruments for changing social norms and building new institutions. Their proactive designs departed from accepted social norms and began a process of evolution away from those norms. My fundamental message thus differs slightly from Cooter's: laws must not diverge too dramatically from social norms, but activist lawmaking can spur change at the margin.

Creating Demand for New Laws

Despite my somewhat greater faith in statutory law as an instrument for change when change is desperately needed, I fully agree with Cooter that new laws may do little by themselves to change behavior. Active lawmaking from above can have a real impact only if supporting institutions develop simultaneously and if economic policy creates a demand for those new laws. For example, creditors in transition economies will not use new bankruptcy or other debt collection laws if they can turn to the state for a bailout when loans turn sour. Similarly, enterprises are unlikely to pay attention to minority shareholder rights unless they need to raise money on capital markets, something they need not do if they can rely on the state for subsidies. These are just two examples of how tight macroeconomic policies that impose financial discipline on banks and firms can help create demand for law from below. This notion of demand for law may in fact be similar to Cooter's notion of social norms; if so, we both agree that demand for law (or social norms) is necessary for statutory law to be effective in practice.

Who are the most likely agents of change through law? Common law can effect change—witness the U.S. Supreme Court decision in *Brown v. Board of Education,* which found that segregated education, long the social norm in the United States, was unconstitutional. But judge-made law is usually conservative, and statutory law attempts to force major change more often. In Central and Eastern Europe, for example, governments, particularly economic ministries, have been the main proponents of economic reforms to change social and business norms, while the courts (particularly the constitutional courts) have resisted rapid change and been active in overturning new economic policies as unconstitutional. My experience indicates that where rapid change is desirable, judges and courts are unlikely to be the agents of such change, particularly in civil law systems where judges have no mandate or tradition as activist agents for change.

What the World Bank Can Do

What does this mean for the role of the World Bank? Although helping build judicial institutions is important—and indeed is already part of several Bank projects—it is not the only, or necessarily even the primary, way in which the Bank can help build the rule of law in borrowing countries. Many countries remain far away from

an ideal equilibrium. They need new and well-structured laws in many economic areas where no laws—or, even worse, counterproductive laws—exist. But lawmaking will be ineffective if it proceeds in a vacuum. Corresponding efforts must be made to build supporting institutions (both legal institutions and more general information and watchdog institutions) and to ensure that sound macroeconomic policies stimulate the demand that will bring these laws to life. Listening to the people is important, not only because some social norms will indeed be efficient (as Cooter argues), but also because human institutions have only limited capacity for change. Nudging economies toward better performance is an incremental process, but I believe that some careful top-down activism—undertaken always with eyes and ears open—can help.

Comment on "The Rule of State Law and the Rule-of-Law State: Economic Analysis of the Legal Foundations of Development," by Robert D. Cooter

Patrick Juillard

Robert Cooter's article deserves close consideration not only because of its complexity, but also because it seems to reflect current thinking in some circles—though I am not sure which circles. Legal circles? Economics-turned-legal circles?

A lawyer would have problems with the substance of Cooter's article, because the principles on which it is based and the methods it follows are foreign to the principles and methods of legal thinking. I thought at first that this divergence was the result of differences in what I call cultural approach—the more dynamic and innovative U.S. approach against the more static and conservative European approach. Such differences are easily reconciled, since the approaches share a common philosophy. But then I began to realize that Cooter raises more basic issues about the nature, sources, and role of law in a modern society, and I concluded that there was a wide gap between his approach, which I would call the *positive* approach, and the more classical, traditional approach, which I would call the *normative* approach. I also realized that some of the differences might well be irreconcilable.

Terminology

Part of the problem I have stems from terminology. Cooter uses some legal terminology in his analysis in ways that lawyers would find unfamiliar. An example is his use of the term *rule-of-law state,* which he defines as a state in which law responds to social norms. Since state law responds to social norms, most people perceive law as just, and many people obey just laws out of respect, which creates a rule-of-law state. I will come back later to the notion that a law is necessarily just because it responds to social norms—a notion I find dangerous. But what I take issue with at this point is the definition of the *rule-of-law state* as one in which people obey the law because they perceive it to be just—that is, responsive to social norms. This definition does not fit with the well-accepted definition of *Rechsstaat* (a state governed

Patrick Juillard is professor of law at the University of Paris-Sorbonne.

Annual World Bank Conference on Development Economics 1996

by the rule of law). As developed by German philosophers and lawyers, Rechsstaat refers to a state that respects fundamental ethical values, which therefore must govern the body politic as a whole. In its contemporary meaning, as developed in domestic and international practice, Rechsstaat refers to a state that is organized as a representative democracy and that protects the fundamental freedoms of its citizens. In that sense the Rechsstaat is not a state shaped by social norms; rather, it is one that shapes social norms so that they respect the ethical values on which the state rests.

Lex Mercatoria

No one would argue with Cooter's assertion that the modern economy creates many specialized business communities or with his characterization of the norms created by these business communities at the *new merchant law.* This characterization is not unfamiliar to legal scholars in Europe and elsewhere. Berthold Goldman of the University of Paris faculty wrote two famous articles on the topic as long ago as 1964 ("Frontières du droit et le mercatoria") and 1979 ("La lex mercatoria dans les contrats et l'arbitrage internationaux"), which have inspired legal scholars the world over. "Nil novi sub sole"—there is nothing new under the sun.

The proponents of lex mercatoria have gone far—indeed, very far—in their defense of this new merchant law. They have gone so far as to submit that *lex mercatoria,* while it did not evict domestic law from the field of domestic transactions, could well be viewed as a legal order unto itself, capable of governing international transactions in the absence of substantive treaty law in this field. A number of arbitral awards rendered under the auspices of the International Chamber of Commerce refer to the principles of lex mercatoria as the *ratio decidendi* (determining factor), even though some of these principles appear to run against well-established principles of domestic law. Under the civil law system one of these well-established principles is the sanctity of contracts, a principle believed essential for legal security (*sécurité juridique*). Hence the wording of Article 1134 in the French Civil Code: "Contracts, whenever entered into in accordance with the law, shall be deemed to be the law of the parties" (*les conventions légalement formées tiennent lieu de loi à ceux qui les ont faites*). One consequence was that a contract could not be amended if one of the parties refused to renegotiate it, even in case of hardship for the other.

This principle was indispensable for economic purposes in the early nineteenth century, when the most frequent kind of contract was the sales contract and the most frequent sale was of unmovable property. But what was true in a rural society during the 1800s is no longer true for international transactions in the twentieth century. Today the magnitude of transactions, the importance of the financial interests at stake, and the length of the contractual period make it impossible to apply the principle embodied in Article 1134. Thus the principle of renegotiation in good faith has evolved through arbitral awards for cases in which the circumstances surrounding performance warrant it. This is a marked departure from the fundamental principle of civil law. French courts will nevertheless grant recognition and

exequatur (enforcement) to such awards which, because they are international in nature, need not respect domestic legal order but only international legal order. This is an example of how lex mercatoria can work against domestic law.

But one should also stress that the concept of lex mercatoria as a legal order still runs into heavy criticism. Leading opponents claim that lex mercatoria is little more than scattered bits and pieces of decisions and behavior that cannot be brought together to create a full-fledged legal order. They also claim that even were lex mercatoria to eventually develop into a full-fledged legal order, it could not stand on a par with domestic legal orders or the international legal order. Why? Because inherent in the concept of legal order is the power to impose sanctions. In other words those who believe this abide by the classical position that law enforcement must be a state monopoly. And this is a position that Cooter strongly questions. With all due consideration, I disagree with him. But I have several areas of agreement as well.

Points of Agreement

1. *Too many laws will kill the law.* The French Council of State recently tried to figure out just how many laws are in force in France. This proved a difficult exercise because of the diversity of sources. As a sovereign state France has the jurisdiction to enact domestic legislation and regulations. As a member of the European Union (EU), France has also accepted to be bound by Community law—EU regulations, directives, and decisions. And finally, as a member of the international community, France enters into international treaties and agreements with other sovereign states, and under the French constitution these treaties and agreements prevail over domestic legislation.

"Nemo censetur legem ignorare"—no one is permitted to be ignorant of the law. That dictum made sense at a time when there were very few laws and when the laws were intended to ensure the proper discharge by the sovereign state of its essential functions. It makes much less sense today, when legislation and regulations proliferate as the functions of the state have expanded from the essential to the nonessential and as legislation and regulations have become the fruits of expediency. It is unreasonable to assume that the average citizen could know anything about all the laws and regulations from this diversity of sources—their purpose and content—even though they may directly affect that citizen's rights. Sadly, law is no longer a matter for citizens but has become a matter only for lawyers.

There is a widespread feeling that the best legal systems are those that create few but enduring laws, laws that are at the same time reasonable and equitable. From that belief flows the conviction that political organizations, be they sovereign states or international entities, ought to impose some restraint on themselves and should not legislate on matters that are not vital to the preservation of collective interests.

A few years ago the French Parliament enacted the law on domestic transportation (*Loi d'orientation des transports intérieurs*). One provision of the law proclaims that every person has a right to transportation. That may well be, but what does that mean? Does it mean a right to free transportation? To safe transportation? Or that the French government must take all necessary steps to ensure that everyone can

enjoy high-speed transportation from any remote location to any other remote location within the national territory? And what happens if the French government fails to ensure that everyone can exercise that right to transportation? The mere fact that the provision raises so may fundamental questions calls its rationality—and rationale—into doubt.

Citizens in member countries of the European Union also resent the proliferation of Community law—all the more since regulations and directives are issued by Community officials who are not subject to any parliamentary control. Such resentment runs especially high when Community laws affect people's day-to-day life. There was some squabbling a few years ago about the public health risk to the Community caused by the consumption of fermented cheese. Should citizens of member countries be allowed to eat fermented cheese? In truth, the question may really have boiled down to a more direct one: Should they be allowed to eat (fermented) French cheese or only (unfermented) Dutch cheese? The average French citizen, whose family has been eating camembert for generations, could not but resent that question, rightly feeling that this was not Community business and that, by raising this question, the Community was intruding on people's private lives. The reaction was quick and strong. Member states inserted in the Maastricht Treaty the famous "subsidiary" provision, which states that Community action should be contemplated only when the common goal would not be better served by action by member states.

The proliferation of laws has a devastating effect on citizens' respect for the law. They come to believe that most laws serve no real purpose and that they are the fruit of expediency rather than of necessity. It follows in many people's minds that because there are too many laws, and because so many of them are useless and senseless, citizens may choose which laws should be complied with and which should not.

2. *Too few laws will kill development.* Another point in Cooter's article is that, while too many laws may kill the law in industrial countries, too few laws may kill development in developing countries. This is absolutely true. Law is crucial to economic development, for a number of reasons.

A reasonably developed legal system is a condition of economic development because the players have to know the rules of the game before they can play. Thus the legal framework for economic activities must exist in order for the economic activities to develop. The degree of sophistication of a legal system will vary according to the level of development. Developing countries, which often start from a rural base, ought first to enact legislation to promote agriculture and agricultural development, rather than legislation to promote industry and industrial development.

The example of the formerly socialist states in Europe shows how difficult it may be to build or rebuild a legal system from scratch. The key concept here seems to be coherence. A legal system rests on a limited number of fundamental principles. These fundamental principles must be clearly identified before the state embarks on the task of building or rebuilding its legal system. Coherence, however, may be difficult to achieve, since the building or rebuilding of the legal system is often entrusted to foreign experts.

I know of an Eastern European country that decided that its civil code should be drafted by an Italian, because Italians are the foremost experts on Roman law, the source of the law in the old days; that its commerce code should be drafted by a German, because the German code of obligations is deemed to be one of the most elaborated codifications in the field; and that its accounting law should follow the French system, because the French have evolved a sophisticated system of accounting laws. No one ever doubted that the end result would be magnificent or that any of these laws would be anything but a masterpiece of legal draftsmanship. What no one seems to have asked, however, is whether the result will be a unified system of law, in which all the elements are in harmony because they all rest on the same fundamental principles.

A reasonably developed legal system is also important to development because it acts as an incentive for foreign investment. The World Bank has played a key role in creating or re-creating environments in developing countries favorable to investment from industrial countries. Investors base their decisions on economic considerations, but they also take into account the political and legal environment, especially its transparency and stability. And for that, foreign investors require no less than a full-fledged legal system that will protect their property and their activities within the host country. Developing countries ought to realize that adherence to a number of international instruments may be considered a prerequisite not only by foreign investors but also by the countries of which they are nationals. Bilateral investment treaties and participation in such multilateral institutions as the International Centre for Settlement of Investment Disputes or the Multilateral Investment Guarantee Agency are essential elements of any legal policy aimed at economic development. The role of the law here should not be underrated.

Points of Disagreement

If I correctly understand Cooter's article, it rests on two basic ideas. One is that social norms must precede legal norms, because social norms generally reflect the sense of justice of the people, and as a corollary, that whenever a social norm reflects the sense of justice of the people, then legislators ought to enact it into law. The second idea is that whenever social norms have been enacted into law, the people are willing to cooperate to enforce the law because these social norms reflect the sense of justice of the people.

1. *The source of the law can be found only in moral norms, not in social norms.* Cooter uses the word *norm* to mean rules imposing obligations, such as the rules of conventional morality. My problem here is that I do not see how a social norm can ever create obligations. A social norm, as Cooter points out, refers to average behavior. Thus a social norm is a fact, and may be observed as such, but it is one to which no value is affixed. In that sense I believe that the use of the word *norm* is incorrect precisely because that word designates an object that is susceptible to a value judgment. But a social norm, while it may or may not reflect preexisting values such as moral values, may neither create nor impose moral obligations. Thus I have prob-

lems with the notion that the source of the law may be found in social norms. The source of the law can be found only in moral norms.

2. *"Average" behavior cannot be the basis of law.* In certain communities average behavior may run contrary to certain moral values of a universal character. There have been numerous dreary examples of communities in which "average" behavior reflected nothing but bigotry and intolerance. Consider nineteenth-century Russia or twentieth-century Germany. Clearly, bigotry and intolerance, though they may be average behavior, cannot possibly create or impose obligations.

Unfortunately, history tells us that such behavior was not viewed by its perpetrators as violating any moral obligation. There is no evidence that antisemitism was perceived as a gross violation of the most basic human rights either by Russians in the nineteenth century or by Germans in the twentieth century. Moral values, not social norms, must impregnate the law, because social norms may or may not conform with moral values, whereas the law must always do so.

3. *Law based on efficiency at the expense of morality is unprincipled law.* A concern expressed in a number of countries, mainly developing countries, is that the law, while responsive to certain needs of the international community, totally ignores the moral values on which that community stands. The argument has often been made about international economic law that it is an unprincipled body of law, in the sense that it does not rest on any moral principles.

International investment law is one often-cited example. It has been alleged that modern international investment law no longer takes into considerations the values from which it emerged. At its inception international investment law was intended to strike a balance between two moral principles: one, that government should ensure that foreign investment benefits the community as a whole, and two, that all foreign investors should be treated in a way that conforms to certain moral and legal standards of a universal character.

The compromise reached initially was that every state was free to choose whether to admit foreign investment into its territory but that once it had decided to admit foreign investment, such investment should be afforded fair and equitable treatment. That compromise reflected certain moral preoccupations. The recent trend, however, is to consider that states should not be free to choose whether to admit foreign investment. Rather, the principle of national treatment should apply, meaning that all foreign investment should be as free to circulate without state interference as is domestic investment.

The principle of national treatment does not reflect any moral values. It is just the legal embodiment of an economic predicament: economic efficiency requires the free movement of investment. It may well be that free movement of investment will foster economic efficiency, but it may also well be that the free movement of investment will come only at a very high social cost in terms of unemployment, for instance. Can it then be said that the law—international economic law—should occupy itself solely with economic efficiency and disregard moral values entirely? Should the principles of international economic law be the principles of economics rather than the principles of ethics? If that were the case,

then in the eyes of most lawyers international economic law would be an unprincipled body of law.

4. *Private enforcement is contrary to legal theory and to the theory of the state.* I also have problems with the notion that enforcement of the law can be entrusted to individuals. The reasoning goes something like this. A given person will accept some social norms but will not accept some others. Whenever that person accepts a certain social norm, that person can be said to have internalized that norm. A person who has internalized a social norm will then be willing to cooperate in order to enforce it, whether that social norm has been enacted into law or not.

These concepts are foreign to legal theory. The law must be obeyed because, by definition, the law is the embodiment of moral values. Thus it cannot be left to individual discretion to determine which laws must be respected and which may be disrespected—except in extreme cases of unjust laws that ought to be disobeyed because they are in gross violation of individual rights. Furthermore, the notion of private enforcement is contrary to the definition of government, which is that a government is a political organization vested with a monopoly on enforcement.

References

Goldman, Berthold. 1964. "Frontières du droit et lex mercatoria." *Archives de philosophie du droit.* Paris : Edit. Sirey.

———. 1979. "La lex mercatoria dans les contrats et l'arbitrage internationaux." *Journal de droit international* : 747

Comment on "The Rule of State Law and the Rule-of-Law State: Economic Analysis of the Legal Foundations of Development," by Robert D. Cooter

Michael J. Trebilcock

The role of the rule of law and of the legal system more generally as an agent of social and economic development is contested terrain. Ghai (1991) argues that law is typically a reflection of underlying social and economic realities and that legal change is not a fruitful way of effecting social change: law is a dependent, not an independent variable. His argument suggests caution in assuming that the choice of legal institution drives the course of economic and social development, rather than the reverse. In contrast, in the context of Central and Eastern Europe, Brenner (1990) argues that economic problems emanate from a weak tradition of law generally and the rule of law specifically and that reform of laws and legal institutions should be accorded a high priority in development efforts. This view assumes (perhaps heroically) that the legal system can play a transformatory economic role, despite the pervasiveness of social, economic, and cultural forces that may oppose it.

The intermediate position, developed by Robert Cooter, is that the rule of law cannot derive from top-down centralized planning, legislation, or regulation making. The policy implications of this view are that international agencies like the World Bank should provide technical assistance to developing countries to help them develop communities of judges and scholars who can shape law to the realities of accepted social norms and business practices.

I find much in Cooter's approach that is compelling. In many former command economies and developing countries many laws that are well drafted by Western standards appear to correspond minimally to the day-to-day economic and social realities in the communities in which they purport to apply. In societies in which the state traditionally has played a dominant role in allocating resources, even governments that are now committed in principle to economic and political liberalization are tempted to turn to centralized edicts rather than decentralized evolution.

Michael J. Trebilcock is professor of law and director of the Law and Economics Program at the University of Toronto.

Annual World Bank Conference on Development Economics 1996

Limitations in Cooter's Conception of the Rule of Law

While noting this basic point of agreement with Cooter, however, I believe that there are substantial limitations to his conception of the role of law in economic and social development. First, Cooter believes that the state should enact laws that respond to social norms that encourage cooperation rather than appropriation, that is, laws that maximize the cooperative surplus and discourage forms of opportunism in which individuals destroy part of the cooperative surplus in order to reap a larger share of it. However, discerning social norms (and recognizing them as law) that maximize the cooperative surplus is often likely to prove a daunting task. Many of the debates in law and economics in industrial country contexts over the past twenty years have related to whether particular rules in tort law, contract law, corporate law, or bankruptcy law are or are not efficient. These debates are not easily resolved and have often been indeterminate. The thesis that the common law has a tendency to evolve efficient rules is widely contested in the law and economics literature. Yet Cooter offers the parallel thesis that social norms evolve in efficient ways.

Second, groups subscribing to social norms may only incompletely internalize the full costs and benefits of the norms. Cooter acknowledges that with many intermediate social and economic institutions the norm that will be adopted is that of efficiency for interests encompassed by the organization and monopoly toward outsiders. Similarly, he acknowledges that in many contexts without the benefit of formal third-party enforcement of laws, enduring relationships that are able to invoke tit-for-tat sanctions for noncompliance with norms may be effective in solving the agency (or prisoners' dilemma) problem, but that these relationships often depend on close ethnic, kinship, or religious ties. As is starkly evident from the history of conflicts in many developing countries, social norms adopted by ethnic, kinship, and religious groups will often exhibit characteristics similar to those adopted by intermediate institutions, that is, efficiency for interests encompassed by the group but monopoly or exclusion toward outsiders. Unless social groups subscribing to particular social norms are coterminous with the whole of the relevant society, the presumption that the norms in question are efficient from a larger social welfare perspective is at best tenuous.

Third, Cooter's concept of efficient social norms is ambiguous. In the conference draft of his article, Cooter argues in a discussion of land law problems and legal reform in Papua New Guinea that the fundamental problem with individual ownership with freehold title, as proposed by the Australian administration, was that it violates social norms that define local justice. Customary law stipulates which relatives will inherit property, rather than allowing owners to designate their heirs. Thus a person who would have inherited land under customary law feels unjustly deprived when the land is converted to freehold and someone else is stipulated as heir or the land is sold. According to this view of social norms, efficient social norms have less to do with maximizing the cooperative surplus and more to do with preserving traditional notions of status and distributive justice, however inefficient these norms may be in the former sense. Thus wealth-maximizing social norms

become, by a methodological slide, utility-maximizing norms, which, in the best Chicago tradition, presumably represent whatever norms one observes people apparently or traditionally subscribing to: these norms must be efficient (in a utility-maximizing sense), or other norms would have been adopted. This formula seems a recipe for paralysis in reforming legal systems that exhibit features that clearly seem dysfunctional in terms of maximizing the cooperative or social surplus, if reforms are antithetical to a long-held sense of individual or group entitlement.

I believe that Cooter's approach is too static in other respects as well. First, it is insufficiently sensitive to the distinctions between insiders and outsiders and between the informal and formal sectors in many developing countries. While social norms may provide an appropriate basis for state laws in many (but not all) interactions within long-established social groups, in a development context these groups increasingly find it necessary or productive to interact in their economic dealings with members of other social groups, with outsiders in the formal sector within the same society, or with the broader international economy (foreign traders or investors). These outsiders will not necessarily or even typically subscribe to the same social norms so that, in the case of a conflict between social norms, state laws can obviously serve an important mediating function by supplying a set of laws that enable productive interactions between insiders and outsiders to occur. As Douglass North (1990) has emphasized, one of the major challenges in economic development is to broaden the contract opportunity set to extend it to impersonal exchanges. In meeting this challenge, the legal system and credible third-party formal enforcement mechanisms are likely to play a critical role.

Another respect in which Cooter's perspective seems to be too static is in his failure to adequately acknowledge that the role of law and the legal system in developing countries cannot be disengaged from changes in the broader social, political, and economic environments in developing countries. The rate and nature of socioeconomic and political change are often dramatic and in many cases are propelled by circumstances, including international forces, over which countries have little or no control. Given that economic development necessarily implies change, it is not obvious that institutionalizing the old is an appropriate prescription for responding to (or facilitating) the new.

Even if long-established social norms were once efficient (in the narrow or the broad sense), these norms will often be severely strained in the face of rapid changes or transitions in people's lives, and more than incremental changes in state laws may be required to adapt to these changes. Cooter has argued, for example, that local land court magistrates in Papua New Guinea should give up the idea that custom is static and that only the legislature can develop a new land law. Rather, they should identify changing customs, raise them to the level of law, extend the customary law of intragroup interactions, and apply it to transactions with outsiders. In contrast, I have argued elsewhere that the rapid proliferation of the cash economy in Papua New Guinea has placed customary land law under such severe strain and rendered it so dysfunctional that formal centralized legislation is required to reduce intragroup and intergroup transactions costs by providing a

more formal structure for informal landowning groups contemplating selling their land; to facilitate a limited range of direct dealings with outsiders, subject to a limited formal approval process; and to set up a system of sporadic registration of title to customary land in order to stabilize and protect property rights (Trebilcock 1984). While each of these reforms can be viewed as incremental, each nevertheless entails more significant and systemic changes than are likely to evolve under Cooter's magistrate-made customary law.

This example leads me to two final points on the role of law in developing countries. As the example of land reform in Papua New Guinea demonstrates, one role for state laws in societies undergoing rapid change is to provide additional institutional options for individuals or groups in adapting to these changes. Here the law plays an important facilitative function in opening up new nonmandatory, noncoercive institutional options that individuals, left to their own devices, may have difficulty devising, given the public goods nature of institutionalizing many of these options. If individuals or groups prefer not to avail themselves of these options and choose instead to continue to govern their social and economic interactions by reference to established social norms, that is their prerogative. But often the state can play an important role in enriching the menu of options open to its citizens, and the use of these options may, over time, change the social understanding of prior norms.

Beyond this facilitative role, and particularly in contexts where lifestyles and belief systems are undergoing dramatic strain or change, the law can play a heuristic or educational role in instilling a new set of social norms that may be internalized by many members of a society over time. It has been widely noted, for example, that the highly personalized economic and social networks in many developing countries often carry over in a dysfunctional way to interactions between citizens and government officials. In such interactions ethnic, kinship, or religious ties can lead to a highly personalized or patrimonial form of public administration that is vulnerable to welfare-reducing forms of favoritism, discrimination, and corruption. In this case long-established social norms that may previously have been efficient, in a broad or a narrow sense, may reduce efficiency. Strong, efficiently enforced laws against favoritism, discrimination, or corruption in government may be necessary to delimit the rights and obligations implied by ethnic, kinship, or religious relationships and to enforce—and possibly internalize over time—a new set of social norms concerning impartiality and due process in public decisionmaking.

Conclusion

In summary, while Cooter's caution against top-down state-imposed law is salutary, I believe that the relationship between informal social norms and formal state laws is much more interactive and dynamic than he assumes. And this implies a significantly larger role for formal state laws than his simple prescription that these laws should largely respond to and reflect existing social norms.

In order to reduce the dissonance between laws as written or administered and citizens' needs and expectations (thereby enhancing compliance and enforcement), I envisage a larger and more activist role for law in the process of economic development. That role would entail more effective participatory and consultative processes in the design and administration of new laws than is often the case. This implication may be consistent with Cooter's framework of analysis but extends it to a domain in which lawyers as "process engineers" should have constructive contributions to make. This view also casts participatory democracy in a more benign light than does Cooter, who worries that majoritarianism will yield incoherence in preference aggregation (social choice theory) or spawn rent seeking by special interest groups (public choice theory). Why social norms, which are also typically the product of group interactions, are not equally afflicted by these problems is not explained.

References

Brenner, Reuven. 1990. "The Long Road from Serfdom and How to Shorten It." *Canadian Business Law Journal* 17: 195–225.

Ghai, Yash. 1991. "The Role of Law in the Transition of Societies: The African Experience." *Journal of African Law* 35: 8–20.

North, Douglass. 1990. *Institutions, Institutional Change, and Economic Performance.* Cambridge: Cambridge University Press.

Trebilcock, Michael. 1984. "Communal Property Rights: The Papua New Guinean Experience." *University of Toronto Law Journal* 34: 377–420.

Floor Discussion of "The Rule of State Law and the Rule-of-Law State: Economic Analysis of the Legal Foundations of Development," by Robert D. Cooter

Jean-Philippe Platteau (discussant from another session) said that he thought that the discussion had overlooked the issue of multiple, conflicting norms and of how the law is to position itself relative to these norms. The main problem of the law is deciding which group, or new emerging or contesting norm, will be supported by the law. In the discussions everyone seemed to implicitly assume that an entire country shared one homogeneous set of social norms, but in fact norms evolve, and different groups defend different norms. And as Cheryl Gray (discussant) had pointed out in her comment, certain new groups, including business groups, often want to see dominant norms change.

A participant from Brazil asked whether the panelists had considered the issue of an independent judiciary. In a number of countries it is relatively easy to get modernizing changes, even legal changes, but it is difficult to develop a truly independent judiciary. Another participant with a related question noted that citizens of industrial countries can participate in both the development of the rules of behavior and their interpretation. But in developing countries legal systems are far from social norms precisely because the people who are affected by social norms are so far removed from legal systems. If judges are the creators of laws, then people who have access to the legal system gain from this separation and people who are denied access lose. In such systems judges have no interest in having a system that is more open, participatory, and efficient, and people who are part of the formal legal system are unlikely to invite people who are estranged from it to participate. So how can the legal system be opened to everyone when there are people who stand to lose from such a change?

Edward Kane (discussant from another session) commented that the theory of conflict between principals and agents that is central to Cooter's thesis has evolved in modern finance into a theory of financial contracting. And part of that theory is that over time, in a somewhat inchoate way but in definite equilibrium, there is a tendency to minimize agency costs through covenants and other arrangements. This

This session was chaired by Ibrahim F. I. Shihata, senior vice president and general counsel at the World Bank.

Annual World Bank Conference on Development Economics 1996

applies even at the social level. Kane wondered how declining agency costs fit into Cooter's theory.

Citing the examples of the effective pooper-scooper regulation in Berkeley, California, and Hungary's bankruptcy law, a participant from the World Bank asked the panelists to say a little more about when things worked and when they did not. Alcohol is an extraordinarily dangerous drug, for example, yet Prohibition in the United States had certainly been a failure. Would the antismoking ordinances that are now common in Western industrial countries have been greeted enthusiastically ten years earlier? Would Hungary's bankruptcy law have been more successful if it had been less draconian?

Gray responded that rules are only enforceable if they fit local social norms. When she is on flight in the United States on a U.S. airline, for example, and someone lights a cigarette, she has only to tell them to put it out and they do. If the same thing were to happen on a flight in Eastern Europe on the same U.S. airline, the smoker would not pay the least bit of attention to her even though the airline's rules were the same in both cases. As for Hungary's bankruptcy law, despite its failings, the fact that it was draconian made it more effective. And it was draconian: 25,000 firms, accounting for almost 10 percent of Hungarian GDP, filed for bankruptcy in two years. The key to this high filing rate was that businesses were required to file if they were in arrears or the manager could be sent to jail. There had been a lot of manipulation of the law, and it was by no means perfect, but the amazing thing was that it had begun to change perceptions and norms. And it had built a trustee profession, which was now quite professional, and it had built the courts. Two years later, however, the law had been so watered down that it was no longer used. Still, to Gray the draconian law had had positive effects in terms of institutions, since it seemed to tilt the system in the right direction. She admitted, however, that other analysts considered the law a complete disaster.

Before asking for Cooter's response to questions from the floor, Ibrahim Shihata (chair) had comments of his own. To the extent that Cooter expressed commonsense words of wisdom, he found little to disagree with, though like all words of wisdom it was subject to qualification. As a theory about the rule of law, however, he had disagreements. True, when state laws respond to social norms they are more likely to be respected and obeyed, but that does not mean that this simple and convincing message should be taken to its extreme. Traditional concepts need not always prevail over modern ones. Case-based common law is not necessarily more commendable than statutory codes in all societies and all sectors.

As much as social norms can and do drive formal law, the relationship can—and at times should—go the other way. Laws have a dual function, as expressions of the current social order and as instruments of its orderly change. In a developing society law is needed not only to restate and refine social norms and local usage but also to introduce new values and structures. In business matters the role of law would be deficient if it followed only what businesses do. In discussing World Bank policy on the issue, Shihata continued, it is important to distinguish between Bank support for legal and judiciary reform and legal reform associated with adjustment lending. In

projects for legal reform, the Bank fully recognizes that sophisticated laws should not be imposed from the top or from the outside regardless of the probable beliefs to the society involved, and that legal reform should as much as possible reflect the general beliefs and opinions of that society. But the Bank also recognizes that developing countries can learn a lot from the experiences of other, especially industrial, countries.

In response to the comments from the floor, particularly comments that his thesis was unbalanced, Cooter tried to put things in perspective. He said that he was broadly engaged in a structural approach to adjudication, of which his theory about the evolution of social norms in interaction with state law—or at least his view of state law—was just a part. Economists understand how to analyze markets and regulation in terms of market failures: if a market has a certain structure, it works; if it does not, it fails. Similarly, in considering the regulation of an economy, one should look first to see if there has been a failure. If there is none, then there is no prima facie case for regulation; it is not possible to improve on a perfectly competitive price. If there is a market failure, however, there might be a case for regulation.

Taking the next step, it had to be determined if there is a regulatory institution that would make the situation better rather than worse. In Cooter's view the liberal tradition in law (which emphasizes government) would apply the same principle. It would look at the system of social norms and ask, where are there failures in that system? It would ask, was the social order failing to produce a norm where one was required, or was it producing a bad norm where a good one was needed? To answer these questions we must take a structural approach to adjudication, to examine the social system, the system of incentives, through which norms are evolving. In Cooter's view, if the structural approach to adjudication were fully developed, we would have an account of normative failure and in light of that we would know where legislation and regulation are required.

It was not his view, Cooter continued, that the normative system is necessarily and always efficient, good, or just. And here critical morality plays a role. In any society there are competing moral norms. Although his article discussed ordinary social morality as if there were a consensus, he also believes that part of the dynamic of politics is that against ordinary morality there is always a critical morality. Against the view, say, that women should have a certain role in the economy there is a view that they should have a very different role in the economy. That tension between critical morality and ordinary morality drives much of politics and deserves fuller discussion in his account of the structural approach to adjudication, which he has not yet developed. In that account he would also stress that it is crucial that judges learn to identify the characteristics of social norms that externalize costs onto other people, because in a pluralistic economy there is no reason to give one group more privileges than another and to say that one group must bear the costs for another group. That would be a case of normative failure, much as pollution is a market failure in the economy as a whole.

After developing an account of structural adjudication and identifying failures in the evolution of norms, the next step is to look at the political order, especially in

majoritarian processes, and see what deficiencies might exist there. Cooter would probably deemphasize majoritarian processes and the consensual processes that produce social norms. Their justification seems to rest more on the fact that the law improves on a social norm through its more effective enforcement than on the view that the law is the will of the majority, a view that has created a lot of the rent-seeking legislation and political problems that others had observed.

As for Michael Trebilcock's (discussant) comments on Papua New Guinea, Cooter said that although the magistrates may have many defects and their ability to develop land law through the common law process might not be perfect, he would urge Trebilcock to consider carefully what Papua New Guinea's Parliament does and how it functions (Cooter 1989, 1991). How many U.S. citizens, Cooter asked, would want constitutional protections against the taking of their property repealed and have Congress reallocate land? He thought Congress would do a pretty bad job of it, and he did not think that Papua New Guinea's Parliament would do any better in terms of justice and fairness.

As for whether his views are conservative or not, Cooter said that in his view the great engine of change in the modern era has been the market, so in that respect he could be described as Marxist. His theory of social norms pointed to the opening up of social change to the free play of market forces in the evolution of social life—and in that respect, he was a great admirer of Max Weber. He said that he considers the morality of capitalism to be crucial to economic development and believes that the law should respond to rather than try to direct that morality. A proper respect for the norms of business life, in the context of modern international competition, would create a dynamic situation with rapid social change.

References

Cooter, Robert. 1989. *Issues in Customary Land Law.* Port Moresby, Papua New Guinea: Institute of National Affairs.

———. 1991. "Inventing Market Property: The Land Courts of Papua New Guinea." *Law and Society Review* 25: 759–801.

Contracting, Enforcement, and Efficiency: Economics beyond the Law

Avner Greif

Neoclassical economics maintains that a legal system is required to foster anonymous exchange and hence efficiency. Social scientists from other disciplines emphasize the importance of informal contract enforcement in fostering exchange. Using a game-theoretic framework and anecdotal evidence, this article argues against the neoclassical theory that only legal contract enforcement facilitates anonymous exchange. At the same time it argues against overstating the extent to which either reputation or personal trust can govern exchange relations efficiently. It proposes that there is no single optimal combination of contract enforcement institutions; different sets of institutions may be optimal under different economic conditions. Moreover, in a given economic situation the nature and efficiency of the existing set of contract enforcement institutions is a function of social, cultural, and political factors. Hence the development of an optimal policy for enhancing contract enforcement in a particular economy must take into account the interrelations among the economic, social, cultural, and political factors that characterize that economy.

Ever since *The Wealth of Nations,* exchange in anonymous markets that facilitates the division of labor has been viewed as essential to economic efficiency and development, implying their dependence on an effective system of contract enforcement. The general equilibrium model implicitly assumes that the legal system enables all mutually beneficial exchange. Indeed, the economic success of Western civilization is often attributed to the emergence of a market economy based on an impartial legal system.

This neoclassical view that places the legal system at the center of contract enforcement in market economies has recently been criticized on the basis of evidence indicating that many contemporary exchange relations in the West and elsewhere are informal. The associated contract enforceability is not provided by the

Avner Greif is associate professor of economics at Stanford University. The author gratefully acknowledges the support of a grant from the National Science Foundation and research assistance from Michala Martens-Daehnfeldt. Masahiko Aoki and Anne Krueger offered helpful comments.

Annual World Bank Conference on Development Economics 1996

legal system but is based on reputation, general morality, and personal trust within social networks. Empirical evidence indicates the importance of two distinct systems of informal contract enforcement: the *individualistic* system of informal contract enforcement prevalent in the West, under which the reputation and morality of the individual are key, and the *collectivist* system of contract enforcement prevalent in most other societies, under which personal trust within social networks is critical.

Previous work on informal contract enforcement has advanced bold claims regarding the relative superiority or importance of the various systems of informal contract enforcement. A long tradition in economics holds that reputation provides a "free lunch" of informal contract enforcement. According to this view, reputation enables contract enforcement without any cost, and so a reputation-based contract enforcement system is more efficient than a system based on law. A long tradition in economic sociology holds that personal trust is the most important determinant of the extent of informal contract enforcement in any society. Hence the Western individualistic system is deficient because it lacks sufficient trust.

This article challenges all these views. It claims that neoclassical theory does not imply that a legal system is necessary for anonymous exchange, while the history of the West does not support the view that the legal system is key for the operation of market economies. In many market economies the legal system may have (and may have had) a limited role. At the same time the article argues that neither reputation nor personal trust provides a free lunch. Each is costly and at times may be more costly than contract enforcement based on the legal system. Finally, theoretical and historical considerations indicate that it is misleading to take the social networks that underlie contract enforcement based on personal trust and reputation as exogenous in analyzing systems of contract enforcement. More generally, a system of contract enforcement is a product of complex historical, cultural, political, social, and economic processes. Hence observed systems are not necessarily efficient, a society's economic and noneconomic aspects determine the relative efficiency of its contract enforcement system, and systems of contract enforcement exhibit path-dependence.

Methodologically, the article combines game-theoretic reasoning and anecdotal evidence. Game theory is used to explore alternative contract enforcement mechanisms; anecdotal evidence is used to illustrate the validity and relevance of these possibilities. The theoretical framework is useful in providing specific insights but clearly has a limited ability to capture the complexity of reality and the interrelations between informal contract enforcement and economic efficiency. In particular, it restricts attention to informal contract enforcement in economic exchange, ignoring the broader issues of the enforcement of property rights, which may be crucial to innovations and particular forms of business organization, and it does not examine the use of coercive power in informal contract enforcement (for an illuminating discussion of this issue, see Gambetta 1993). The content of the "law" that governs informal contract enforcement (an issue discussed in Ellickson 1991) and mechanisms for resolving disputes are discussed only briefly.

Formal and Informal Contract Enforcement in Theory and Practice

For an economist an exchange transaction is a contract governing property rights among specific individuals. In deciding whether to exchange, an individual considers the physical attributes of the goods to be exchanged, the distance in time and space between the *quid* and the *quo,* and the identity of the other party to the transaction. A necessary condition for (voluntarily) exchanging goods is that the transaction be mutually beneficial. Individuals will perceive an exchange as mutually beneficial only if they expect the other party to the exchange to fulfill its part of the agreement. The set of mutually beneficial exchanges thus depends on the institutions that ensure compliance with the terms of the exchange. To see how contract enforcement institutions determine the feasibility of exchange, consider an economy in which exchange is supported only by social ties within each of the small communities that constitute the society. Exchange among members of a community can take place, but markets for goods exchanged between members of different communities cannot function.

The Role of the Legal System in Neoclassical Economics

A common interpretation of neoclassical economics is that it implies that the legal system is the only institution relevant to contract enforcement. Its centerpiece, the general equilibrium model, implicitly assumes the operation of such a system and postulates that all mutually beneficial exchange is feasible. The identities of the individuals involved and the distance in time and place between the *quid* and the *quo* impose no restrictions on the decision to exchange.

Recent work in information economics indicates that asymmetric information and incomplete contracts limit the extent to which legal systems can facilitate exchange. Neoinstitutional economics argues that when this limitation implies efficiency loss, private order institutions will emerge to govern exchange (Williamson 1985). These institutions enforce the agreed-on exchange without relying directly on the legal system. Instead they rely on various contractual arrangements—such as sharecropping, ownership structure, arbitration and termination clauses, deferred compensation plans, and mandatory retirement—to provide the appropriate incentives for compliance. Thus neoinstitutional economics explicitly rejects the central role of the legal system posited by neoclassical economics, but it maintains that the legal system plays a large indirect role in governing exchange. (For further discussion of legal centralism see Williamson 1985, pp. 20–21, and Ellickson 1991, pp. 138–40.)

Empirical Evidence of Informal Contract Enforcement

Recent empirical studies indicate that many exchange relations in historical and contemporary markets and developing economies are not governed—directly or indirectly—by the legal system. Greif (1989, 1993) found that relations between

merchants and their overseas agents in the Mediterranean region during the late medieval Commercial Revolution were not governed by the legal system. Macaulay's (1963) survey of sixty-eight businesspeople and lawyers, mainly from Wisconsin, is among the best-known empirical studies of exchange that is not governed by the legal system. Macaulay found that "many, if not most, exchanges reflect no [contractual] planning or only a minimal amount of it" (p. 60) and that "disputes are frequently settled without reference to the contract or potential or actual legal sanctions" (p. 61). (For additional empirical studies of such "relational contracting," see Macaulay 1985 and Macneil 1985.) Similarly, international diamond traders in New York transact business without legally enforceable contracts (Bernstein 1992).

Contracting in the absence of law seems to be even more widespread in developing economies. In Peru 48 percent of the economically active population works in the informal sector, which is based on a "system of extralegal norms" (de Soto 1989). A survey of 224 randomly selected manufacturing firms in Kenya found that 58 percent borrowed from informal sources without legal contracts or collateral (Fafchamps and others 1993). In Russia informal exchange is extensive (Greif and Kandel 1995). Some of the most economically advanced sectors in some developing countries function without relying on the legal system. Indeed, exchanges in one of the most successful forms of capitalism in Pacific Asia—the family firms of Chinese immigrants—are not enforceable by the legal system (Redding 1995). Thus understanding the factors that determine the set of feasible exchange relations requires analyzing informal contract enforcement institutions, that is, institutions that ensure contract enforcement without relying on the legal system.

Foundations of Informal Contract Enforcement: General Morality, Personal Trust, and Reputation

Social scientists have provided empirical evidence that general morality, personal trust, and reputation underlie informal contract enforcement institutions. From an economic perspective, morality can be viewed as the effect of the actions taken to obtain a particular outcome on the utility associated with that outcome. The utility associated with a good that is stolen may differ from the utility associated with the same good if purchased. Moral values are acquired through a process of primary and secondary socialization that shapes an individual's preferences. Once particular values have been internalized, they affect behavior. The importance of such "first party enforcement" has been emphasized by Arrow (1974), who argues that "societies in their evolution have developed implicit agreements to certain kinds of regard for others . . . [agreements that] contribute greatly to the efficiency of its working. It has been observed, for example, that among the properties of many societies whose economic development is backward is a lack of mutual trust" (p. 26). More recently, Platteau (1994b) provides an extensive discussion of the transmission, evolutionary stability, and contribution to efficiency of general morality (see also Ellickson 1991, pp. 147–49).

By contrast, economic sociologists emphasize the importance of personal trust in informal contract enforcement. They argue that "the anonymous market of neoclassical models is virtually nonexistent" (Granovetter 1985, p. 495) and that personal trust based on social relations is the most important contract enforcement institution. The social content of personal relations "carries strong expectations of trust and abstention from opportunism" (Granovetter 1985, p. 490); people tend not to cheat people they know personally. According to this view the factors that make these expectations self-enforcing are part of human nature. An individual's utility depends on the approval of "those who know his behavior and communicate their feelings to him" (Holländer 1990, p. 1159). Hence social structures (or networks of relations) "are the structures that fulfill the function of sustaining order" in modern market economies because they structure personal relations (Granovetter 1985, p. 491). Economic efficiency is a function of a society's social structures because these structures and the personal trust embodied in them determine the extent of economic exchange (Smelser and Swedberg 1994).

To lend empirical support to these claims, economic sociologists point to the interaction between social and economic relations in both industrial and developing economies and the economic success of various "network" societies, particularly in East Asia (Hamilton 1994). Empirical evidence of the importance of personal trust is complemented by experimental evidence: even very short personal interactions between subjects in experiments substantially increased cooperation in situations in which conventional economic theory predicted that no cooperation would take place (Dawes and Thaler 1988).

Economists and political scientists have recently focused on the role of reputation in supporting informal contract enforcement. Reputation would induce an individual to forgo current gains from cheating in exchange, because cheating implies diminished future gains. Informal contract enforcement institutions based on reputation mechanisms can differ on several dimensions: whether the gains from exchange and the sanctions for reneging on a contractual agreement are economic or social, how gains and sanctions are enforced, whether personal or corporate reputation is at stake, whether punishment of a cheater is inflicted by the cheated party (a second-party enforcement, or bilateral reputation, mechanism) or by others (a third-party enforcement, or multilateral reputation, mechanism), how information on past conduct is transmitted (particularly the extent to which formal organizations, such as credit bureaus, rather than social networks transmit information), the mechanisms that make the threat of punishment following cheating credible, and the extent to which formal organizations with functions other than information transmission are required for the operation of a reputation mechanism (Greif 1992, 1994b; Platteau 1994a).[1]

Many studies have demonstrated the empirical relevance of reputation mechanisms. Adam Smith (1896 [1763], pp. 253–54) noticed their operation during the eighteenth century, and recent studies have examined reputation-based institutions in the context of late medieval Mediterranean and European long-distance trade (Milgrom, North, and Weingast 1990; Greif 1993, 1994a; and Greif, Milgrom,

and Weingast 1994). Macaulay (1963) stresses the importance of the second-party reputation mechanism among the businessmen he studied in Wisconsin. "The final type of non-legal sanction [that enabled exchange without legal contracts] is the most obvious. Both business units involved in the exchange desire to continue successfully in business and will avoid conduct which might interfere with attaining this goal" (p. 63). Portes (1994) notes that in Latin America the informal economy is based on "the [social] sanctioning power of the community [which] is a key force guaranteeing individual compliance" with contracts (p. 438). Fafchamps (1995) identifies the operation of the multilateral reputation mechanism in Zimbabwe.

Theory, History, and Governance of Anonymous Exchange by the Legal System

Contract enforcement institutions can be, and are, based on various mechanisms: the legal system, morality, personal trust, and reputation. Political considerations—namely, the ability and willingness of the state to provide efficient and impartial legal contract enforcement—substantially affect the relative efficiency of legal contract enforcement and its use. In Peru, for example, the high cost of using the legal system forced the adoption of informal means of governing exchange (de Soto 1989). Yet evidence that the legal system does not govern particular exchange relations in a particular time and place does not imply that it is not the most efficient general means of doing so. Indeed, a common implicit interpretation of the general equilibrium model is that it lends support to the importance of the legal system since it is this system that makes anonymous exchange feasible among *homo economicus*.

The claim that a legal system is required for a market economy in which exchange is not limited to nonanonymous exchange based on reputation and personal trust has gained support from historians, who have emphasized the centrality of an impartial and efficient legal system in the rise of Western market economies. Weber (1981 [1927]), for example, extolled the benefits of such a legal system by arguing that "the capitalistic form of industrial organization . . . must be able to depend upon calculable adjudication and administration [of the law]" (p. 277). More recently, North (1991) notes that the "technology associated with the growth of manufacturing [in the nineteenth-century United States] . . . required effective factor and product markets. Underlying such markets are secure property rights, which entail a polity and judicial system to permit low-cost contracting" (p. 101). Clearly, these theoretical and historical arguments regarding the need for an effective legal system do not preclude supplementing it with other contract enforcement institutions. Platteau (1994b) recognizes the complementarity among anonymous exchange, the legal system, and general morality and argues that the rise of the West is attributable to "the emancipation of the individual from old ties of social and political allegiance . . . and the development of generalized morality in which abstract principles or rules of conduct are considered equally applicable . . . beyond the narrow circle of personal acquaintances" (p. 802).

Yet these theoretical and historical arguments in support of the notion that the legal system is required for a market economy in which exchange is not limited to nonanonymous exchange based on reputation and personal trust are flawed. First, neoclassical economics implicitly assumes that institutions other than the legal system can support anonymous exchange. Second, the historical experience of the West reveals the importance of institutions not based on the legal system in supporting anonymous exchange and suggests that these institutions have at times been more efficient than exchange based on bilateral, legally enforceable contracts.

To see why the implicit assumptions of neoclassical economics do not imply that only the legal system can support anonymous exchange, assume that anonymous exchange is indeed efficient (and that personal trust and reputation cannot govern exchange). For a legal system to govern anonymous exchange, information about the transacting parties must be available at the society level. But if such information is available, the legal system is not necessarily the only contract enforcement institution that can take advantage of this information to facilitate anonymous exchange—institutions based on reputation can take advantage of this information as well. The assumptions of neoclassical economics are thus consistent with anonymous exchange that is governed by contract enforcement other than the legal system.

Milgrom, North, and Weingast (1990) show that theoretically the key to facilitating anonymous exchange is not the coercive power of the court but the availability of information at the society level. In their model the ability of the party with information at the society level to activate a multilateral reputation mechanism provides the transacting individuals with incentives to uphold their contractual obligations. Empirical evidence indicates that reputation-based institutions did and still do enhance contract enforcement among (otherwise) anonymous individuals in Western market economies by generating information, creating identities, relying on a third party (other than the court), and constraining the ability of the transacting parties (or their agents) to act opportunistically by using various bureaucratic procedures, regulations, and incentive schemes. (In some informal contract enforcement institutions reputation and legal contract enforceability complement each other; see Greif 1994b.)

To grasp the nature of such institutions, consider the New York Stock Exchange (NYSE), a voluntary, for-profit association established in 1792 that facilitates exchange among anonymous buyers and (in the absence of the NYSE) anonymous sellers of stocks. Reputation considerations motivate the NYSE to supervise information disclosure by firms and motivate brokers to follow the procedures it sets—procedures aimed at curtailing firms' ability to act opportunistically in selling stocks. Other examples of institutions that establish rules and resolve disputes are the trade associations prevalent in the United States (Bernstein 1995). Credit bureaus and credit card issuers facilitate a credit market among anonymous economic agents by providing appropriate information and creating an intermediary (Klein 1992). Brand names, escrow accounts, external public auditing, better business bureaus, credit rating agencies, and accreditation of professionals are additional examples of such institutions.[2] Many such institutions emerged in the United States during the

second half of the nineteenth century, when the economy became increasingly dominated by market forces (Zucker 1986). Such institutions may, in fact, have fostered the transition to a market economy. The extent to which these mechanisms were and are used in an economy with a relatively well-developed legal system suggests that they enhance efficiency relative to exchange based exclusively on legally enforceable bilateral contracts.

Toward a Theory of Systems of Informal Contract Enforcement

Since contract enforcement can and does take place in the absence of a legal system, is the legal system needed at all for contract enforcement? Is it an important contract enforcement institution? Does informal contract enforcement always dominate, in theory and in practice, a (feasible) legal system?

Various Informal Contract Enforcement Systems and Their Costs

There is a perception among social scientists that the legal system is less efficient than informal contract enforcement. Some economists hold that reputation provides the "free lunch" of contract enforcement without cost, even in anonymous exchange. Ellison (1994) demonstrates that under relatively weak restrictions, even in a repeated, random-matching prisoners' dilemma game in which players are unable to recognize their opponents, there is an equilibrium in which cooperation is sustained through third-party "contagious" punishment (cheating by one player induces both players to cheat subsequent partners). Some economic sociologists argue that exchange is generally nonanonymous and governed by personal trust rather than by the legal system. Personal trust is their free lunch, although it is not always satisfying. According to this view the relative efficiency of a society's informal contract enforcement institutions depends on its exogenously determined social structures or networks, within which personal trust prevails (Granovetter 1985).

These different views among economists and sociologists underlie distinct interpretations of the relative efficiency of existing systems of contract enforcement. Social psychologists have identified two such systems, the individualist and the collectivist (Bellah and others 1985; Reynolds and Norman 1988; and Triandis 1990). The individualist system that prevails in the West is characterized by relatively low levels of personal trust (associated with an integrated social structure in which individuals frequently shift from one group to another), relatively high levels of general morality, an effective legal system, and the prevalence of reputation mechanisms.

In contrast, the collectivist system that prevails in most other countries is characterized by informal contract enforcement based mainly on personal reputation and trust. The social structure is segregated in the sense that individuals interact socially and economically mainly with members of specific religious, ethnic, or familial groups. Neither the legal system nor other organizations play a large role in contract enforcement. At the same time there is little cooperation between members of different groups, and general morality is relatively low, as indicated by several studies

showing that Asian businessmen are perceived by their peers as being more opportunistic than their Western counterparts (Pyatt and Redding 1995; Yamagishi and Yamagishi 1994). (Greif 1994a documents relative collectivism and its economic impact in the premodern Islamic world; Hamilton 1991 does the same for premodern China.)

The costs associated with reputation-based enforcement in the individualist system are transparent. Contract enforcement requires various organizations (such as credit bureaus and trade associations) that are costly to operate. The theoretical need for such organizations in an individualist system has been demonstrated by several studies (Milgrom, North, and Weingast 1990; Greif, Milgrom, and Weingast 1994). Empirically, Wallis and North (1986) found that transactions costs that include the cost of contract enforcement increased in the United States during the period in which the informal contract enforcement institutions discussed earlier emerged.

In contrast, in the collectivist system the cost of organizations is low and reputation and personal trust seem to provide a free lunch. Can it then be concluded that collectivist informal contract enforcement always dominates the legal system? Are collectivist societies more efficient than individualist ones? Do the nature and effectiveness of a society's contract enforcement system reflect only economic factors? Based on the experiences of economically prosperous collectivist economies such as Japan, Singapore, and Taiwan (China), some social scientists have argued that collectivist informal contract enforcement is more efficient than the individualist system. Yet the fact that most collectivist societies are low-income economies and most individualist societies are high-income economies suggests a need to analyze the factors underlying the relative efficiency of the collectivist informal contract enforcement system, the legal system, and the individualist system. Particularly in need of analysis are the factors influencing the extent to which reputation mechanisms (without facilitating organizations) can and do govern exchange efficiently, the factors determining social networks (and hence personal trust), and the factors determining a society's level of general morality.

The model presented here examines these factors, considers a possible process leading to the development of individualist and collectivist systems of informal contract enforcement, and identifies the factors determining their efficiency relative to each other and to the legal system. The analysis begins by considering only economic motivations and the relations between reputation-based informal contract enforcement and social structures. It is then extended to incorporate personal trust, general morality, and the legal system.

Endogenous Contract Enforcement Systems and Their Associated Social Structure

Exchange relations require contract enforceability because at least one party can cheat the other. Since (in the absence of appropriate institutions) exchange is always sequential—that is, some time elapses between the *quid* and the *quo*—exchange is

modeled as a one-sided prisoners' dilemma game in which only one party can cheat the other. In this game one economic decisionmaker (the principal) can offer another (the agent) a contract specifying an exchange and a share in the surplus generated by the exchange (a wage). By acting opportunistically, the agent can earn more than the highest wage the principal finds profitable to offer. This specification captures the essence of exchange between workers (including managers) and their employers, buyers and sellers of goods (whose quality is revealed only after the purchase), and lenders and borrowers.

Since the agent can act opportunistically, the principal would not initiate an exchange (even if it were efficient) in the absence of a contract enforcement institution. Reputation can provide such an enforcement institution to facilitate exchange. Suppose that the exchange can be repeated each period and that the principal offers the agent a per period wage higher than the (time average) per period income available to the agent if the agent is fired.[3] Suppose also that the principal threatens never to exchange with the agent again if the agent acts opportunistically. If the short-run gain from cheating is less than the long-run gain from being honest, the agent will not act opportunistically.

Now suppose that there are many principals and agents. The number of potential agents is larger than the number of principals, and an agent who does not exchange receives the reservation utility of zero. (These assumptions enable principals to punish particular agents by not hiring them.) Suppose also that exchange is uncertain in the sense that because of the principal's changing needs there is some per period probability that exchange with an agent will be terminated even though the agent is honest.

The agent's wage must be high enough to induce honesty and thus enable exchange. The lowest wage required to keep an agent honest depends on both the potential gains from acting opportunistically and the threat of second- or third-party (collective) punishment. It is easy to see why a principal would be expected to retaliate against a dishonest agent. When there is an equilibrium in which second-party punishment governs exchange, the wage paid to agents and the agents' unemployment rate are such that no hired agent finds it optimal to cheat. If a principal did not fire a dishonest agent, that agent would find it optimal to cheat again unless the agent's wage were higher than the equilibrium wage. But since the cheated principal can hire another agent from the pool of unemployed agents at the equilibrium wage, the principal's best course is to fire the cheater and hire another agent.

But why would a principal who had not been cheated find it optimal not to hire an agent who had cheated another principal? Clearly, if past behavior is an indication of future behavior, the credibility of collective punishment is ensured. But this explanation for collective punishment fails to take into account the segregation common in collectivist societies. It implies that exchange between groups is limited only by transmission of information about past actions. Such information is embodied in patterns of exchange within a group, since exchange with a particular member indicates that the other members of the group found that member to be honest. Such repeated interactions are easily observable, however, implying that if this type

of information were the key to patterns of exchange, exchange would not be confined to social structures.[4] Accounting for segregation requires focusing on another source of credible collective punishment.

Why should a principal participate in third-party enforcement even if agents are identical? A third-party principal would find it optimal to retaliate against an agent who had cheated another principal because collective punishment is self-enforcing. As long as the relations between a principal and an honest agent can be terminated, the equilibrium wage that principals must pay their agents is lower under third-party enforcement than under second-party punishment because collective punishment increases the penalty for cheating—the agent will not be hired in the future by any principal. Hence, other things being equal, the expected cost of cheating is higher under collective punishment, and a lower wage is therefore sufficient to induce honesty. The wage required to ensure the honesty of an agent who is not expected to be hired by other principals is higher than the wage required to ensure the honesty of an agent who is expected to be hired elsewhere. Collective punishment is thus self-enforcing: if every principal expects other principals to practice collective punishment, it is optimal for each to do so.

The expectations for intragroup collective punishment have profound implications for the formation of social structures. To grasp why, suppose that there are two identical groups of principals and agents that in the past could not exchange (or among whom it was not efficient to exchange). Suppose also that exchange has become (technically) possible and efficient. Will principals from one group hire agents from the other, or will the society remain segregated?

Collective punishment practiced within a group creates a wedge between efficient and profitable exchange relations that reduces the profitability of intergroup exchange and can thus lead to a segregated society in which efficient intergroup exchange does not take place. If there is some doubt whether collective punishment governs intergroup exchange, a principal initiating an intergroup exchange will have to pay a wage higher than the intragroup wage to keep an agent honest. The wage will be higher because any uncertainty about collective punishment in intergroup exchange reduces the probability that an agent who cheats in such relations will be punished, thereby increasing the wage required to keep him honest. Since the principal's cost of initiating intergroup exchange is higher than the cost of initiating intragroup exchange, only intragroup exchange will be initiated, and segregation will result. Principals will initiate intergroup exchange only if the efficiency gains are sufficiently large.

This is not the case, however, when intergroup exchange becomes possible between two groups that practice second-party punishment. In that case expectations do not create a wedge between profitable and efficient exchange relations, implying the emergence of an integrated society in which (efficient) intergroup exchanges are initiated. Although uncertainty over punishment is likely to exist, the intergroup and intragroup optimal wages are identical. Hence any efficiency gains from intergroup exchange will motivate principals to initiate such exchange.

These results are strengthened when the assumption that agents' past actions are common knowledge is relaxed. Suppose that before the game begins each principal

can either invest or not invest in an information network and that the action the principal takes is common knowledge. By investing in the network, the principal bears the cost of information gathering for each period, in return for which the principal is informed of the histories of all the other principals who also invested in information acquisition. Principals who do not invest in the network know only their own history.

When the decision to acquire information is endogenous and collective punishment is expected, agents' histories have value and principals are motivated to invest in gathering information. The wage required to keep an agent honest is a function of that agent's history because an agent who cheated in the past will cheat if rehired and paid the going wage. Similarly, if an agent is hired by a principal who is not attached to the information network, the agent will cheat the principal unless paid more than the going wage. The availability of information, in turn, makes collective punishment possible. Note that in the absence of expectations of collective punishment, an agent's wage is not a function of his or her history, and principals are not motivated to invest in information networks. But if principals do not gather information, collective punishment cannot be expected to be practiced. (Note, however, that a collectivist enclave may persist within an individualist society with second-party enforcement because of the very large benefit of collective punishment to the members of the collectivist group.)

Personal Trust, Social Sanctions, and Moral Considerations

How can personal trust, social sanctions, and moral considerations be integrated with the analysis? A collectivist society is characterized by self-enforcing collective punishment, segregation, and an intragroup social communication network; the social structure is thus conducive to fostering personal trust (Granovetter 1985). Furthermore, since economic exchange is relatively frequent among members of a collectivist group, it may foster informal contract enforcement based on social rather than economic sanctions. A mechanism that links patterns of economic interactions and social sanctions was proposed by the sociologist Homans (1950), who argued that frequent (economic) activity carried out among particular individuals increases the level and intensity of "friendliness" among them, leading them to further interact socially and economically. This mechanism implies that a relatively high level of friendliness would emerge among members of a collectivist group. Friendliness, in turn, is likely to facilitate further informal contract enforcement by enabling group members to retaliate for economic opportunism by imposing social sanctions. (This is an example of transactions being linked to enhanced contract enforceability; see Greif and Kandel 1995.)

The distinct social and economic patterns of interactions in individualist and collectivist societies are also likely to affect moral enforcement mechanisms—that is, enforcement based on the tendency of individuals to derive utility from acting according to their values (first-party enforcement). One of the processes through which individuals develop a value system is their attempt to find moral justification for their behavior (Davis 1949; Homans 1950; and Sugden 1989); this view of cul-

ture as a "legitimizing" mechanism is fundamental to the theories of Karl Marx and Emile Durkheim. Hence different patterns of social and economic interactions are likely to lead to the development of distinct value systems. In particular, the pattern of social and economic interactions in an integrated individualist society is likely to lead to the development of general morality, the notion that people should act in a trustworthy manner toward all individuals. The pattern of social and economic interactions in a segregated collectivist society is likely to lead to the development of personalized trust, the notion that one should act in a trustworthy manner only toward individuals one knows.

Origin and Relative Efficiency of Individualist and Collectivist Informal and Legal Contract Enforcement

The model indicates the interrelations between the economic and noneconomic features of a society and its informal contract enforcement institutions. Particular expectations about punishment and group membership, which are of cultural rather than economic origin, determine whether second- or third-party enforcement governs exchange and define the initial boundaries within which third-party enforcement is practiced. Once crystallized within a society at a point in time, these expectations—together with the historical process through which various groups began interacting with each other—may have a profound impact on the evolution of the social structures, level of personal trust, and value systems of the society, further influencing its system of contract enforcement. Hence an observed system of informal contract enforcement is not necessarily efficient.

If the appropriate expectations are lacking, even the relatively simple second-party enforcement mechanism may fail to arise. Szanton (1971) observes that the common belief in Estancia, the Philippines, was that "people are primarily concerned with their own personal welfare or short-term benefit and will take advantage of a situation of trust, whenever possible." There is a "pervasive mistrust of others fortified by innumerable accounts of economic double-dealing in the town. Even within a nuclear family, when it comes to business matters, trust between individuals is often limited. The expectation that others work only for their own advance is so powerful that even partnerships between kinsmen or close friends rarely last more than a few weeks or months"(p. 89).

Once expectations associated with second-party enforcement take hold, they are difficult to alter. In Ghana, for example, past conduct "plays little role in identifying reliable clients. There seems to be no mechanism whereby information about clients' trustworthiness is shared among firms. . . .When prompted directly, most firms declare that they never bother warning another firm about a particular untrustworthy customer. . . . Sharing valuable information with competitors would provide them with an undue advantage, and no effort is made in that direction by most respondents" (Fafchamps 1993, pp. 16–17).

Note that such second-party enforcement may be very inefficient, since many efficient exchange relations would not be profitable enough to be initiated. When

only second-party punishment is inflicted following cheating, the wage that the principal has to pay to keep the agent honest is higher than that required when the agent is subject to collective punishment. Thus the principal will find it profitable to hire agents only in situations in which the efficiency gains from exchange are relatively high. This implies that a legal system, even if it only partially limits cheating, may substantially increase the efficiency of contract enforcement. Thus informal contract enforcement based on second-party punishment is more likely to emerge and survive in a society with an effective legal system. Collectivism is more likely to emerge and survive in a society without such a system.

The differences between East and West are noteworthy in this regard. In Europe, following a long tradition of Roman and German law, a legal system was developed to govern economic exchange, and medieval Christianity placed the individual and general morality, rather than the social group, at the center of its theology. As a result what prevails in the West is an individualist system of contract enforcement based on legal enforceability, general morality, second-party enforcement, and formal organizations to support exchange (see Greif 1994a). By contrast, Confucianism in China emphasizes responsibility among relatives, while the "the state has not traditionally maintained order by jurisdiction [and] civil law was not developed" (Redding 1990, p. 45). The collectivist system, which seems to have emerged in Imperial China (Hamilton 1991), continues to dominate in the Chinese business community today.

Although collectivism is an alternative to the individualistic system, there is no reason to believe that it should dominate such a system. In collectivist systems group boundaries are determined by the process through which the initial expectations regarding group membership were established, and this process does not necessarily imply efficient group size. Once these expectations are established, each member has an incentive to exchange within the group, and the group as a whole has an incentive to prevent intergroup exchange. The functioning of informal contract enforcement within the group depends on not punishing a member for cheating a nonmember. Unless this is the case, members can establish intergroup exchange relations based on their intragroup reputation. Doing so, however, undermines the intragroup reputation mechanism, since it is based on expected intragroup exchange.

Chinese and Japanese immigrants to the United States "erected an elaborate structure of organizations [for mutual assistance] based on the exact locale of emigration. These organizations were ascriptive . . . [and] the obligations owed to one another by those from the same home area were matched by the reluctance of those from other areas to help" (Granovetter 1994, p. 141). Similarly, Asian populations in Kenya are segregated in business communities in which collective punishment is exercised. These communities distrust each other about as much as they distrust non-Asians (Fafchamps 1993).

Note, however, that in contrast to the assumption in economic sociology that group boundaries are exogenous to contract enforcement, the model outlined here indicates that these boundaries may yield to economic pressure. When the efficiency

gains from transacting outside the group are large enough, the group will dissolve. Indeed, this is the process that Zucker (1986) describes as occurring in the United States starting in the mid-nineteenth century. The large gains from intergroup exchange and from exchange over large geographical areas led to disintegration of informal contract enforcement within relatively small groups. Schmitz's (1995) study of an industrial district in the Sinos valley in Brazil reveals that industrial growth increased the gains from transacting outside the original industrial group within the valley, leading to a decline in intragroup informal enforcement. Kennedy (1988) describes how members of the Fra Fra community in northern Ghana left their collectivist community for the city once the gains from acting individually increased enough.

An important conclusion, then, is that theory alone cannot predict which system of societal organization—the integrated individualist system based on second-party enforcement, formal organizations, general morality, and legal enforceability, or the segregated collectivist system based on personal trust and third-party enforcement—is more efficient in promoting exchange. "The collectivist system is more efficient in supporting [intragroup] agency relations and requires less costly formal organization (such as law courts), but it restricts efficient [intergroup] agency relations. The individualist system does not restrict [intergroup] agency relations but is less efficient in supporting [intragroup] relations and requires costly formal organizations" (Greif 1994a, p. 942).

The analysis provides a framework for examining the economic, social, and cultural factors that determine the relative efficiency of governing exchange by each type of system. Each of these sets of factors is examined in the following sections.

Economic Factors

Suppose that a legal system exists but that it is costly to use and lacks relevant information regarding actions. In such a case the relative efficiency of governing exchange through a collectivist system or a legal system depends on the nature of relevant potential exchanges, the extent of segregation, and the cost of operating the legal system. If intergroup exchange does not yield efficiency gains and collectivism can support all efficient intragroup exchange, then collectivism imposes no costs and is more efficient than the legal system. But when segregation prevents efficient intergroup exchanges from being initiated, collectivism may be less efficient than the legal system.

The relative efficiency of second- and third-party reputation mechanisms depends on the gain from principal-agent relations and the agent's reservation utility. To prevent opportunism, agents have to receive a high enough wage relative to their reservation utility. The principal would initiate relations with an agent only if this wage premium were lower than the total gain from exchange. Although there may be transactions in which exchanges are efficient—that is, the total gain is positive—they will not be initiated based on informal contract enforcement if the wage premium is too high. Thus the relative efficiency of such informal contract

enforcement depends on the relationship between the wage premium and the total gain from exchange. This, in turn, may depend on the structure of the economy. Informal contract enforcement is more efficient in dual economies in which an agent's reservation utility is a very low agricultural wage, for example. It is less efficient in economies in which agents can find lucrative alternative work relatively quickly.

The relative efficiency of informal contract enforcement also depends on the importance of human capital that is specific to relations between a particular principal and agent. Suppose that repeated employment by the same principal allows an agent to acquire human capital specific to that relationship. Suppose also that exchanging through a single agent is efficient. After acquiring the human capital the agent may act opportunistically and expropriate a much larger share of the surplus generated by the exchange. Anticipating this behavior, the principal may be deterred from entering into exchange relations to begin with. Williamson (1985) examines how parties can mitigate this contractual problem by using private institutions that rely indirectly on the legal system. This may not be an option in a society in which exchange is governed by informal contract enforcement, and the principal may have to resort to (inefficiently) employing several agents to curtail each agent's ex post ability to act opportunistically. By hiring several agents, the principal can lend credibility to the threat to terminate relations in case of opportunistic behavior. Indeed, de Soto (1989) observes that in Peru "owners of informal businesses tend to diversify their sources of supply and sale markets more than [do] formals" (p. 164).

Collectivism is more efficient than a legal system in situations in which an agent's actions are observable but unverifiable by the legal system, since observability is sufficient for the operation of collective punishment but verifiability is required for the operation of the court. Ellickson (1991) points out that disputes among whalers regarding ownership of a whale that had drifted away from the ship whose crew had killed it could not be resolved by the court. Informal third-party contract enforcement governed these disputes in an efficient manner.

For other information structures, however, collectivism is not necessarily more efficient than a legal system. Suppose that there is asymmetric information about an agent's ability to execute a contract—the ability is known to the agent but not to the principals and the legal system. Under collectivism an agent can be motivated to reveal this ability only by the threat of terminating future relations. This threat is not effective, however, in situations in which revealing one's ability means not being able to exchange again. Hence informal contract enforcement may lead to inefficiency because of its failure to match principals and agents with the appropriate abilities. A legal system can provide agents with additional means of revealing their ability in the form of a penalty for failure to execute a contract. De Soto (1989) notes that in Peru "contracts which are legally enforceable lend credibility to people as pledges . . . [and] the mere fact that contracts can be enforced encourages the parties to make reasonable commitments that they can fulfill" (p. 164). Clearly, in a collectivist society social (rather than economic) sanctions can substitute for the court (see Besley 1995b for a theoretical discussion and Udry 1994 for an empirical study

of northern Nigeria). Yet reliance on social sanctions enhances potentially inefficient segregation.

The observation that the limited scope of social sanctions can restrict exchange reflects an important deficiency of collectivism, namely, its limited flexibility in the face of changing situations. Social sanctions among particular individuals, information revealed through past actions (rather than through legal contracts), segregation and associated expectations, and personal trust are sunk in particular relations. This investment is both sunk and specific to the particular individuals involved. As de Soto (1989) notes with respect to Peru, business people "invest time, effort, and money in cultivating long-term friendships" (p. 165). This investment cannot be recovered if the relations are terminated, and it cannot be used in exchanges with others, implying that even if a more economically efficient exchange opportunity arises, the principal may find it more profitable to exchange with the agent with whom an investment in relations has been made. Fafchamps (1993) notes that in Kenya "the absence of an information sharing mechanism [implies that even] established firms also find it difficult to shift their activities to respond to changes in relative prices" (p. 77). By contrast, adjustments under the legal system can be made instantaneously.

Social and Cultural Factors

The relative efficiency of collectivism also depends on social and cultural factors. In particular, social and cultural factors are likely to determine the initial boundaries of the groups that exercise collective punishment. In collectivist societies such groups are often defined by a particular social or cultural feature. In Africa, for instance, merchant communities often share a particular religious faith—Islam (Ensminger 1992) or, more recently, Evangelicalism (Poewe 1989). In the Asian community in Kenya caste and religion serve as dividing lines between business communities (the Shahs and Patels, the Sikhs and Ismaelians, and so on) within which informal contract enforcement prevails (Fafchamps and others 1993). The relative efficiency of informal contract enforcement based on collective punishment thus depends on the extent to which these social and cultural factors happen to lead to segregation that is economically efficient.

The nature and role of the family also affect the efficiency of basing exchange on collectivism. Continuity in relations is crucial to the effectiveness of personal reputation mechanisms; hence people with short time horizons may be more willing to cheat. In Ghana, for example, "most of the reported cases of non-payment . . . belong to this category of good business relations turned sour" because of a customer who died, went bankrupt, moved to another city, or retired without paying his bills (Fafchamps 1993, p. 18). This "end of the game" problem implies that it may be impossible to exchange to begin with.[5]

The actual severity of this problem in a particular society depends on the role and nature of the family. Suppose that intrafamily relations are such that children support their parents in their old age if they are able to do so. If children are considered morally responsible for their parents' business dealings, a parent's temptation

to cheat in old age will be reduced. Similarly, a group that practices informal contract enforcement based on collective punishment has further leverage in deterring opportunistic behavior when it also provides its members with safety nets in case of illness, old age, or bankruptcy. In China—where the culture emphasizes intrafamily obligations and the family's continuity, and the extended family traditionally provides the locus of information, safety nets, and employment to its members (Hamilton and Biggart 1988; Hamilton 1990)—collectivism is effective in governing intragroup exchange in the business community.

Business Organization

Although the issue is outside the scope of the model, the relative efficiency of collectivism also depends on the optimal form of business organization—in particular, whether a large organization with multiple layers of management is efficient. The optimal form of business organization depends on many time- and location-specific factors, such as production and monitoring technology, cost of capital, flexibility, scale, and uncertainty (Milgrom and Roberts 1992; Aoki 1988). Particular forms of business organization may be more efficient under particular circumstances. Though the large corporations that emerged in the United States toward the end of the nineteenth century seem to have been more efficient than the small craft production systems they replaced, large firms were unable to compete with small firms in twentieth-century Italy, for example.

To the extent that it constrains the selection of forms of business organization, informal contract enforcement may affect efficiency. Kennedy (1988) notes that collectivism among the successful Hausa traders of northern Nigeria "may increasingly inhibit the emergence of more advanced types of business behavior in the future" (p. 146). In particular, the choice of forms of business organization may be restricted because of a limited ability to hire managers. Hamilton (1990) and Redding (1990) note that key managerial positions in Chinese business firms are often entrusted to family members. Clearly, the extent to which such organization implies efficiency loss (if at all) depends on the optimal form of business organization.

Customary Law and Informal Contract Enforcement

It has been argued that flexibility is one of the efficiency-promoting attributes of informal contract enforcement. In exchange situations in which agents have to make decisions, informality implies that their choice of actions is not restricted by a legal contract. Benson (1989) argues that the informally enforced code of conduct that governed long-distance European trade during the late medieval period was adjusted to optimally meet the needs of that trade and that the later codification of this informal code of conduct was inefficient because it constrained flexible adjustment. Ellickson (1991) shows how the social norms that governed relations among whalers between 1750 and 1870 were adjusted in a wealth-promoting manner to the changing needs of the industry.

Ellickson's distinguished study notwithstanding, there are sound theoretical reasons to doubt whether flexible adjustment of informal codes of conduct always occurs in a collectivist society. For collective punishment to be effective, individuals should share a perception of what behavior triggers punishment. In any exchange in which agents have to make a decision, their behavior is guided by neither the actual profit it will entail for the principal nor the principals' perceptions regarding the appropriateness of the action. Rather, their behavior is guided by their expectations of the principals' perceptions of their actions. The agent may therefore prefer actions that were taken in the past without incurring punishment to more efficient actions that were not taken in the past. Agents' need to consider others' perceptions of their actions under collective punishment can hinder the efficient adaptation of customary law and may therefore make informal contract enforcement less efficient in situations requiring rapid adjustment (Greif 1993). Thus collectivism may entail an ex post cost of maladaptation, while reliance on legal contracts imposes a high ex ante cost of contracting, since firms tend to perform only to the letter of the contract. Once again, the optimal contract enforcement institution is case specific.

Policy Implications

In summary, a society's contract enforcement institutions—formal and informal alike—are a product of a historical process in which economic, social, cultural, and political factors interact. Since the relative efficiency of a particular system depends on the economic environment as well as these other factors, there is no reason to expect a particular system in a particular society to be economically efficient. Indeed, the empirical evidence indicates that some societies seem to have relatively efficient contract enforcement systems while others do not. Thus there is considerable scope for policy aimed at enhancing contract enforcement in general and informal contract enforcement in particular.

Actual policies aimed at enhancing either legal or informal contract enforcement must be country- and time-specific. The general policy considerations proposed here are based mainly on theoretical insights and the historical experience of the West.

Taking Advantage of Existing Informal Contract Enforcement

Economic and social collective punishment within a group provides a means of enforcing contracts entered into by members of the group with members of other groups. This method was used to facilitate exchange in Europe as early as the medieval period, when every member of certain trading communities bore responsibility for the contractual performance of all members of the community. If a trader failed to pay a debt to a creditor from another town at the specified place and time, the goods of any member of the debtor's community could be seized as compensation. The trader whose goods were confiscated was sent to recover the losses from the original debtor. Traders used intracommunity enforcement mechanisms to support intercommunity exchange (Greif 1995).

Taking advantage of intragroup contract enforcement may be particularly useful for improving credit and risk sharing in low-income countries (Besley 1995b). In the Grameen Bank group lending program in Bangladesh, loans are made to individuals but groups of borrowers assume some joint liability. The success of the program has been attributed to the information available to the group members and to the incentives provided by joint liability. As Besley (1995a) shows, under certain conditions a sufficient condition for improving repayment rates is the ability of the group to impose social sanctions on its members.

History provides other examples in which informal contract enforcement has been used to attract credit from outside sources. In nineteenth-century Europe new agricultural techniques to improve productivity required capital investment that was beyond the means of most peasants. Banks were reluctant to provide credit because of the high cost of monitoring such investment, as well as moral hazard problems. To circumvent this problem, credit cooperatives were established in Germany to borrow from banks and lend to their members. The information, monitoring ability, and informal contract enforcement mechanisms available to members of small agricultural communities, as well as the incentive structure of the cooperatives, enabled them to succeed where bank lending supported by the legal system had failed (Banerjee, Besley, and Guinnane 1994). Yet as the failure of similar cooperatives in Ireland indicates, a cooperative's success depends on local conditions (Guinnane 1994).

Improving the Operation of Informal Contract Enforcement

Policy can attempt to increase the efficiency of informal contract enforcement institutions by improving such enforcement within the originating group, by improving intergroup enforcement, and by promoting organizations that foster informal contract enforcement in general.

INFORMAL CONTRACT ENFORCEMENT WITHIN GROUPS. Inefficiencies may result when collective punishment governs relations among members of a group. The group may be much smaller than optimal, it may lack well-established procedures for resolving disputes, or it may lack a well-defined code of conduct that coordinates expectations about which behaviors should be subject to collective punishment. (For a discussion of this last deficiency in Peru, see de Soto 1989.)

Because these deficiencies result from a lack of coordinated expectations, they can be overcome to some extent by establishing an organization to communicate with its members. The benefit of such an organization is suggested by the success of organizations that foster informal contract enforcement among Japanese and Chinese immigrants in the United States and Southeast Asia. By facilitating internal communication, such organizations allow expectations about who is a member and who is not to be coordinated, especially where new members of the group are concerned. A coordinating organization can also facilitate dispute resolution by defining and supervising an appropriate set of procedures and by making the decisions

of this dispute resolution mechanism legally binding. Such a procedure was adopted in England during the late medieval period. In the case of disputes between two merchants, other merchants were called on to pass judgment, which was enforced by the law.

A coordinating organization can also foster the establishment of a code of conduct by determining standards of behavior and quality. The benefit of such an organization is suggested by the experience of the early U.S. automobile industry. Members of the industry established an informal organization that determined its standards. Standards were established by automobile engineers, who made decisions based solely on technical merit, rather than by producers, whose interests might have distorted their choice of standards.

A coordinating organization may also be important in situations in which informal contract enforcement governs the relations between one large economic player and many small players, as in the case of a large firm with many small suppliers. In this case the large firm may be unable to commit to its contractual obligations to the small suppliers at the efficient level of trade, even if its reputation is at stake. (For an analysis of this issue in the context of medieval trade, see Greif, Milgrom, and Weingast 1994.) Creating a formal organization to coordinate suppliers' responses to the firm's behavior and to enforce its decisions on all suppliers may be Pareto optimal. Toyota did just that. (In this case the principal seems to have had to guarantee that the agents not take advantage of their relation-specific investment.)

IMPROVING INFORMAL CONTRACT ENFORCEMENT BETWEEN GROUPS. If segregation is one of the main costs of informal contract enforcement, facilitating contracting between groups should be a policy priority. Parties can overcome the deficiencies of segregation by using the court, particularly when exchange is profitable and the nature of the relevant information is such that the legal system can adjudicate it (that is, when conduct is easily verifiable, as, for example, with respect to credit). In many cases this is not an option, however, because the sums involved are too small or the court is ineffective. When the sums are too small to justify the cost of a regular court, establishment of a small claims court should be considered. One approach—adopted in premodern England—would be to allow litigants to choose which court to turn to, and to pay courts by the case. Alternatively, it may be possible to establish regulations that foster contract enforceability in intergroup exchange, such as having the provider of credit for the purchase of a particular machine own the machine until full payment is made.

When it is not possible to use the court, intragroup organizations can foster intergroup cooperation by providing information and coordinating expectations about responses to misconduct. Establishment of such organizations is difficult, however, because individual members of distinct groups (or of no group in the case of second-party reputation) have to be induced to provide information that can be useful to others, including potential competitors. Milgrom, North, and Weingast (1990) present the best analysis of inducing individuals to provide proprietary information to a third party that, in turn, would use it for informal contract enforcement. They

show how a "court" without coercive power can foster contract enforcement among individuals who do not expect to trade more than once. Suppose that each pair of individuals is matched only once and that each individual knows only his or her own experience. Assume that the court is capable only of verifying past actions and of keeping records of individuals who cheated in the past and that acquiring information and appealing to the court is costly for each individual. Despite these costs a symmetric sequential equilibrium exists in which cheating does not occur and individuals are induced to provide the court with the information required to support cooperation. It is the court's ability to activate a multilateral reputation mechanism by controlling information that provides the appropriate incentives. Hence a court can ensure contract enforcement over time even if it cannot use coercive power against cheaters.

This analysis is of limited applicability, however, since it assumes that the extent to which contract enforcement is achieved by a particular pair does not reduce the profitability of exchange by any other pair. In other words it assumes that competition does not exist. By assumption, the model also ignores the complexity introduced by having traders select their trading partners.

Intergroup exchange can also be fostered by having an individual from each group handle all intergroup exchange. If exchange between these individuals is expected to be sufficiently frequent, intergroup cooperation can be sustained based on the operation of a reputation mechanism between these individuals. Clay (1995) notes the operation of such a mechanism in California in the early nineteenth century. Exchange, particularly credit relations, between the native communities (within which contract enforcement was based on personal trust) and the American merchants (among whom economic third-party enforcement sustained exchange) became possible once a particular merchant settled in a community and became part of the indigenous system of contract enforcement.

Similar phenomena have been observed in contemporary developing countries. Hayami and Kawagoe (1993) note that in rural Indonesia intermediaries with access to both the informal contract enforcement system within villages and formal contract enforcement promote efficiency. Affluent intervillage collectors (of agricultural products) raise capital in the formal credit market (by using their assets as collateral or by pawning), which they then lend to smaller collectors (who have no access to formal credit), thus facilitating exchange. Similar arrangements in which intermediaries use two systems of contract enforcement have been observed in contemporary Ghana (Fafchamps 1993).

ORGANIZATIONAL SUBSTITUTES FOR AND COMPLEMENTS TO INFORMAL CONTRACT ENFORCEMENT. Various types of organizations can enhance informal contract enforcement. In developing countries consumer watch groups and better business bureaus can provide consumers with objective information on past actions and quality. The limitations of such institutions should be recognized, however. Such institutions have been successful in the United States partly because of the presence of brand names and large firms. In many developing countries small-scale

production dominates and producers, especially in the informal sector, do not have brand names, suggesting that such organizations may have to be modified for local conditions.

History may provide some guidance on how such modification should be made. In premodern Europe small-scale production without brand names was also the norm. The Europeans attached a brand name to the product of a group of producers (usually on a regional basis) and took advantage of local informal and formal contract enforcement within a group to induce each producer to ensure production quality. Similar organizations existed in late Imperial China. In large cities craft associations based on the region of origin of their members set quality standards and supervised product quality in an explicit attempt to acquire and preserve reputation (Hamilton 1985). Similarly, one interpretation of township and village enterprises in contemporary China (a form of enterprise organization in which local government owns the enterprise but local individuals hold implicit property rights) is that the local government's concerns about its reputation in external credit markets enables it to commit to the quality of local investment opportunities and thereby raise the capital required for such investment (Che 1994).

Enhancing Contract Enforcement by Incorporating Informal Contract Enforcement Institutions

Besley (1995a) notes that traditional institutions for credit based on informal contract enforcement "seem in general to disappear as capital markets develop" (p. 121). It may well be that despite the cultural and social differences among various societies, development leads to convergence in at least some aspects of contract enforcement mechanisms. Policy may be able to ease the transition.

A policy aimed at enhancing contract enforceability may be counterproductive if it is adopted at an inappropriate stage of development. In a society based on a particular set of informal contract enforcement institutions, the introduction of an effective legal system, or even alternative informal contract enforcement institutions, may undermine rather than promote exchange. In collectivist societies the limit on intergroup exchange facilitates intragroup exchange. Enhancing intergroup exchange may thus decrease the extent of possible intragroup exchange. Policies that weaken ties within a group may lead to a decline in contract enforcement and efficiency. Similarly, introducing a legal system may undermine the implicit understanding that governs transactions under informal contract enforcement. The need to follow the letter of the law may increase transactions costs.

Not intervening with contract enforcement institutions also may retard growth, particularly when contract enforcement is based on reputation. For a reputation mechanism to foster exchange, the exchange must be sufficiently profitable. To the extent that development eliminates excessive profits, it may undermine the operation of the reputation mechanism. At the same time, as the model indicates, there is no reason to expect that economies will be able to endogenously adjust their contract enforcement institutions to the new situation. The effects on informal contract

enforcement of policies that promote growth and competition should thus be considered, and appropriate remedies implemented.

The relationship between expected future gains from exchange and informal enforcement based on reputation implies that such enforcement is sensitive to economic downturns. An economic downturn that sufficiently reduces the expected future return from exchange may lead to a reduction in contract enforceability that exacerbates the downturn. It may not be a coincidence that the recent recession in Japan was associated with accusations that implicit contracts had been reneged on. Since reestablishing a system of informal contract enforcement may be time consuming, a policy aimed at providing alternative means of contract enforcement may be warranted.

Conclusion

Because a society's contract enforcement institutions are a product of distinct historical, economic, cultural, political, and social processes, different societies have different contract enforcement systems, each of which may be more or less efficient in particular situations. Furthermore, the interdependence of contract enforcement and other aspects of a society implies that attempts to implement a particular feature of the contract enforcement system of one society in another may not enhance efficiency. Contract enforcement institutions are not necessarily transferable from one society to another. But since systems of contract enforcement may not be efficient, intervention can be beneficial. Empirical and theoretical analysis, directed by the methodologies and insights of various disciplines within the social sciences, is likely to highlight the emergence, operation, and efficiency implications of the various systems of informal contract enforcement of particular societies, inspiring appropriate policies.

Notes

1. In incomplete information models (first introduced by Kreps and others 1982), cheating reveals information about the cheater as a type of "lemon" who will cheat again. Another mechanism is discussed later.

2. Some of these institutions are based not only on reputation but also, indirectly, on the legal system.

3. For a discussion of such an "efficiency wage" mechanism, see Shapiro and Stiglitz (1984). This discussion implicitly assumes an infinite time horizon. The results also hold when the horizon is finite but there is uncertainty when termination will occur or when there are multiple Pareto ranking equilibria in the stage game (see Pearce 1992).

4. Clearly, members of a group can try to manipulate the information transmitted to other groups.

5. If it is certain that agency relations will terminate at some point in time, there is no subgame perfect equilibrium in which agents are hired and do not cheat. It is always optimal for an agent to cheat in the period before employment is terminated. Hence if it is known that employment will be terminated in period t, the agent will cheat at $t-1$, implying that the principal will find it best to terminate the relations in $t-2$. If, however, the agent recognizes that the principal will terminate relations at $t-2$, the agent will cheat at $t-3$.

References

Aoki, Masahiko. 1988. *Information, Incentives, and Bargaining in the Japanese Economy.* Cambridge: Cambridge University Press.

Arrow, Kenneth. 1974. *The Limits of Organization.* New York: Norton.

Banerjee, V. Abhijit, Timothy Besley, and Timothy W. Guinnane. 1994. "Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test." *Quarterly Journal of Economics* (May): 491–515.

Bardhan, Pranab, ed. 1989. *The Economic Theory of Agrarian Institutions.* Oxford: Oxford University Press.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1985. *Habits of the Heart: Individualism and Commitment in American Life.* Berkeley: University of California Press.

Benson, Bruce L. 1989. "The Spontaneous Evolution of Commercial Law." *Southern Economic Journal* 55 (3): 644–61.

Bernstein, Lisa. 1992. "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry." *Journal of Legal Studies* 21 (January): 115–57.

———. 1995. "The Newest Law Merchant: Private Commercial Law in the United States." Working paper. Georgetown University, Georgetown Law Center, Washington, D.C.

Besley, Timothy. 1995a. "Nonmarket Institutions for Credit and Risk Sharing in Low-Income Countries." *Journal of Economic Perspectives* 9 (1): 115–27.

———. 1995b. "Savings, Credit and Insurance." In Jere Behrman and T. N. Srinivasan, eds., *Handbook of Development Economics,* vol. 3. Amsterdam: North Holland.

Che, Jiahua. 1994. "Community Government as an Enforcement Mechanism with Collective Responsibility: An Analysis of Township-Village Enterprises." Working paper. Stanford University, Department of Economics, Stanford, Calif.

Clay, Karen. 1995. "Coalitions and Contract Enforcement: Trade on the California Coast, 1830–1846." Working paper. University of Toronto, Department of Economics.

Davis, Kingsley. 1949. *Human Society.* New York: Macmillan.

Dawes, Robyn M., and Richard H. Thaler. 1988. "Anomalies: Cooperation." *Journal of Economic Perspectives* 2 (summer): 187–97.

de Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World.* New York: Harper & Row.

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes.* Cambridge, Mass.: Harvard University Press.

Ellison, Glenn. 1994. "Cooperation in the Prisoners' Dilemma with Anonymous Random Matching." *Review of Economic Studies* 61 (3): 567–88.

Ensminger, Jean. 1992. *Making a Market: The Institutional Transformation of an African Society.* Cambridge: Cambridge University Press.

Fafchamps, Marcel. 1993. "The Enforcement of Commercial Contracts in Ghana." Working paper. Stanford University, Food Research Institute, Stanford, Calif.

———. 1995. "Trade Credit in Zimbabwe." In Anita Spring and Barkaro McDade, eds., *African Entrepreneurship.* Bloomington: Indiana University Press.

Fafchamps, Marcel, Tyler Biggs, Jonathan Conning, and Pradeep Srivastava. 1993. "Enterprise Finance in Kenya." World Bank, Africa Region, Regional Program on Enterprise Development, Washington D.C.

Gambetta, Diego. 1993. *The Sicilian Mafia.* Cambridge, Mass.: Harvard University Press.

Geertz, Clifford. 1973. "Suq: The Bazaar Economy in Sefrou." In Clifford Geertz, Hildred Geertz, and Lawrence Rosen, eds., *Meaning and Order in Moroccan Society.* Cambridge: Cambridge University Press.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedeness." *American Journal of Sociology* 91 (3): 481–510.

———. 1994. "The Economic Sociology of Firms and Entrepreneurs." In Alejandro Portes, ed., *The Economic Sociology of Immigration.* New York: Russell Sage Foundation.

Greif, Avner. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *Journal of Economic History* 44 (4): 857–82.

———. 1992. "Institutions and International Trade: Lessons from the Commercial Revolution." *American Economic Review* 82 (2): 128–33.

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

———. 1994a. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." *Journal of Political Economy* 102 (5): 912–50.

———. 1994b. "Trading Institutions and the Commercial Revolution in Medieval Europe." In Abel Aganbegyan, Oleg Bogomolov, and Michael Kaser, eds., *Economics in a Changing World,* vol. 1. Proceedings of the Tenth World Congress of the International Economic Association. London: Macmillan.

———. 1995. "Markets and Legal Systems: The Development of Markets in Late Medieval Europe and the Transition From Community Responsibility to an Individual Responsibility Legal Doctrine." Stanford University, Department of Economics, Stanford, Calif.

Greif, Avner, and Eugene Kandel. 1995. "Contract Enforcement Institutions: Historical Perspective and Current Status in Russia." In Edward P. Lazear, ed., *Economic Transition in Eastern Europe and Russia: Realities of Reform.* Stanford, Calif: Hoover Institution Press.

Greif, Avner, Paul Milgrom, and Barry Weingast. 1994. "Coordination, Commitment and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* 102 (4): 745–76.

Guinnane, Timothy W. 1994. "A Failed Institutional Transplant: Raiffeisen's Credit Cooperatives in Ireland, 1894–1914." *Explorations in Economic History* 31 (January): 38–61.

Hamilton, Gary G. 1985. "Why Not Capitalism in China? Negative Questions in Historical Comparative Research." *Journal of Developing Societies* I (2): 187–211.

———. 1990. "Patriarchy, Patrimonialism, and Filial Piety: A Comparison of China and Western Europe." *British Journal of Sociology* 41 (1): 77–104.

———. 1991. "The Organizational Foundations of Western and Chinese Commerce: A Historical and Comparative Analysis." University of California at Davis, Department of Sociology.

———. 1994. "Civilizations and the Organization of Economies." In Alejandro Portes, ed., *The Economic Sociology of Immigration.* New York: Russell Sage Foundation.

Hamilton, Gary G., and Nicole Woolsey Biggart. 1988. "Market, Culture, and Authority: A Comparative Analysis of Management and Organization in the Far East." *American Journal of Sociology* 94 (supplement): S52–S94.

Hayami, Yujiro, and Toshihiko Kawagoe. 1993. *The Agrarian Origins of Commerce and Industry.* New York: St. Martin's Press.

Hofstede, Greet. 1984. "The Cultural Relativity of the Quality of Life Concept." *Academy of Management Review* 1 (3): 389–98.

Holländer, Heinz. 1990. "A Social Exchange Approach to Voluntary Cooperation." *American Economic Review* 80 (December): 1157–67.

Homans, George C. 1950. *The Human Group.* New York: Harcourt.

Kennedy, Paul. 1988. *African Capitalism: The Struggle for Ascendancy.* Cambridge: Cambridge University Press.

Klein, Daniel B. 1992. "Promise Keeping in the Great Society: A Model of Credit Information Sharing." *Economics and Politics* 4 (2): 117–36.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory* 27 (4): 245–52.

Macaulay, Stewart. 1963. "Noncontractual Relations in Business: A Preliminary Study." *American Sociological Review* 28: 55–67.

———. 1985. "An Empirical View of Contract." *Wisconsin Law Review* 3: 465–82.

Macfarlane, Alan. 1978. *The Origins of English Individualism.* Oxford: Basil Blackwell.

Macneil, Ian R. 1985. "Relational Contract: What We Do and Do Not Know." *Wisconsin Law Review* 3: 482–524.

Milgrom, Paul R., and John Roberts. 1992. *Economics, Organization, and Management.* Englewood Cliffs, N.J.: Prentice Hall.

Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics & Politics* 2 (1) : 1–23.

North, Douglass C. 1991. "Institutions." *Journal of Economic Perspectives* 5 (1): 97–112.

Pearce, David G. 1992. "Repeated Games: Cooperation and Rationality." In Jean-Jacques Laffont, ed., *Advances in Economic Theory,* vol. 1. Proceedings of the Sixth World Congress of the International Economic Association. Cambridge: Cambridge University Press.

Platteau, Jean-Philippe. 1994a. "Behind the Market Stage Where Real Societies Exist. Part I: The Role of Public and Private Order Institutions." *Journal of Development Studies* 30 (3): 533–77.

———. 1994b. "Behind the Market Stage Where Real Societies Exist. Part II: The Role of Moral Norms." *Journal of Development Studies* 30 (3): 753–817.

Poewe, Karla. 1989. *Religion, Kinship, and Economy in Luapula.* Lewinston, NY: Edwin Mellen Press.

Portes, Alejandro. 1994. "The Informal Sector and Its Paradoxes." In Alejandro Portes, ed., *The Economic Sociology of Immigration.* New York: Russell Sage Foundation.

Pyatt, Graham, and S. Gordon Redding. 1995. "A Survey of Chinese Businessmen in Thailand." Preliminary report presented at the Center for Advanced Studies in the Behavioral Sciences. Stanford University, Stanford, Calif..

Redding, S. Gordon. 1990. *The Spirit of Chinese Capitalism.* Berlin: Walter de Gruyter.

———. 1995. "Overseas Chinese Networks: Understanding the Enigma." *Long Range Planning* 28 (1): 61–9.

Reynolds, Charles H., and Ralph V. Norman, eds. 1988. *Community in America.* Berkeley: University of California Press.

Schmitz, Hubert. 1995. "Collective Efficiency: Growth Path for Small-Scale Industry." *Journal of Development Studies* 31 (4): 529–66.

Shapiro, Carl, and Joseph E. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74 (3): 433–44.

Smelser, Neil J., and Richard Swedberg. 1994. "The Sociological Perspective on the Economy." In Neil J. Smelser and Richard Swedberg, eds., *The Handbook of Economic Sociology.* New York: Russell Sage Foundation.

Smith, Adam. 1896 [1763]. *Lectures on Justice, Police, Revenue, and Arms.* Edited by Edwin Cannan. Oxford: Clarendon Press.

Sugden, Robert. 1989. "Spontaneous Order." *Journal of Economic Perspectives* 3 (4): 85–97.

Szanton, David L. 1971. *Estancia in Transition: Economic Growth in a Rural Philippine Community.* Quezon City, Philippines: Ateneo de Manila University Press.

Triandis, Harry C. 1990. "Cross-Cultural Studies of Individualism and Collectivism." In J. Berman, ed., *Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press.

Udry, Christopher. 1994. "Risk and Insurance in a Rural Credit Market: An Empirical Investigation in North Nigeria." *Review of Economic Studies* 61 (3): 495–526.

Wallis, John Joseph, and Douglass C. North. 1986. "Measuring the Transaction Sector in the American Economy, 1870–1970." In Stanley L. Engerman and Robert E. Gallman, eds., *Long-Term Factors in American Economic Growth.* National Bureau of Economic Research Studies in Income and Wealth, vol. 51. Chicago: University of Chicago Press.

Weber, Max. 1981 [1927]. *General Economic History.* New Brunswick, N.J.: Transaction Books.

Williamson, Oliver. 1985. *The Economic Institutions of Capitalism.* New York: Free Press.

Yamagishi, Toshio, and Midori Yamagishi. 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18 (2): 129–66.

Zucker, Lynne G. 1986. "Production of Trust: Institutional Sources of Economic Structure, 1840–1920." *Research in Organizational Behavior* 8: 53–111.

Comment on "Contracting, Enforcement, and Efficiency: Economics beyond the Law," by Avner Greif

Robert C. Ellickson

Avner Greif, a leading economic historian of informal contract enforcement, provides a cogent overview of the subject. His expositional strategy of mixing game theory with anecdotes should placate both theorists and those skeptical of theory. His basic theses are, if anything, insufficiently controversial. In my view "economic sociologists"—a felicitous term of Greif's—have already won the battle against the legal-centralists who claim that the rules of formal law are necessary and sufficient foundations of a market economy. Nor will anyone quarrel that it is better to make social structures endogenous rather than to take them as givens.

Greif focuses on contract enforcement among traders. It is not evident that his insights would carry over to other substantive contexts, such as family and inheritance rules, the establishment of property rights, and protection of individuals' bodily integrity. Informal enforcement may work less well in some of these areas. For example, Hernando de Soto (1989), a well-known admirer of the informal sector, asserts that the Peruvian experience shows that the legal system is superior to the informal sector in ensuring property rights in residential neighborhoods.

One can quibble with some of Greif's theoretical points. He predicts that, in order to maintain informal contract enforcement within the group, members would not be punished for cheating an outsider. This assertion may be true for utterly segregated groups, such as gypsies, but is implausible for more typical groups, whose members engage in both internal and external trade and would therefore want to maintain their reputations with outsiders.

A tone of caution pervades Greif's article. He repeatedly states that no single approach to contract enforcement is likely to be optimal in all social contexts, in part because societies have evolved along different paths that constrain their options. I commend his prudence. Current social-scientific theory of the informal sector is not yet sufficiently mature to provide economic development specialists with marching orders, although it can prevent them from making some obvious errors.

Robert C. Ellickson is Walter E. Meyer Professor of Property and Urban Law at the Yale Law School.

Annual World Bank Conference on Development Economics 1996

The Intriguing Concept of Social Capital

The issues of spontaneous order that fascinate Greif have excited interest in many academic disciplines—the core social sciences certainly, but also game theory, philosophy, and psychology. To complement Greif's stress on the tools of economics, I strive in this comment to connect his ideas to the closely related theory of social capital that is emerging within political science and sociology. Two notable contributors have been Robert Putnam and James Coleman.

Putnam (1993) distinguishes between the vertical social ties present in patron-client, parent-child, and other hierarchical relationships, and the (usually thinner) horizontal ties existing among merchants and others who interact as presumptive equals on a basis other than kinship. (This is a variant of Greif's distinction between collectivist and individualist social relations.) Just as Greif argues that individualist relations and general morality tend to foster intergroup trade, Putnam asserts that strong horizontal ties are essential for the emergence of a strong market economy and a healthy state. An economy performs less well, claims Putnam, when it relies heavily on intrafamily and intrahierarchy exchanges. The upshot is that the enhancement of social capital in horizontally organized groups is essential to economic development.

The notion of social capital remains obscure, however. According to Coleman (1990) social capital, unlike physical and human capital, inheres in relations among people and depreciates from lack of use. Putnam repeatedly refers to certain examples of social capital—"norms of reciprocity," "networks of civic engagement," "trust"—but offers no definition of the term. He regards formal associations such as bowling leagues, which have designated leaders, membership lists, periodic meetings, and official records, as repositories of social capital (see also McPherson and Rotolo 1996). While a bowling league is indisputably a better incubator of social exchange than bowling alone would be, it is not obvious why bowling in leagues would foster social capital better than would bowling regularly with an informal group of friends.

A Definition of Social Capital

Drawing in part on prior work (Ellickson 1991), I see a group's social capital as consisting of two components: the stock of previous cooperation among members and the presence of a credible system for enforcing norms through self-help and other nonlegal means. Both conditions foster trust and future cooperation.

The first component, the stock of previous cooperation, nourishes expectations of future cooperation. These expectations are then partly self-fulfilling. This stock could be measured by cumulating instances in which the members of a group had achieved a cooperative outcome and subtracting instances in which members had defected and caused a suboptimal outcome. This formulation obscures many subsidiary issues that must be resolved to operationalize the calculus. High-stakes instances warrant more weight than low-stakes ones. The significance of a particu-

lar cooperative or uncooperative outcome also presumably declines as times passes. (This may be one basis for Coleman's statement that social capital depreciates over time.)

The second component—the presence of a credible informal social control system—also fosters cooperative behavior. Greif observes that both first- and third-party enforcement systems may be significant. A first-party system involves internalized norms against defection that individuals enforce against themselves. A trader who is aware that other traders have been socialized to behave cooperatively would be more trusting and less likely to defect. The likelihood of third-party punishment of a defection increases when an actor is embedded in social networks that provide ready opportunities for others to administer sanctions. Informal sanctioning is fostered when actors belong to a closely knit group, that is, when information about the behavior of all group members is readily available, the power to administer informal sanctions is broadly distributed, and member exit from the web of continuing relationships is difficult. A group may have constitutive norms that directly serve to maintain its close-knittedness; these govern the symbols, rituals, and ideologies of the membership (Ellickson 1991).

A simple example can highlight the distinction between these two components of social capital. Suppose twenty strangers end up sharing a lifeboat after a ship capsizes in the middle of the ocean. As strangers, these people would have no stock of previous cooperation. But they would have much social capital of the second sort. For the duration of their journey they would be embedded in a social situation characterized by excellent information, easy sanctioning, and scant means of exit.

This definition of social capital suggests a number of promising avenues for research. Many game theorists anticipate that a previously cooperative person will defect as an iterated game approaches its end. This supposes that the credibility of future social controls is a far more important component of social capital than is a history of previous cooperation. In what contexts is this true?

Durkheim originated the notion that the presence of a few deviants enhances the well-being of a group (Ellickson 1991). Although a defection reduces the stock of previous cooperation, apprehension and punishment of deviants might augment the credibility of the informal system of social control enough to increase total social capital. Publicity about recent accomplishments of third-party enforcers enhances enforcers' reputations for effectiveness. By creating opportunities for well-publicized punishments, deviants therefore may make net contributions to social capital. In addition, practice in exercising social controls may be important. In a group with a unblemished history of cooperation, skill in using tit-for-tat and related control techniques might deteriorate. This possibility provides another basis for Coleman's assertion that social capital depreciates when not used.

Nurturing Social Capital from Above

A state—or an international organization such as the World Bank—may aspire to enhance stocks of social capital both to promote economic development and to

strengthen civil society. These sorts of efforts are paradoxical because they attempt to use hierarchical planning to improve spontaneous social structures. As officials at the World Bank well know, the track record of interventions of this sort is hardly encouraging.

Like Greif, both Coleman and Putnam are deeply skeptical of outside efforts to enhance social capital, which they envision as an asset that slowly accumulates through indigenous interactions within groups. Coleman's skepticism is particularly notable because he regards social capital as a public good that self-interested actors are apt to undersupply. He notes that groups are strengthened by closure (attainment of reciprocal relations across all pairs), stability, and the sharing of a supportive ideology. It is implausible that an outsider could readily manipulate any of these conditions. Indeed, Coleman believes that economic prosperity tends to reduce the social capital of traditional groups by destabilizing them.

In an important article Eric Posner (1996) offers a somewhat more optimistic assessment of the possibility of constructive legal manipulation of informal groups. Posner distinguishes between legal rules aimed at influencing the behavior of groups and legal rules aimed at aiding individuals in the demographic category from which a group draws its members. Although Posner does not discuss economic development as such, his analysis provides some support for a program of the following sort. First, the intervening entity would identify groups whose efforts it wanted to bolster; in the development context these would be social groups connected by horizontal ties. Second, the entity would make use of grants, tax exemptions, and the like to transfer resources to the groups themselves or to the groups' members on the basis of their group membership. These actions could be expected to strengthen the social capital of the favored groups by making membership more valuable and exit less likely. Finally, the intervening entity would refrain from policies that might weaken favored groups. The two counterproductive policies Posner repeatedly highlights are provision of non–group-based aid to the individuals in the population from which a group draws and regulation of a group's internal operations. The first reduces social capital by easing exit; the second interferes with, or supplants, the group's internal system of social control.

The upshot of Posner's analysis is that economic development specialists might strengthen social capital by providing aid to merchant associations, consumer associations, sports groups, and social clubs and also by not interfering in the operations of these groups. The definition of social capital offered above points to the possible merit of other types of interventions. Informal enforcement systems depend in part on reliable information about member behavior. To enhance information flows in developing countries, an intervening agency might subsidize the creation of new informal meeting places—perhaps a common water pump at a central crossing or a new multipurpose meeting room. In a more developed setting the agency could support telephones, photocopiers, and desktop publishing. To help ensure a recipient group's continued close-knittedness, the agency might disburse aid in installments conditioned on a group's continued existence, rather than in an initial lump sum award.

Cautions for Development Specialists

I have little confidence in the wisdom of these suggestions. In some contexts these policies might prove worthless, perhaps because outsiders are unable to fathom the social conditions they confront. Worse, policies may backfire. They may end up weakening groups or converting them into aggressive special interests. Too often, outside experts imagine that they will succeed in jolting a social group trapped at a local optimum to move to a global optimum (Rowe 1996). To do so consultants would have to have better information and incentives than group members do, which is unlikely. Members have their ears closer to the ground and have greater stakes in the consequences of reforms. In addition, as Putnam (1993) notes, changes in social capital are "measured in decades" (p. 184). Is the time horizon of an ambitious World Bank official long enough to spur a personal interest in glacial change?

I tentatively conclude that economic development specialists at most should attempt no more than a few small experiments with policies designed to augment social capital. The theory of social capital is not well developed and, as Greif stresses, implementation tends to be perilous because social contexts vary and are hard to assess. Development specialists are more likely to be operating with a comparative advantage when they engage in their more traditional pursuit of reforming legal systems. Although legal centralists have long overestimated the importance of law, Greif and others would agree that the legal framework is important in some areas of trade. Law may also be highly influential in contexts such as land tenure, inheritance, and family law, not to mention business associations and taxation. If the theory of social capital is not yet adequate to enable consultants to reform informal groups from on high, it is rapidly becoming robust enough to pinpoint misguided legal reforms that undermine civil society.

References

Coleman, James S. 1990. *Foundations of Social Theory.* Cambridge, Mass.: Harvard University Press.

de Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World.* New York: Harper & Row.

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes.* Cambridge, Mass.: Harvard University Press.

McPherson, J. Miller, and Thomas Rotolo. 1996. "Diversity and Change in Voluntary Groups." *American Sociological Review* 61 (2): 179–202.

Posner, Eric A. 1996. "The Regulation of Groups: The Influence of Legal and Nonlegal Sanctions on Collective Action." *University of Chicago Law Review* 63: 133–97.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton, N.J.: Princeton University Press.

Rowe, Mark J. 1996. "Chaos and Evolution in Law and Economics." *Harvard Law Review* 109: 641–68.

Comment on "Contracting, Enforcement, and Efficiency: Economics beyond the Law," by Avner Greif

Sally Falk Moore

Avner Greif's article is subtle and stimulating. It is splendidly context minded to assert that policy regarding institutions of contract enforcement must take into account economic, social, cultural, and political factors. And I am almost completely disarmed by the disclaimer at the beginning of his article that "the theoretical framework . . . clearly has a limited ability to capture the complexity of reality" (p. 240). Greif's warning that the contextualized realities and the theoretical models should not be expected to match is almost enough to fend off any critique.

There are nevertheless some dimensions that might be added to the analysis as well as some shifts of emphasis that might strengthen the argument.

Enforcing Contracts: Timing and Incentives

Contract enforcement by courts differs in important ways from informal, socially based contract enforcement. Depending on the setting, the difference is often between what might be called the delayed single-transaction remedies of the courts and the much more durable and continuous discipline exercised in long-term, multistranded transactional fields. At best, recourse to a court takes place after contractual failure. A complainant goes to court when a dispute arises. Even if redress is given, it is likely to come only after substantial delay. Do such legal remedies have much effect on other contracting parties and on subsequent contracts? At the time that an agreement is made, how much effect does the possibility of future lawsuits have on performance? Those making informal agreements are usually unaware of, and therefore unswayed by, case law and legal precedent or the likelihood of future lawsuits. Redress from the courts always comes *after* the deal has gone sour.

By contrast, the features of the social milieu in which informal norms and agreements are generated are in place before the contract comes into existence. From the start, the existence of these conditions encourages participants to meet their obligations and work out their differences. Where access to the social field and member-

Sally Falk Moore is professor of anthropology at Harvard University.

Annual World Bank Conference on Development Economics 1996

ship in good standing are not only prerequisites to going about one's normal business but provide opportunities for mounting advantageous transactions, participants cannot take the risk of losing access or reputation either in the short run, when a single transaction is at stake, or in the long run, when a series of transactions is involved. Playing the game more or less by the rules is mandatory for staying in the game. The pressures and inducements that together produce informal contracts are ongoing and operate to prevent breach long before there is a breakdown (as well as helping to settle matters afterwards).

This raises a basic question about the way the topic has been framed. If our primary interest is in economic development, is the focus on enforcement the best way to approach the issue? Why not concentrate on inducements to perform—the rewards associated with meeting contractual obligations—rather than on penalties for failure? In many settings the threat of being disciplined may be a less immediate motivator than the pleasure of respectability and the vision of profit.

In developing countries, legislative bodies, courts, and administrative agencies seldom epitomize the rational order envisioned by Max Weber (1978) when he wrote about the role of a formal and calculable law in fostering the development of capitalism. Access to courts and officials is often problematic. And in many settings once access is obtained, there is little hope of finding Weberian rationality and impartiality. To note the vitality and usefulness of certain informal social milieus that ensure transactional performance is by no means to say that there is never any need for effective formal structures. Developing countries would be well served by more effective formal legal institutions.

Conjunctions and Intersections: Social Fields and Theoretical Types

However theoretically tidy Greif's dichotomy between two types of society may be—one individualist and one collectivist, one integrated and one segregated, one in which formal legal structures set the norms and enforce contracts and one in which informal relationships prevail—this oversimplification provides no means of analyzing the many settings in which several order-producing systems coexist. One has only to think of the many great trading and banking houses (past and present) that have been the internal monopolies of particular ethnic groups. These institutions are often internally closed for certain purposes but transact readily and as often as possible in the impersonal market with outsiders. As Greif notes, many informal contractual arrangements are made in the United States despite the presence of a formal legal system. Not only does the existence of a formal legal system in the United States not preclude informal arrangements, the legal and economic systems could not function without them.

To my mind what is needed is an approach grounded in the inspection of social fields in action. Such an approach does not prejudge what, in any particular social field, might be found to be the operative sources of social order.

For my own purposes, I have long found the concept of the semiautonomous social field to be a useful way of thinking about the relation between formal and informal obligation-setting and norm-enforcing systems (Moore 1978). Working in

rural East Africa during the period of high socialism in Tanzania, I was struck by the limits on the reach of the state. Despite the establishment of a variety of party and state institutions, many zones of local autonomy remained.

It is useful to establish a simple framework within which to conceptualize the coexistence and interaction of multiple types of control. As I conceive it, the semiautonomous social field is one that has rule-making capacities of its own and the means to induce or coerce compliance internally, but is embedded in a larger world. Among other influences of external provenance, the law and legal institutions of the state can create, invade, and shape the social fields within it. Yet this rarely takes place exactly as intended, and is seldom completely encompassing. The interaction among semiautonomous social fields and between them and the legal system is complex, variable, and generally selective.

A social field may even normalize illegality. In Taiwan (China) small businesses that have problems enforcing contracts sometimes hire thugs to beat up or threaten to beat up defaulters (Winn 1994). This is informal contract enforcement. It is effective. But it is not very attractive. Greif specifically excludes the use of coercive power from his discussion. But there are many forms and degrees of symbolic violence, reminders of the real thing, that can be used to exact conformity (Weber 1978; Bourdieu 1977; and Foucault 1986). To omit coercive dimensions may be to severely limit and unbalance the discussion, particularly since the option of using force is not open to the courts.

As a framing concept, the semiautonomous social field has the advantage of requiring that one consider several dimensions at once: how transactions are actually negotiated and carried out in a particular milieu and how that "site" is affected by the sociopolitical and economic environment. The concept helps crystallize thinking on the issues that need to be investigated.

In developing countries social fields of informal contract enforcement that also, on occasion, interact actively with impersonal markets are especially deserving of attention. In some of these the informal structure absorbs many of the transactions costs involved in raising and managing capital, yet the capital can be carried across the threshold between personal and impersonal settings. If properly encouraged and channeled, such threshold crossings could have a significant economic impact on development. My purpose is to underline conjunction and intersection as empirical realities and to criticize the exaggeratedly pure typologizing that tends to be used in abstract theoretical models.

By way of illustration, consider how rotating credit associations have contributed to economic development. Such institutions suggest some of the possibilities bubbling up from "below." In a rotating credit association all members meet each week or month (or some other agreed interval) and contribute a fixed sum to a pooled fund, all of which is given to a different member at each meeting. This rotation gives members periodic access to a more substantial bundle of cash than they could normally accumulate through their individual savings.

Such rotating credit associations have sprung up all over the world. There are associations of men, associations of women, and mixed associations. There are asso-

ciations of poor people and still more of people who have steady incomes. They are commonplace in urban areas and characteristically are not assemblages of kinspeople. They are found from Mexico City to London, from Taiwan (China) to Africa (Ardener and Burman 1995).

How do these mutual-benefit associations contribute to larger-scale economic development? In addition to the kitty for distribution, some rotating credit associations establish a common capital fund to which they also contribute at each session. Any member can borrow from this fund, but the borrower must pay a substantial rate of interest on the loan, and the loan must be approved by the association's members.

One such African association with which I have had contact has accumulated enough money in the common fund to start investing in urban real estate and construction. Despite the large sums of money involved, there is no question of formal legal involvement in the association's internal affairs. The members of the association trust each other—and with good reason.

In its urban land transactions the association tends to deal in an impersonal land market, in which transactions are conducted with members of other ethnic groups. These arrangements are not at all comparable to the social relations that are continuously reaffirmed within the credit association. For these transactions the association turns to the law. They hire a lawyer. They make formal contracts. They deal with local bureaucracies.

Thus these men are straddling two social fields—one in which close personal networks govern transactions and one in which markets are impersonal. The capital generated in the common fund of the association can be used for productive investment in a larger world. This example has parallels all over the world. Policies could be devised to encourage the grouping together of such associations to aggregate still larger capital resources for investment. At one point Taiwan (China) created bank-like formal institutions for such purposes, and they are reported to have enjoyed a good deal of success (Winn 1994).

It is clear that there are potentially entrepreneurial groups in developing countries that could further economic development while serving their own interests and "enforcing" their own internal agreements. Too great a theoretical preoccupation with formal contract enforcement might well miss the positive dynamic possibilities existing in such social fields. Awareness of such efforts is crucial to formulating policies to make such groups more effective on a large scale and in the long run. Other types of organizations that could potentially contribute to development in the same way should also be examined.

I strongly agree with Greif's general proposals—especially his recommendation that local, context-specific empirical and theoretical analyses precede the formulation of policy. Unfortunately, the subtext of a purely theoretical preoccupation with contract enforcement seems to be that defaulting and cheating are the principal obstacles to development. But are they? I doubt it.

References

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

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice.* Cambridge: Cambridge University Press.

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes.* Cambridge, Mass.: Harvard University Press.

Foucault, Michel. 1986. "Disciplinary Power and Subjection." In Steven Lukes, ed., *Power.* New York: New York University Press.

Macaulay, Stewart. 1963. "Non-Contractual Relations in Business: A Preliminary Study." *American Sociological Review* 28: 55–67.

Moore, Sally Falk. 1978. "Law and Social Change: The Semi-Autonomous Social Field as an Appropriate Subject of Study." In Sally Falk Moore, eds., *Law as Process.* London: Routledge.

Weber, Max. 1978. *Economy and Society,* vol. 2. Berkeley: University of California Press.

Winn, Jane Kaufman. 1994. "Relational Practices and the Marginalization of Law: Informal Financial Practices of Small Businesses in Taiwan." *Law and Society Review* 28 (2): 193–232.

Floor Discussion of "Contracting, Enforcement, and Efficiency: Economics beyond the Law," by Avner Greif

Greif began the discussion by responding to Robert Ellickson's (discussant) point that the concept of social capital advanced by James Coleman (1990) and Robert Putnam (1993) was poorly defined. In fact, Greif said, some scholars who had written extensively on the subject had not defined it at all. Economists like to identify exogenous and endogenous variables, and if none exist then it is hard for them to think about the issue. Ellickson had provided a framework in which to think more concretely about the issue. What Greif found especially intriguing are the mechanisms through which social structures are generated. One of the aims of his analysis had been to understand the relationship between informal contract enforcement and the scope of exchange—that is, whether it took place within a close-knit group or not. He had not wanted to simply take one and explain the other; rather, he wanted to examine the interaction between them to see how informal enforcement reinforces segregation or, in turn, how a close-knit relationship reinforces informal contract enforcement. To the extent that his focus had been on examining the relationship between the two, the work might be closer to what Ellickson had in mind than it might appear.

Greif found Ellickson's point about a principal and agent's prior history of cooperation interesting because economists tend to assume that the world starts from zero at some point in time, which is never true. But Greif did not think it was enough of an explanation to say that because someone cooperated in the past they likely would keep on cooperating, even if they knew there would be no future relationship.

Greif said that he had intentionally refrained from elaborating on two points brought up by Sally Falk Moore (discussant) on the importance of corruption in the legal system and mafia-like use of coercion as an informal means of enforcement. He had dealt with those issues in another paper, about contract enforcement in Russia, but he thought that including it here would complicate the current article (Greif and Kandel 1995). Moore also had raised the issue of the desire for respectability as an element of contract enforcement, and he agreed that people sometimes forgo mate-

This session was chaired by Johannes F. Linn, vice president, Europe and Central Asia, at the World Bank.

Annual World Bank Conference on Development Economics 1996

rial gains to gain the respect of others. But if the exchange is anonymous—with a taxi driver, for example—people might not care what the other person thinks about them. Issues of respectability, personal trust, and personal relationships are not unrelated to the issue of segregation, he said. Greif thought that he should clarify one point, about the ability of informal contract enforcement (as opposed to ex post punishment in the legal system) to prevent a breach of contract. The division was not that clear-cut, he said. Deterrents—the fear of punishment—were at work in both cases. People honor contracts because they fear losing future business or being punished in court. Fear of punishment is the controlling factor either way.

A participant from the World Bank asked the panelists to elaborate on the discussion's implications for competition policy, especially in transition economies, where contract enforcement is largely informal, legal systems are underdeveloped, and competition is weak. If procompetition policies were implemented in those systems, it could break up relationships or networks of firms that are involved in informal contract enforcement. Where that is the case, is it important to push for procompetition policies in the absence of working legal institutions, or should countries wait until a legal system (or an alternative contract enforcement system) is in place? Greif agreed that procompetition strategies could hinder informal contract enforcement. There is a tradeoff involved. It also depends on the type of transaction involved. Is it a consumer market? An input-output purchase? Wholesalers and retailers? The extent to which it is an anonymous market is an important determinant of whether procompetition policy undermines informal enforcement systems. Other policies might have to be adopted, possibly even giving some organizational features to informal contract enforcement or explicitly using the legal system as much as possible in that economy.

Karla Hoff (discussant in another session) considered this to be a very important point. For her it suggested another question, about Greif's comment that informal enforcement mechanisms (or reputations) are not a free lunch because there is a segregation cost. She wondered if there was also a second cost, or tradeoff, because informal mechanisms work best with small groups while competition works best with large groups. Hence the more competition there is, the less free the lunch. Ellickson responded that smallness helps in informal cooperation, but informal social control is possible in large groups too. Concerns about reputation—informal penalties—can be important in surprisingly large groups. Academics, for example, worry about their reputations in circles that may encompass tens or even hundreds of thousands of others.

A participant from the World Bank noted that there seems to be a difference between the informal sector and the informal mechanisms that can operate between highly formal organizations. The World Bank, for example, trades hundreds of millions of dollars on the New York Stock Exchange in verbal deals, and if there is a misunderstanding about what has been said, the misunderstanding is settled through highly informal enforcement mechanisms.

A participant from the U.S. Treasury asked whether there was a bit of a departure between Greif's focus on segregation and Ellickson's and Moore's comments. After

all, very different kinds of groups and social fields produce rules. Would the panelists not distinguish between an organization like the National Association of Securities Dealers and specific ethnic kin groups that form a savings association and invest in real estate? Is there not a normative difference? Moore agreed that it was important to make distinctions, but in fact rotating credit associations are rarely among kin and often are formed by business associates. In the case she had discussed, the members were government bureaucrats in a capital city and their only intimate association was in connection with the money. When they invested their pool in the anonymous market, they did not withdraw from the formal system; rather, they hired a lawyer because they had to define their relationships among themselves and with the seller, tenant, and so on. The legal system needs repair work, but so do other systems. Organizing a way to accumulate capital is fundamental to economic health, and systems should be developed to stimulate such endeavors.

On the same topic, Timothy Besley (presenter from another session) shared an anecdote about a rotating credit association in Los Angeles that illustrates how complex the interactions between informal and formal enforcement can be. A member of the group had found some fault in the association and took his fellow members to court. The judge threw the case out, however, on the grounds that rotating credit associations constituted illegal gambling, so contract disputes among members could not possibly be settled in a formal court of law. Ellickson observed that although the legal ground for throwing the case out might sound silly, the basic point was that such groups cannot prosper if their members are able to fall back on the legal system. To remain strong, their internal enforcement systems should not be interfered with, so the court was probably right not to get involved.

Greif had the final word on rotating credit associations. They are not a modern innovation, he said; they existed in the Middle Ages and at some point Western society shifted away from them. Why, in the nineteenth century, had Germans come out with cooperatives that replaced rotating credit associations and made possible a good deal of agricultural development in a short period of time? Why had no similar arrangements developed in other countries? During that same century the U.S. economy also shifted from one system of contract enforcement to another. We know almost nothing about these transitions, Greif said; hardly any information has survived to enlighten us about why they occurred. It would be helpful to know what led one society to respond in a way that improved economic efficiency while another society remained stuck with largely informal, inefficient systems of contract enforcement.

References

Coleman, James S. 1990. *Foundations of Social Theory.* Cambridge, Mass.: Harvard University Press.

Greif, Avner, and Eugene Kandel. 1995. "Contract Enforcement Institutions: Historical Perspective and Current Status in Russia." In Edward P. Lazear, ed., *Economic Transition in Eastern Europe and Russia: Realities of Reform.* Stanford, Calif.: Hoover Institution Press.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton, N.J.: Princeton University Press.

Labor and Environmental Standards in International Trade

International Labor Standards and Trade

Alan B. Krueger

From theory alone it is difficult to generalize about the efficiency and equity of linking international labor standards to trade. Some economists contend that international labor standards are merely disguised protectionism. But an evaluation of the factors influencing support for legislation that would ban imports to the United States of goods made with child labor provides little support for the prevailing political economy view. In particular, members of Congress representing districts with a high proportion of unskilled workers, who are most likely to compete with child labor, are less likely to support a ban on imports made with child labor. Another finding is that the prevalence of child labor declines sharply with national income. Finally, an analysis of compulsory schooling laws, which are often suggested as an alternative to prohibiting child labor, finds a tremendous amount of noncompliance in developing countries.

As national economies have become more integrated, efforts to coordinate international labor standards have become more prominent. Opponents of international labor standards argue that such efforts are an unnecessary and counterproductive interference with the workings of free markets, representing either disguised protectionism or misplaced compassion. Proponents argue that a set of minimal labor standards is necessary to promote fair competition and to facilitate efficient operation of the labor market. In addition, a growing undercurrent of resentment in industrial countries toward trade with countries with low labor costs threatens the viability of international trade agreements. In the United States, for example, this opposition has been galvanized by presidential candidate Patrick Buchanan, but it also has been voiced by several union leaders and liberal politicians.

Alan B. Krueger is Bendheim Professor of Economics and Public Affairs at the Woodrow Wilson School of Public and International Affairs at Princeton University and research associate at the National Bureau of Economic Research. The author is grateful to Varuni Dayaratna for excellent research assistance and Richard Freeman, Gail Grossman, Larry Katz, Jorge Perez-Lopez, Andrew Samet, and Ken Swinnerton for helpful discussions.

Annual World Bank Conference on Development Economics 1996

The core areas of labor standards typically include freedom of association, collective bargaining, prohibition of forced labor, elimination of exploitative child labor, and nondiscrimination.[1] The International Labor Organization (ILO) has been the main institution concerned with international labor standards since its inception in 1919. The ILO establishes conventions that are binding only on the countries that ratify them. It is not empowered to enforce compliance with ratified conventions; instead it relies on international pressure, advice, and monitoring to encourage compliance. Additionally, several bilateral and multilateral trade agreements cover labor and environmental standards. For example, labor side agreements were critical to ratification of the North American Free Trade Agreement (NAFTA).

This article critically evaluates the arguments for and against international labor standards. It reviews theoretical perspectives on labor standards and discusses issues related to the use of international trade linkages as a lever for influencing labor standards. Because there are valid arguments on both sides of this theoretical debate, empirical evidence is necessary to sort out the validity of the cases for and against international labor standards and to determine the desirability of linking labor standards to trade policy.

Theoretical Perspectives on International Labor Standards

A variety of theoretical arguments have been made for and against international labor standards. Some of these arguments are very old; others have been developed more recently.

Efficient Competitive Markets and the Political Economy Model

A starting point for most economic analyses is the efficient competitive markets model. Ehrenberg (1994) provides an overview of the implications of this model for international labor standards. In this model the total compensation (monetary and nonmonetary) workers receive is equal to their marginal contribution to the value of output. Each country's economy operates at a Pareto-optimal position—no government policy will make a person better off without making another person worse off. Wage differentials compensate workers for the varying health risks and other disamenities they face on the job. Child laborers are assumed to be paid a wage commensurate with their contribution to output and to rationally choose between working and pursuing formal education or other activities. In this model labor standards cannot raise the welfare of a country as a whole, although they can make some workers better off at the expense of other workers, consumers, or employers. Ehrenberg (1994) and others argue that the cost of meeting standards is likely to be borne by workers, in the form of lower wages or devalued currency. This model is clearly a simplified view of the world, but it may capture some critical effects of introducing labor standards.

In this model trade between nations is based on comparative advantage. Countries specialize in the activities in which they have a comparative advantage in

terms of physical or human resources. The more different are nations, the more they stand to gain from trading with each other. A reduction in trade barriers will create winners and losers in each country, although the aggregate gain to the winners will exceed the loss to the losers in each country. According to this model, if a common set of labor standards were imposed on countries, the net gains from trade would be reduced. Prices of goods produced by labor-intensive technologies will rise if labor standards add to the cost of labor. Because industrial countries tend to specialize in capital-intensive goods, this model predicts that the welfare of consumers in industrial countries will decline if minimum labor standards are imposed worldwide, although the welfare of workers in labor-intensive industries in these countries may increase. (See Brown, Deardorff, and Stern forthcoming for a theoretical treatment of these and related issues.)

If this model is a good description of the economy, why would some industrial countries seek to impose international labor standards? A widely held view is that labor standards are pursued by vested interests in these countries (labor unions, employers in certain industries) to prevent competition from developing countries based on comparative advantage (Hansson 1983; Bhagwati 1994; and Srinivasan 1994). For example, if child labor is used extensively in the textile industry in developing countries, then textile companies in industrial countries would benefit from an international ban on child labor and so would have an incentive to lobby for such policies. Srinivasan (1994) boldly argues that "the demand for linkage between trading rights and observance of standards with respect to environment and labor would seem to arise largely from protectionist motives" (p. 38). I call this the prevailing political economy view of international labor standards.

Redistribution

An efficient private market may fail to generate a distribution of incomes or working conditions that is desired by the public. For example, wages for the least skilled workers may be so low as to impoverish a large segment of the workforce. Thus society may wish to redistribute income toward people with very low incomes. In an economy that is operating at the efficient frontier, any redistribution will entail some deadweight loss, ruling out lump sum transfers. For example, it is often argued that a minimum wage reduces the employment of some groups of workers, causing deadweight loss. But a minimum wage may still be desirable because the total income of low-paid workers increases if the elasticity of demand for labor is less than one. The desirability of labor standards as a redistributive tool would depend on a society's interest in redistributing income and on the comparative strengths and weaknesses of other programs that could be used to redistribute income.

The comparative advantage of a minimum wage, for example, depends critically on the elasticity of labor demand—the lower is the elasticity, the smaller is the distortion created by a minimum wage.[2] In the United States, where most minimum wage workers are employed in the nontraded goods sector, research generally finds that minimum wage increases have had little or no effect on employment (Card and

Krueger 1995). This may not be the case in developing countries, however. But the distortionary effect of a minimum wage and other labor standards in developing countries may be diminished by endogenous compliance; when the efficiency loss is great, firms and workers have a stronger incentive to avoid the minimum wage through noncompliance or by moving to the uncovered sector (Squire and Suthiwart-Narueput 1995).

It is also important to compare the net benefits of labor standards with the net benefits of feasible alternative policies. Labor standards often are not targeted to the poorest in society because the very poor either are not working or are working in the informal sector, where labor standards are not followed.

Much economic research focuses on the adverse side effects of public policies, to the exclusion of the effectiveness of the policies themselves. But just because policies may have negative side effects is not reason enough to conclude that the interventions are undesirable. Any unintended consequences must be weighed against the intended consequences. An analogy to medicine is instructive. Chemotherapy is used to treat certain forms of cancer even though chemotherapy often has adverse side effects. Indeed, to treat these adverse side effects, doctors often prescribe additional medications that may themselves have adverse side effects. Similarly, multiple economic policies to bring about a desired level of redistribution may be more effective than a single policy.

Market Failure

Some observers have emphasized market failures, especially in the labor market, as a justification for international labor standards.[3] In some cases labor standards can improve efficiency as well as equity if the market has failed. There are several possible reasons for market failures. Information in the labor market is often imperfect and asymmetric. For example, employees may lack adequate information about safety conditions. Employers may have an incentive to conceal safety risks, especially in casual labor markets in which reputational effects are small.

Unequal market power also may lead to market failure. Forced labor is an obvious example in which employers have market power, and the abolition of forced labor would enhance economic efficiency. As another example, child laborers are often imperfectly mobile, which confers some monopsony power to employers. If employers have monopsony power over workers, a skillfully set minimum wage could increase employment, wages, and welfare. Of course, if the minimum wage is set too high, it could reduce employment and efficiency, even in a monopsonistic labor market.

A related issue concerns discrimination, broadly defined to mean the existence of equally productive workers who are paid different wages based on characteristics unrelated to their productivity (such as race or gender). In a perfectly competitive market there is a strong economic incentive against discrimination because nondiscriminating firms will gain a competitive advantage. Yet social customs and market power may enable discrimination to persist (Akerlof 1976). The elimination of dis-

crimination would improve economic efficiency and would be morally justified. Swinnerton (1996) argues that core labor standards (such as prohibitions against forced labor and discrimination) are always economically efficient, while other standards (such as a fixed minimum work age) are efficient in some countries and inefficient in others, depending on the level of development.

Externalities, resulting from the failure of the parties involved to internalize all the costs and benefits of particular actions, also could lead to market failure. Although externalities probably provide a stronger justification for environmental standards than for labor standards, they may justify some labor market standards.

Recent economic modeling of standards focuses on potential multiple equilibria (Fields 1995; Basu and Van 1996). The economy could settle down into one of several equilibria, some of which may be Pareto inferior to others. Standards could potentially move the economy to a more efficient equilibrium, or they could reinforce an inefficient one. Little evidence is available to test whether standards help overcome market failures, however.

Race to the Bottom

One version of the race to the bottom model assumes that in some countries labor is exploited by the political or economic leadership (through forced labor, for example). If some countries exploit labor in this fashion, competitive forces will induce other countries to lower their labor standards as well or else risk higher unemployment. Echoing Gresham's law, Marshall (1994) argues that "a basic principle of highly competitive markets is that bad standards tend to drive out the good" (p. 67). Collectively, people in all countries could be better off with a minimal level of labor standards if some countries exploit labor and this exploitation leads to lower labor standards abroad.[4]

Freeman (1994) and Ehrenberg (1994) argue that, as a practical matter, a race to the bottom is unlikely. As evidence they note that states in the United States have maintained widely divergent labor standards for decades. Thus it is improbable that a race to the bottom will cause all countries to converge to a common, negligible set of labor standards. Still, on the margin a low level of standards in one country could put downward pressure on standards in other countries, especially as the world trading system becomes more integrated.[5] For example, U.S. Speaker of the House of Representatives Newt Gingrich and presidential candidate Patrick Buchanan both have argued against an increase in the minimum wage in the United States because the Mexican minimum wage is less than one-quarter the U.S. level.

Consumer Sovereignty

Consumers may consider the process by which products are made an important attribute of the product. Freeman (1994) argues that consumers often are willing to pay more for products that are made in socially responsible ways (without forced labor, for example), and this desire is likely to increase with income. A 1994 poll

provides some support for Freeman's view: 84 percent of U.S. shoppers said they would be willing to pay $1 extra for a $20 garment if it were made without sweatshop labor (Haq 1996). Freeman concludes that a proper role of government is to provide information on socially responsible companies and thereby to induce companies to provide better working conditions by altering market demand for their products. In its recent campaign against sweatshop labor the U.S. Department of Labor tried exactly this tactic, to some effect.

Taken to an extreme, individuals in one country may feel that it is morally unacceptable for a country to gain a comparative advantage based on certain labor practices, such as forced labor (Charnovitz 1992; Bhagwati 1994). Just as individuals may choose not to buy a particular product because they dislike some attributes of the product, so too may society collectively express these preferences by pursuing national and international labor standards. But it may or may not be more efficient for the government to collect the necessary data to pursue these preferences through international labor standards, as opposed to letting consumers express their preferences individually.

Rodrik (1995) draws an interesting parallel between international labor standards and restrictions on domestic technology. Formally, liberalized trade is equivalent to an improvement in technology, because trade enables goods to be "produced" at lower costs. Rodrik notes that nations often restrict the type of technology that domestic firms can use. Labor standards are a type of restriction on technology. Rodrik argues that "it is difficult to see why a particular sort of technology, that which is embodied in international trade, should be immune from the same type of considerations" (p. 10). To a worker in an industrial country who is displaced by an underage child laborer, it makes little difference whether that child works at home or abroad.

Enhanced Labor Market Institutions

Some labor standards are desirable because they enhance the efficient operation of the labor market. For example, Freeman (1993), Marshall (1994), and others argue that protected collective bargaining could enhance the operation of the labor market. Piore (1994) argues that labor market standards enhance the stability of social relationships and may lead to more efficient production strategies.

Labor market institutions may have positive spillovers to the rest of society. It is arguable that the Solidarity movement and Western pressure for free trade unions in Poland had as much of an impact on the rise of democracy and markets in Central and Eastern Europe as the U.S. military buildup. Many analysts argue that policies that protect free and democratic unions have collateral political benefits. Take the U.S. ordeal of forced labor as another example. In the nineteenth century European countries made little effort to link trade to the abolition of slavery. Had Britain refused to purchase cheap U.S. cotton produced with slave labor, there is a chance that the U.S. Civil War could have been averted or shortened. When judged against the small increase in prices that may result from international labor standards, the collateral political and social benefits could be quite large.

Why Link International Trade and Labor Practices?

In a well-functioning democracy the government has a strong incentive to set labor standards so as to overcome market failures, achieve desired redistribution, and enhance efficiency. Elected officials therefore would have a strong incentive to choose the "right" level of labor standards given their country's norms, culture, and level of economic development, because by choosing the right standards they maximize social welfare and increase their chances of being reelected. For the country itself the optimal labor standards will depend on the significance of market failures, the comparative efficacy of standards, the desired level of redistribution, and other factors. Technical assistance and nonbinding advice may help countries establish the set of standards that is best for them, but if the political system functions well, external influences should have little reason to interfere with labor standards in sovereign countries.

An exception to this optimistic conclusion would occur if the political system in a particular country were not responsive to the welfare of its citizens. For example, if the leadership of a country exploits its workers (say, through forced labor), this exploitation will put downward pressure on wages and working conditions in other countries (although consumers may pay less for goods). International pressure for labor standards could improve the welfare of workers in both countries if one country unfairly exploits its workers. This observation underscores U.S. Labor Secretary Robert Reich's 1994 position that, if a country "lacks democratic institutions and fails to disseminate the benefits of growth, other countries might justifiably conclude that low labor standards are due not to poverty itself, but to political choices that distort development and warp the economy's structure"(p. 4).

If one assumes that minimal labor standards are desirable, what are the arguments for or against using international trade as a point of leverage to enforce minimum labor standards in other countries? Some trade economists have taken the extreme position that any policy that interferes with free trade is disguised protectionism and so must be bad. I have six pragmatic observations on this issue.

- Multinational and other organizations often play an informational and monitoring role with respect to labor standards and working conditions. Better information should make political and economic markets work better. If the provision of information prods nations to strengthen their labor standards or enforcement, so be it.
- If the pressure for standards emanates solely from a desire to protect workers and companies in advanced economies, the standards may not be in the best interest of less advanced economies. In particular, a legitimate concern is that developing countries will be pressured to accept standards that exceed their economic capacity.
- In any event it is unlikely that minimal labor standards will provide much protection to workers in industrial countries. The gap between wages and working conditions for unskilled workers in industrial and developing economies is so great that any realistic set of minimal labor standards is

unlikely to have much impact on trade flows (Grossman and Krueger 1993). If the sole goal of labor standards is to help workers in industrial countries, there are probably more direct and more effective ways than international labor standards.

- Ehrenberg (1995) proposes the novel idea that industrial countries compensate developing countries for accepting and enforcing higher labor standards. Whatever the economic merits of this idea, it is unlikely to attract much support in industrial countries, which can barely muster the political support for trade agreements or for aid to developing countries.
- Labor standards strike me as a legitimate subject of bargaining in trade negotiations. Presumably, a well-intentioned government will not accept an agreement unless, in total, it is expected to make the country better off. No country has the right to impose its laws on another sovereign nation. Yet trade agreements are voluntarily agreed to. Side agreements may help produce a more equitable distribution of the surplus resulting from expanded trade. The NAFTA side agreements, which create institutions to ensure that parties to the agreement enforce their own labor laws, may serve as a model. Also note that if trade agreements increase national income, as is expected, countries will be able to afford more stringent labor standards following the agreements. Because the demand for labor standards tends to rise with national income, many countries will choose on their own to strengthen and enforce their standards following trade agreements.
- Political support for free trade is tenuous in many industrial countries, which have experienced rising income inequality and high unemployment. In these countries labor and environmental side agreements are likely to enhance political support for trade agreements. Given a choice between no trade agreement and an agreement that requires more vigorous enforcement of labor laws already on the books, my guess is that even the most hardened trade economist would choose the second option.

Disguised Protectionism?

Although a major concern with international labor standards is that they may be used as an excuse to inefficiently protect interests in industrial countries from competition from developing countries, there is surprisingly little empirical evidence—one way or the other—on this issue. One way to investigate whether legislators support international labor standards as a means of protecting domestic interests would be to examine whether support for such legislation is greatest among legislators whose constituents would benefit most from international labor standards. This hypothesis is particularly difficult to test, however, because labor standards are typically bundled with other trade legislation and rarely come up for a separate vote, such as was the case for the NAFTA side agreements.

One bill that provides an opportunity to study the determinants of political support for international labor standards is the proposed U.S. Child Labor Deterrence

Act of 1995 (S. 706 and H.R. 2065). If passed, this legislation would prohibit U.S. imports of goods produced abroad with child labor.[6] Senator Tom Harkin (Democrat-Iowa) sponsored this legislation in the Senate, and Representative Barney Frank (Democrat-Massachusetts) in the House of Representatives. Although the bill has not come to a vote, it has thirty-five cosponsors in the House and seven in the Senate. Cosponsoring legislation is an indication of strong support for a bill. I have assembled a data set to study the determinants of support for this trade-linked child labor standard.

Specifically, I relate whether a member of the House is a cosponsor of the Child Labor Deterrence Act to characteristics of the representatives' districts and the representatives' political background.[7] Plausibly, constituents with a low level of education (and their employers) are those most likely to benefit directly from this act because imported products made with child labor are most likely to compete with domestic products produced by less-educated workers in the United States. So if support for banning imports of products made with child labor in part represents a concealed desire to protect constituents from foreign competition, support for this legislation would be expected to be strongest among legislators who represent districts with relatively more high school dropouts. To test this proposition, I collected data for each congressional district on the share of the population age twenty-five and older with less than a high school degree.[8]

Estimation results of several linear probability models relating cosponsorship of the Child Labor Deterrence Act (dependent variable) to other characteristics of the representative and his or her district are reported in table 1.[9] In addition to the share of high school dropouts, I control for a number of other variables. The union participation variable measures the proportion of the workforce in the state that belongs to a union, based on Hirsch (1993). Votes on NAFTA and the General Agreement on Tariffs and Trade (GATT), the representative's rating by Americans for Democratic Action (ADA), party affiliation, the popular vote for the representative in the 1994 election, and the number of terms served by the representative are all from Duncan and Lawrence (1995). The sample consists of 434 members of the 104th Congress. (The sample size is not 435 because one seat was open in 1995.)

The results indicate that representatives from districts with a high concentration of high school dropouts are less likely to cosponsor the Child Labor Deterrence Act. And the magnitude of the effect is fairly large: going from a district with 10 percent high school dropouts to one with 30 percent lowers the probability of sponsoring the act by roughly 8 percentage points, other things being equal. This finding is contrary to what I would expect from a simple political economy model that says members of Congress whose constituents would benefit most from the Child Labor Deterrence Act are most likely to support the act.

Several other variables reported in table 1 are also of interest. First, representatives from states that have a higher union participation rate are more likely to cosponsor the act. There are two plausible explanations for this finding: unionized workers receive rents, and unions therefore try to prevent foreign competition from eroding those rents; and unionized workers and their representatives are concerned

Table 1. Determinants of Congressional Support for a Ban on Imports of Goods Produced with Child Labor

Variable	Mean [standard deviation]	All 1	All 2	All 3	Democrats 4	Democrats 5	Democrats 6
Proportion of high school dropouts	0.25 [0.09]	−0.29* (0.17)	−0.44* (0.20)	−0.41* (0.19)	−0.59* (0.31)	−0.62* (0.31)	−0.56* (0.30)
Union participation	0.16 [0.07]	0.56* (0.18)	0.49* (0.23)	0.20 (0.22)	1.02* (0.40)	0.65 (0.44)	0.26 (0.45)
For NAFTA and GATT	0.44 [0.50]	—	−0.11* (0.04)	−0.11* (0.04)	—	−0.16* (0.06)	−0.16* (0.06)
For NAFTA, not GATT	0.09 [0.29]	—	−0.08* (0.03)	−0.07* (0.03)	—	−0.13* (0.06)	−0.11* (0.07)
For GATT, not NAFTA	0.20 [0.40]	—	−0.02 (0.06)	−0.03 (0.06)	—	−0.03 (0.07)	−0.03 (0.07)
Americans for Democratic Action rating	0.45 [0.36]	—	—	0.31* (0.07)	—	—	0.35* (0.09)
Democrat	0.46 [0.50]	0.18* (0.03)	0.17* (0.03)	−0.03 (0.04)	—	—	—
Vote in 1994	0.66 [0.13]	0.12 (0.10)	0.16 (0.13)	0.11 (0.12)	0.30 (0.27)	0.29 (0.27)	0.18 (0.26)
Number of terms	4.84 [4.03]	0.001 (0.004)	0.002 (0.004)	0.002 (0.004)	0.002 (0.006)	0.003 (0.006)	0.003 (0.006)
Sample size	434	434	345	345	188	188	188

* Statistically significant at the 0.10 level.
Note: The mean of the dependent variable is 0.08 in column 1, 0.09 in columns 2 and 3, and 0.17 in columns 4, 5, and 6. Data pertain to members of the U.S. House of Representatives in 1995. Estimates are from a linear probability model. Numbers in parentheses are White standard errors. The dependent variable equals one if the representative is a cosponsor of the Child Labor Deterrence Act and zero if he or she is not.
Source: Author's calculations.

about labor rights generally. As discussed below, because unionized workers are unlikely to compete with child labor, the first explanation is questionable.

For representatives who also served in the 103rd Congress, it is possible to examine the relationship between support for NAFTA and GATT and support for the Child Labor Deterrence Act. Specifically, columns 2 and 3 include dummy variables that measure whether the representative voted for NAFTA and for GATT, for NAFTA and against GATT, and for GATT and against NAFTA. The results indicate that representatives who supported NAFTA and GATT were 11 percentage points less likely to support the Child Labor Deterrence Act than were representatives who opposed both NAFTA and GATT. Thus representatives who opposed expanding trade opportunities through NAFTA and GATT were also more likely to support the Child Labor Deterrence Act. This finding suggests that representatives who support international labor standards are more likely to favor protectionist policies generally, but the findings for the education variable make it unclear whether those protectionist policies are in the narrow interest of their constituents. Below I examine the determinants of support for NAFTA and GATT to gain further insights into this issue.

Members of the Democratic Party (defined to include the one Independent member of Congress) are more likely to support child labor standards. Indeed, none of

the cosponsors of the Child Labor Deterrence Act in the House is a Republican (although two Republicans cosponsored the bill in the Senate). In columns 4–6 I reestimated the models for the subsample of Democrats. The results for this subsample are qualitatively similar; most important, representatives of districts with many less educated workers continue to be less likely to support the Child Labor Deterrence Act.

Support for international labor standards may just reflect representatives' ideology. Unfortunately, ideology is hard to define and quantify. The ADA rating measures the "liberalness" of the representative's voting record, as reflected by votes for bills supported by the ADA. This variable, intended to reflect representatives' ideology, is positively associated with support for the Child Labor Deterrence Act.[10] Including this variable hardly alters the effect of the other variables except for the dummy variable measuring party affiliation, which becomes statistically insignificant. Finally, note that the variables measuring representatives' share of the vote in 1994 and number of terms in office both have statistically insignificant effects in all the specifications.

Comparison with NAFTA and GATT Votes

It is useful to contrast the model results reported here with the results of comparable models of the determinants of support for NAFTA and GATT. Specifically, are representatives from districts with many less educated workers more or less likely to vote for NAFTA and GATT? Estimates of linear probability models are presented in table 2. The sample consists of the subset of the members of Congress who were in office in 1994 and 1995. Even though the vote on GATT was held in a lame duck session of Congress, this sample could be expected to be responsive to constituents' interests because it consists exclusively of members who returned to Congress.

The independent variables are much more successful at explaining support for NAFTA than for GATT.[11] Interestingly, representatives with a large share of less educated workers in their district are more likely to oppose NAFTA and GATT, even though they are less likely to support international child labor standards. This finding suggests that the share of less educated workers in a district at least partly reflects a constituency that stands to benefit from protectionist policies.

A higher union participation rate and membership in the Democratic Party are also strongly negatively related to votes for NAFTA. McArthur and Marks (1988) similarly find that a high union participation rate and union political contributions are strong predictors of votes for the 1982 automobile industry domestic content bill, and Baldwin (1985) finds that union political contributions are correlated with opposition to the Trade Act of 1974. The finding of a negative relationship between unionization and support for NAFTA is not surprising, since the American Federation of Labor–Congress of Industrial Organizations (AFL-CIO) strongly opposed NAFTA. The ADA rating is insignificantly related to support for NAFTA or GATT, but in other results I have found that the AFL-CIO's political rating scale has a significant negative effect on votes for NAFTA and GATT.[12]

Table 2. Determinants of Support for NAFTA and GATT

	NAFTA		GATT	
Variable	*1*	*2*	*3*	*4*
Proportion of high school dropouts	–0.52*	–0.53*	–0.50	–0.47
	(0.28)	(0.30)	(0.31)	(0.31)
Union participation	–1.68*	–1.64*	0.18	0.10
	(0.36)	(0.37)	(0.39)	(0.41)
Americans for Democratic Action rating	—	–0.05	—	0.05
		(0.13)		(0.15)
Democrat	–0.37*	–0.34*	–0.04	–0.09
	(0.06)	(0.10)	(0.06)	(0.11)
Vote in 1994	–0.06	–0.05	–0.22	–0.26
	(0.19)	(0.19)	(0.21)	(0.21)
Number of terms	0.004	0.004	–0.001	–0.002
	(0.006)	(0.006)	(0.007)	(0.007)
Sample size	345	345	349	347

* Statistically significant at the 0.10 level.
Note: 53 percent of the sample voted for NAFTA; 63 percent voted for GATT. Coefficients are from a linear probability model. Numbers in parentheses are White standard errors. The dependent variable equals one if the representative voted in favor of NAFTA (or GATT) and zero if he or she voted against.
Source: Author's calculations.

A consideration of the major players in the legislative battles over NAFTA and GATT yields additional insights. Senator Ernest Hollings (Democrat-South Carolina) led the unsuccessful campaign against GATT, although he did succeed in postponing the vote to a lame duck session. A majority of the South Carolina congressional delegation also voted against GATT and NAFTA. It was widely believed that Senator Hollings was motivated by a concern to protect textile and apparel firms, which are major employers in South Carolina. The textile and apparel industry is also a major employer of children abroad. Yet not one of the current cosponsors of the Child Labor Deterrence Act is from South Carolina. Since South Carolina stands to benefit as much as any U.S. state from a ban on imports produced by child labor, the lack of support by South Carolina representatives also suggests that support for the act is not motivated by disguised protectionism.

A broader literature on the political economy of tariffs also finds that opposition to trade liberalization is related to constituents' economic interests, similar to the results of the NAFTA and GATT regressions. Baldwin (1985), for example, finds that representatives from districts with relatively many workers in import-sensitive industries were more likely to oppose the Trade Act of 1974. In related work Tosini and Tower (1987) examine support for the Textile Act of 1985, which would have established quotas to restrict imports of textile goods to the United States. Legislators were much more likely to support this bill if their district had a high percentage of workers in the textile industry and if they received funds from the Amalgamated Clothing and Textile Workers Union. They were more likely to oppose the bill if their districts had a high percentage of workers in export industries. Thus although support for child labor standards does not appear to be related

to constituents' economic interests, support for tariffs and quotas does appear to reflect constituents' economic interests.

Interpretation

There are a variety of ways to interpret these results. In my view, however, the weight of the evidence does not support the hypothesis that advocacy of international labor standards reflects disguised protectionism. Representatives from districts that stand to gain the most from the Child Labor Deterrence Act—those with a large number of unskilled workers—are least likely to cosponsor the act. Moreover, representatives from districts with a high fraction of less-skilled workers tend to be more opposed to NAFTA and GATT (see table 2). An alternative explanation for these results is, as Freeman (1994) argues, that demand for international child labor standards is a normal good (meaning that demand rises with income) and that people with higher socioeconomic status select representatives who are more supportive of placing limits on child labor. Of course, it is possible that support for other types of labor standards represents disguised protectionism. But given the results for the Child Labor Deterrence Act, it seems to me to be incumbent on those who view international labor standards as disguised protectionism to provide econometric evidence to support that conclusion.

Another issue concerns the strong effect of the union participation variable on support for the Child Labor Deterrence Act. One may be tempted to conclude from this result that unions support international labor standards out of self-interest (to raise the incomes of their members). Studies of legislative support for the minimum wage have reached this conclusion from similar evidence; for example, Bloch (1980) and Cox and Oaxaca (1981) interpret a positive correlation between a state's unionization rate and support for a minimum wage increase as evidence that (generally high-wage) unionized workers benefit from minimum wage legislation because union workers are substitutes for minimum wage workers. I find that interpretation to be strained in this case, however, because unionized workers are not obvious beneficiaries of a ban on imports made with child labor—and certainly stand to benefit less than high school dropouts from such a ban. An alternative interpretation is that unions are pursuing policies that strengthen worker rights generally, rather than merely maximizing the self-interest of their members. Sometimes, seemingly altruistic behavior stems from altruistic motives.

My own experience suggests that union leaders actively support labor standards and that, in many instances, labor standards would not receive any attention if it were not for unions. U.S. labor unions have supported stronger labor side agreements in NAFTA, lobbied Congress to increase the budget of the ILO, and pressed the administration of President Clinton to add labor standards to the agenda of international summits. Yet in many cases I do not think that the union leadership effectively furthers its members' narrow interests by pushing these policies. Indeed, in many instances I am surprised that the AFL-CIO uses its limited political capital to press for international labor standards that are of little benefit to its members instead of pur-

suing policies with much greater direct benefit to its membership. Indeed, in the recent AFL-CIO presidential election Lane Kirkland was criticized for pursuing international labor standards at the expense of domestic union issues. This is not to suggest that unions never pursue legislation that benefits their membership at the expense of others. They do. But it does suggest that a positive association between unionization and support for international labor standards does not prove that unions support standards for the sole reason of enhancing the economic position of their members.

Two final points are in order. First, the conclusion that support for a child labor ban does not stem primarily from disguised protectionist motives does not mean that such a ban is economically efficient or desirable. Second, if support for the ban were motivated primarily by disguised protectionism, it might still be economically efficient. To evaluate the desirability of standards, it is necessary to examine their actual impact. Are labor standards complied with? What is their economic effect? Are there better ways of achieving the same ends? These questions are partially addressed in the next section, in the context of child labor and compulsory schooling.

Child Labor Standards and Compulsory Schooling

It is widely agreed that exploitative child labor and the employment of very young children should be prohibited. First, child workers, like all other workers, should not be exploited. Furthermore, very young children are not capable of making rational employment and schooling decisions on their own, although their families often help them to make sensible choices. The primary economic approach to modeling child labor is to assume that rational time allocation decisions are made jointly by children and their families. As Grootaert and Kanbur (1995) emphasize, child labor standards could alter the bargaining power and welfare of children while weakening the economic position of their families. Basu and Van (1996), by contrast, note that a ban on child labor could push up the wages of adults, possibly to the point where families are wealthy enough that they no longer want their children to work. Another concern, however, is that a prohibition on child labor in one sector could force children into less desirable activities in other sectors, such as the underground economy.

The effect of education policy on child labor supply has largely been ignored in this literature. If the government provides higher-quality education, the incentive for students to acquire education and postpone work will be greater. If schools are not available nearby or are of low quality, then child labor becomes a more attractive choice. It is possible for children to make rational decisions to work rather than attend school based on their existing set of schooling opportunities, but the schooling opportunities may be suboptimal.

Patterns of Child Labor Activity

How widespread is child labor? The question is hard to answer because the practice is difficult to define and measure. Statistical agencies often do not collect information on labor force status for children below the minimum work age. ILO Convention 138 on

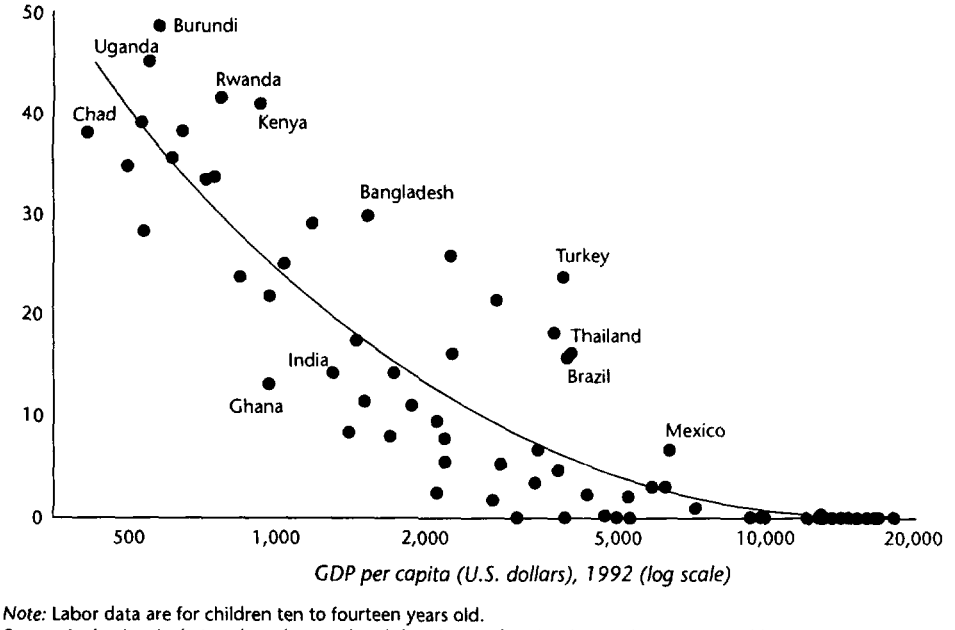
child labor sets the minimum work age at fifteen but permits a lower age for developing countries. (The ILO convention also permits light work for children age thirteen to fifteen, provided it does not interfere with their educational activities.) Forty-six countries have ratified this convention. Yet based on a collection of data from 124 countries, the ILO estimates that in 1990 some 78.5 million children ages fourteen and under worked worldwide (Ashagrie 1993). This high rate of employment suggests a significant amount of noncompliance with the child labor standard in many countries.

Not surprisingly, as the data in figure 1 indicate, employment of young children is common in low-income countries and uncommon in high-income countries.[13] Employment rates are highest for children in the low-income countries of Burundi (49 percent), Uganda (45 percent), and Rwanda (42 percent). In countries where GDP per capita exceeds $5,000, including most of Western Europe and North America, employment of young children is negligible. Cross-country differences in log GDP per capita and its square account for an impressive 80 percent of the variability in child employment rates worldwide.

Although child labor is considered a necessary source of production and income in many developing countries, it is considered an abomination in many other countries. The ILO finds that nearly 80 percent of child laborers work in agriculture, hunting, forestry, and fishing. In many cases this work may not interfere with a child's normal social and educational development. An important lesson is that the minimum work age should be tailored to each country's economic and social conditions.

Figure 1. Child Labor and GDP per Capita

Percentage of children working, 1995



Note: Labor data are for children ten to fourteen years old.
Source: Author's calculations based on updated data from Ashagrie 1993 and Summers and Heston 1991 (Penn World Table 5.6).

Another important lesson of figure 1 is that a higher level of economic development is associated with a decline in child labor. Many goods are normal goods, meaning that their consumption rises as income rises. Child labor could be thought of as a "normal bad," a practice that is tolerated when societies are poor but not when they are wealthy. Thus one would expect more developed economies to have stronger child labor standards, and they typically do. Also note that the number of child laborers under the age of fifteen declined by 11 percent between 1980 and 1990 (although much of this decline may well represent a reporting phenomenon; Ashagrie 1993). Economic growth appears to be an important way to reduce child labor. If trade agreements increase the wealth of nations, then developing countries that are a party to such agreements would be expected to more readily adopt child labor standards after trade has expanded.

Compulsory Schooling

Compulsory schooling laws and minimum work age requirements are usually complementary policies. By requiring that all children attend primary school, for example, the state effectively removes them from the labor force (Weiner 1991). The ILO recommends that children receive full-time schooling or vocational training at least until they reach the minimum work age. A number of authors have recently suggested that compulsory schooling be emphasized as a policy to reduce child labor (see, for example, Weiner 1991). Given the wide prevalence of compulsory schooling laws, I examined evidence on compliance with such laws, using data from the 1990–93 World Values Survey, a set of international cross-sectional surveys conducted by the European Values Systems Study Group. In each country respondents were asked at what age they had completed (or will complete) full-time education.

A look at the school leaving age distribution for people born between 1959 and 1974 in selected low-income countries and in the United Kingdom and the United States shows a tremendous amount of noncompliance with compulsory schooling laws (table 3). In Brazil 80 percent of children left school before reaching their thirteenth birthday, even though the compulsory schooling age is fourteen, and in Mexico and Portugal a quarter of the population left school before reaching the minimum schooling age. In India, where the compulsory schooling age varies across regions and often is as low as eleven, 40 percent of the population left school at age twelve or younger. Weiner (1991) argues that India's lack of commitment to a national compulsory schooling policy is a major reason a large share of the population is illiterate. In the United States, where sixteen is the most common compulsory schooling age across states, few students leave school before age seventeen. In the United Kingdom half of students leave school at age sixteen, the compulsory age.

Although there is clearly noncompliance with compulsory schooling laws, the laws still may lead to higher educational attainment than would otherwise be the case. The compulsory schooling age has affected educational attainment in the United Kingdom, as table 4 clearly shows. The compulsory schooling age was raised from fourteen to fifteen in 1947 and then from fifteen to sixteen in 1973. The com-

parison in table 4 of the school leaving age distribution for three cohorts—one covered by the age fourteen law, one by the age fifteen law, and one by the age sixteen law—shows that for each cohort the modal school leaving age equals the minimum compulsory level. Moreover, for each cohort no more than 5 percent of individuals leave school before reaching the minimum age. Harmon and Walker (1995) simi-

Table 3. Distribution of School Leaving Age, 1959–74 Birth Cohorts
(percent)

School leaving age	*Argentina*	*Brazil*	*Chile*	*India*	*Mexico*	*Nigeria*	*Portugal*	*United Kingdom*	*United States*
Younger than 13	6.6	80.2	2.4	40.5	25.2	4.2	17.4	0.0	0.6
13	5.9	5.8	2.0	4.2	1.3	2.0	7.4	0.4	0.2
14	5.9	3.7	2.7	5.1	1.8	7.5	10.7	0.0	0.0
15	3.5	3.5	3.7	3.2	5.1	8.6	4.6	5.3	1.8
16	4.5	2.1	5.2	3.6	4.0	6.4	5.4	50.1	2.6
17	10.8	2.1	12.3	4.3	3.8	6.6	4.6	10.3	15.0
18	16.0	1.0	21.9	6.5	7.8	14.3	7.8	11.2	29.7
19	8.0	0.7	10.4	4.9	5.3	7.8	0.0	0.0	4.4
20	4.2	0.4	8.1	4.5	6.5	8.5	10.6	7.4	4.9
Older than 20	34.5	0.4	31.3	23.3	39.1	34.1	31.6	15.2	40.8
Average age	18.0	12.6	18.5	15.8	17.4	18.2	17.0	17.3	19.1
Compulsory schooling age, 1975	14	14	15	11–14	14	n.a.	14	16	16–18
Sample size	312	691	597	1,116	789	600	407	370	395

n.a. is not applicable (no compusory schooling law).
Note: In India and the United States the compulsory schooling age varies across regions. Data have been weighted to adjust for nonrandom sampling.
Source: Author's calculations based on World Values Surveys 1990–93 and UNESCO *Statistical Yearbook* 1975.

Table 4. School Leaving Age in the United Kingdom for Birth Cohorts Affected by Varying Compulsory Schooling Laws
(percent)

	Birth cohort		
School leaving age	*Before 1933*	*1933–56*	*1958–74*
Younger than 13	0.7	0.6	0.0
13	3.4	0.7	0.4
14	**63.5**	2.5	0.0
15	5.1	**40.4**	5.4
16	10.7	26.1	**50.5**
17	4.8	9.7	10.5
18	3.5	5.3	10.1
20	1.9	1.7	8.0
Older than 20	6.5	13.1	15.1
Compulsory schooling age	14	15	16
Sample size	450	574	400

Note: Data have been weighted to adjust for nonrandom sampling.
Source: Author's calculations based on World Values Surveys.

larly find that educational attainment shifted out with the increases in the compulsory schooling age in the United Kingdom.

In developing countries it is less clear that the compulsory schooling age has much effect on educational attainment. None of the low-income countries in table 3 shows much of a spike in school completion around the compulsory schooling age. Child labor and low educational attainment are particularly a problem in Brazil. An estimated 10 percent of Brazilian children work on the streets for their own or their family's survival (Myers 1988). Brazil increased its compulsory schooling age from eleven to fourteen in 1971, yet table 5 shows that there is hardly any difference in school leaving ages for the cohort (born after 1962) covered by the age-fourteen law and the cohort (born before 1958) covered by the age-eleven law. Regardless of the compulsory schooling age, roughly 85 percent of children left school before reaching the age of fourteen.

In the United Kingdom and the United States compulsory schooling has been found to lead to higher earnings. Angrist and Krueger (1991) and Harmon and Walker (1995) find that the earnings payoff to years of compulsory schooling tends to exceed the payoff to years of schooling beyond the compulsory level. For several reasons policies that increase educational attainment in developing countries are likely to have large rewards as well. First, primary education pays a higher return than secondary and higher education in developing countries (Psacharopoulos 1994), and compulsory schooling laws typically pertain to primary schooling. Second, fertility rates tend to decline with maternal education in developing countries. Finally, infant and child mortality rates tend to decline with maternal education as well (World Bank 1995a).

By themselves, compulsory schooling laws are unlikely to increase educational attainment or to reduce child labor. Increasing educational attainment would

Table 5. School Leaving Age in Brazil for Birth Cohorts Affected by Varying Compulsory Schooling Laws
(percent)

	Birth cohort	
School leaving age	*Before 1958*	*After 1962*
Younger than 13	82.2	80.2
13	2.7	5.8
14	2.4	3.7
15	3.3	3.5
16	3.3	2.1
17	1.9	2.1
18	1.4	1.0
19	0.6	0.7
20	1.5	0.4
Older than 20	0.8	0.4
Compulsory schooling age	11	14
Sample size	1,091	691

Note: Data have been weighted to adjust for nonrandom sampling.
Source: Author's calculations based on World Values Surveys.

require adequate educational facilities, vigorous enforcement, and a commitment on the part of parents and policymakers to fostering education. Compulsory schooling laws can form an important component of child labor policy, but unless communities have adequate schools and families have the financial resources and will to send their children to school, noncompliance will be rampant.

More generally, low compliance with labor standards is often a major issue in developing countries (World Bank 1995b). If countries lack the capacity or will to enforce their labor standards, there is little value in the international community's pressing for more stringent standards. I believe that this energy would be better spent encouraging nations to enforce the laws that they already deem adequate. For example, industrial countries could provide technical assistance and financial support to developing countries to aid in the enforcement of labor standards.

Conclusion

A review of the theoretical literature suggests that labor standards could enhance the efficiency of the labor market and improve the distribution of income in some situations but could prove counterproductive for efficiency and equity in others. Interestingly, after noting the unequal political power of employers and employees, Adam Smith (1937 [1776]) concluded that "when the regulation, therefore, is in favor of the workmen, it is always just and equitable; but it is sometimes otherwise when in favor of the masters" (p. 142). In modern democracies, however, I would second Fields's (1995) observation: "There is no easy generalisation, and the 'less is better' view is as imbalanced on one side as is the 'more is better' view on the other" (p. 11).

Support for international labor standards in advanced economies does not necessarily represent disguised protectionism. Although a large literature finds that political support for tariffs and quotas at least in part reflects a desire to protect constituents' economic interests, the evidence does not suggest that politicians support international labor standards out of a desire to further the narrow economic interests of their constituents. Thus pressure for international labor standards cannot be dismissed automatically as disguised protectionism. Standards may or may not serve a useful purpose, but they must be evaluated on their merits.

Rich countries tend to have more stringent labor standards and better working conditions than do poor countries. In particular, economic development is inversely related to the use of child labor. The costs of labor standards are probably borne by the country with the standards, in the form of lower wages, higher product prices, or devalued currency. Many labor standards are normal goods, for which demand is likely to increase with economic growth. Thus standards that sacrifice economic growth could have a negative effect on working conditions in the long run.

Unless there is a concerted effort to enforce standards, and unless the standards are appropriate for the economic conditions in a country, standards are unlikely to have much impact. If standards are set too high, they will be widely ignored. Also, reliance on compulsory schooling laws as an alternative to labor market standards

will have little effect unless the laws are enforced, sufficient schools are available, and attendance is sufficiently valued by parents and children.

From an analytical standpoint there is much to be said for treating labor standards as normal goods, which are desired and consumed in greater quantity when income is higher. Although the political economy model does not adequately explain why some U.S. legislators support the Child Labor Deterrence Act, a view of labor standards as a normal good does. Representatives from better-off districts behave as though their constituents have a stronger desire than those in less well-off districts to avoid products made with child labor, even though such actions may require their constituents to pay more for the products they consume. Similarly, rich countries tend to impose stronger labor standards on themselves and are less likely to employ child labor. An important unresolved question is whether there are more efficient ways of satisfying people's demands for better treatment of workers and children than by pressuring foreign nations to adopt labor standards or by refusing to purchase goods made under conditions deemed substandard. For example, an industrial country could transfer income directly to poor children in developing countries, admit more immigrants, or subsidize employers to improve the conditions of workers and children in developing countries. If, however, labor standards enhance efficiency, then the cost of exercising those preferences through international labor standards could be small or nonexistent.

Notes

1. Some observers draw a distinction between labor standards (such as minimum work ages) and labor rights (such as the right to bargain collectively). Although this distinction is meaningful, in this article the term labor standards covers both standards and rights.

2. Competition over rents created by redistributional policies also may reduce efficiency because resources are devoted to rent seeking rather than to production (see Krueger 1974).

3. World Bank (1995b) lists several market failures as a rationale for labor standards. Maskus, Rutherford, and Selby (1995) provide a simulation of the effect of labor standards in Mexico that assumes workers are misinformed about work-related hazards.

4. Davis (1996) provides an interesting theoretical model in which Europe is assumed to have a binding minimum wage and the United States is assumed to have flexible wages. His model predicts that Europe incurs enough unemployment to raise the wage of low-skilled U.S. workers to the European minimum wage if trade arbitrages goods prices between the countries.

5. Freeman (1994) writes, "I do not accept the premise of some that bad standards drive out good standards. Any country that wants higher labour standards for itself can have them....if it is willing to pay" (p. 87). My point is that lower standards abroad may alter the price that the country wanting higher standards will have to pay.

6. The Child Labor Deterrence Act would urge the president to seek an agreement with other governments to secure an international ban on trade in goods produced by children under age fifteen. In addition, it would require the secretary of labor to identify foreign countries that do not comply with national laws that prohibit child labor and that use child labor to produce exports. After consultations with the U.S. trade representative and secretaries of state, treasury, and commerce and at least one public hearing, the importation of such products from these countries could be prohibited.

7. With only 7 cosponsors in the Senate, this type of statistical analysis does not make much sense for the 100 members of the Senate.

8. This variable was derived from the STF3 file of the 1990 Census, which pertains to the 103rd Congress.

9. Since the dependent variable equals either zero or one, and the mean is rather low, a linear probability model is not strictly speaking appropriate. To partially address this issue, I have reported standard errors that correct for heteroskedasticity. More important, I also have reestimated the regressions using

a logistic model and found qualitatively similar results. For simplicity, I report the linear probability models.

10. Results were quite similar when I used the American Federation of Labor–Congress of Industrial Organization's (AFL–CIO) political rating scale instead of the ADA rating.

11. Indeed, a chi square test finds that the variables in the GATT equations are jointly statistically insignificant.

12. This may result, in part, because voting on NAFTA enters into the AFL–CIO's evaluation criteria.

13. In addition, numerous microdata studies have found that child labor is negatively related to family income within countries (Rosenzweig and Evenson 1977).

References

Akerlof, George. 1976. "The Economics of Caste and of the Rat Race and Other Woeful Tales." *Quarterly Journal of Economics* 90 (4): 599–617.

Angrist, Joshua, and Alan Krueger. 1992. "Does Compulsory Schooling Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106 (4): 979–1014.

Ashagrie, Kebebew. 1993. "Statistics on Child Labour: A Brief Report." *Bulletin of Labor Statistics.* International Labor Office, Geneva, Switzerland.

Baldwin, Robert. 1985. *The Political Economy of U.S. Import Policy.* Cambridge, Mass: MIT Press.

Basu, Kaushik, and Pham Hoang Van. 1995. "The Economics of Child Labor: Theory, Experience, and Legislation." Cornell University, Department of Economics, Ithaca, N.Y.

Bhagwati, Jagdish. 1994. "Policy Perspectives and Future Directions: A View from Academia." In *International Labor Standards and Global Economic Integration: Proceedings of a Symposium.* Washington, D.C.: U.S. Department of Labor, Bureau of International Labor Affairs.

Bloch, Farrell. 1980. "Political Support for Minimum Wage Legislation." *Journal of Labor Research* 1 (2): 245–54.

Brown, Drusilla, Alan Deardorff, and Robert Stern. Forthcoming. "International Labor Standards and Trade: A Theoretical Analysis." In Jagdish Bhagwati and Robert Hudec, eds., *Harmonization and Fair Trade.* Cambridge, Mass.: MIT Press.

Card, David, and Alan Krueger. 1995. *Myth and Measurement: The New Economics of the Minimum Wage.* Princeton, N.J.: Princeton University Press.

Charnovitz, Steve. 1992. "Environmental and Labour Standards and Trade." *World Economy* 15 (3): 335–56.

Cox, James C., and Ronald L. Oaxaca. 1981. "The Determinants of Minimum Wage Levels and Coverage in State Minimum Wage Laws." In Simon Rottenberg, ed., *The Economics of Legal Minimum Wages.* Washington, D.C.: American Enterprise Institute.

Davis, Donald. 1996. "Does European Unemployment Prop Up American Wages?" Harvard University, Department of Economics, Cambridge, Mass.

Duncan, Philip, and Christine Lawrence. 1995. *Congressional Quarterly's Politics in America, 1996: The 104th Congress.* Washington, D.C.: Congressional Quarterly.

Ehrenberg, Ronald. 1994. *Labor Markets and Integrating National Economies.* Washington, D.C.: Brookings Institution.

Fields, Gary. 1995. "Trade and Labour Standards: A Review of the Issues." Organization for Economic Cooperation and Development, Paris.

Freeman, Richard. 1993. "Labor Market Institutions and Policies: Help or Hindrance to Economic Development?" In Lawrence H. Summers and Shekhar Shah, eds., *Proceedings of the World Bank Annual Conference on Development Economics 1992.* Washington, D.C.: World Bank.

———. 1994. "A Hard-Headed Look at Labour Standards." In Werner Sengenberger and Duncan Campbell, eds., *International Labor Standards and Economic Interdependence.* Geneva, Switzerland: International Institute for Labor Studies.

Grootaert, Christiaan, and Ravi Kanbur. 1995. "Child Labour: An Economic Perspective." *International Labour Review* 134 (2): 187–204.

Grossman, Gene, and Alan Krueger. 1993. "Environmental Impacts of a North American Free Trade Agreement." In Peter Garber, ed., *The U.S.-Mexico Free Trade Agreement.* Cambridge, Mass.: MIT Press.

Hansson, Göte. 1983. *Social Clauses and International Trade.* New York: St. Martin's.

Haq, Farhan. 1996. "Critics Link Immigration Laws to Sweatshops." Inter Press Service, March 26.

Harmon, Colm, and Ian Walker. 1995. "Estimates of the Economic Return to Schooling in the U.K." *American Economic Review* 85 (5): 1278–86.

Hirsch, Barry T. 1993. *Union Membership and Earnings Data Book 1993.* Washington, D.C.: Bureau of National Affairs, Inc.

Krueger, Anne. 1974. "The Political Economy of the Rent-Seeking Society." *American Economic Review* 64 (3): 291–303.

Marshall, Ray. 1994. "The Importance of International Labour Standards in a More Competitive Global Economy." In Werner Sengenberger and Duncan Campbell, eds., *International Labour Standards and Economic Interdependence.* Geneva, Switzerland: International Labor Office.

Maskus, Keith, Thomas Rutherford, and Steven Selby. 1995. "Economic Implications of Changes in Labor Standards: A Computational Analysis for Mexico." University of Colorado, Department of Economics, Boulder.

McArthur, John, and Steven Marks. 1988. "Constituent Interests versus Legislator Ideology: The Role of Political Opportunity Costs." *Economic Inquiry* 26 (3): 461–70.

Myers, William. 1988. "Alternative Services for Street Children: The Brazilian Approach." In Assefa Bequele and Jo Boyden, eds., *Combating Child Labor.* Geneva, Switzerland: International Labor Office.

Piore, Michael. 1994. "International Labor Standards and Business Strategies." In *International Labor Standards and Global Economic Integration: Proceedings of a Symposium.* Washington, D.C.: U.S. Department of Labor, Bureau of International Labor Affairs.

Psacharopoulos, George. 1994. "Returns to Investment in Education: A Global Update." *World Development* 22 (9): 1325–43.

Reich, Robert. 1994. "Keynote Address." In *International Labor Standards and Global Economic Integration: Proceedings of a Symposium.* Washington, D.C.: U.S. Department of Labor, Bureau of International Labor Affairs.

Rodrik, Dani. "Labor Standards In International Trade: Do They Matter and What Do We Do About Them?" Columbia University, Department of Economics, New York.

Rosenzweig, Mark, and Robert Evenson. 1977. "Fertility, Schooling and the Economic Contribution of Children in Rural India: An Economic Analysis." *Econometrica* 45 (5): 1065–80.

Smith, Adam. 1937 [1776]. *An Inquiry into the Nature and Causes of the Wealth of Nations.* New York: Random House.

Squire, Lyn, and Sethaput Suthiwart-Narueput. 1995. "The Impact of Labor Market Regulations." Policy Research Working Paper 1418. World Bank, Policy Research Department, Washington, D.C.

Srinivasan, T. N. 1994. "International Labor Standards Once Again!" In *International Labor Standards and Global Economic Integration: Proceedings of a Symposium.* Washington, D.C.: U.S. Department of Labor, Bureau of International Labor Affairs.

Summers, Robert, and Alan Heston. 1991. "Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–88." *Quarterly Journal of Economics* 106 (May): 327–68.

Swinnerton, Kenneth. 1996. "An Essay on Economic Efficiency and Core Labor Standards." U.S. Department of Labor, Bureau of International Labor Affairs, Washington, D.C.

Tosini, Suzanne, and Edward Tower. 1987. "The Textile Bill of 1985: The Determinants of Congressional Voting Patterns." *Public Choice* 54 (1): 19–25.

Weiner, Myron. 1991. *The Child and the State in India.* Princeton, N.J.: Princeton University Press.

World Bank. 1995a. *Priorities and Strategies for Education: A Review.* Washington, D.C.

———. 1995b. *World Development Report 1995: Workers in an Integrating World.* New York: Oxford University Press.

World Values Survey. 1994. *World Values Survey 1981–84 and 1990–93.* Computer file. Inter-University Consortium for Political and Social Research. Ann Arbor, Mich.

Comment on "International Labor Standards and Trade," by Alan B. Krueger

Adolfo Figueroa

Alan Krueger poses several questions about labor standards. First, do higher labor standards have a positive effect on efficiency and equity? The literature provides no clear answer, Krueger concludes. Implicitly, this question considers labor standards as an exogenous variable. Second, what factors determine labor standards? This question makes labor standards an endogenous variable. In the final analysis, are labor standards endogenous or exogenous to the economic process?

Krueger posits that a country's labor standards may depend on several factors, including culture, degree of democracy, significance of market failures, and level of development. He concentrates on level of development. His basic theoretical proposition is that labor standards improve with economic development because they are a normal good. The rationale behind this proposition, also put forward by Freeman (1994), is simple: as real incomes rise, people demand more commodities of high quality and fewer of low quality, and commodities produced under poor working conditions are considered of low quality.

Krueger uses the case of child labor to test the theory. His results show that, contrary to what is generally believed, the pattern of political support in the United States for the Child Labor Deterrence Act is not consistent with the hypothesis that support for the bill is a case of disguised protectionism. This finding, Krueger argues, proves that his results are consistent with the hypothesis that labor standards are a normal good. He concludes that rich societies tend to impose stronger labor standards on themselves and are less likely to use child labor. Two additional pieces of evidence are presented to support this hypothesis. One is the negative correlation between countries' child labor participation rate and per capita GDP. The other is the tremendous amount of noncompliance with compulsory schooling laws (an alternative to prohibiting child labor) in developing countries.

Krueger generalizes from the case of child labor to other kinds of labor standards and presents the empirical observation—consistent with theory—of a positive correlation between per capita income and labor standards. In sum, labor standards are

Adolfo Figueroa is professor of economics at the Catholic University of Peru.

Annual World Bank Conference on Development Economics 1996

an endogenous variable to the economic process. A consequence of this theory is that each country must find optimal labor standards in line with their level of development. If developing countries set labor standards too high, they will be widely ignored.

Other Hypotheses for Raising Labor Standards

The logic of Krueger's argument seems impeccable. But there are problems. The observation that labor standards are higher in richer countries, which is the empirical basis of his argument, can also support other theoretical hypotheses.

Even assuming, as Krueger does, that labor standards are set by the market, the demand side cannot be considered the only factor. As I have argued elsewhere, we can also find elements on the supply side that could lead to endogenously determined labor standards (Figueroa 1994). Not only are labor standards higher in richer countries, but so are real wages. The efficiency wage theory suggests that producers will set minimum wages in order to ensure productivity and maximize profits. Efficiency wages should be higher in richer countries; otherwise these countries' high levels of technology, productivity, and profits would not be socially viable and income distribution would not be socially acceptable. The same could be said about work conditions. My point is that labor standards may also be supply-driven, a hypothesis that is also consistent with a positive correlation between labor standards and per capita GDP.

If labor standards are endogenously determined, labor institutions also would be considered endogenous to the economic process. The policy implication of this view is that nothing can be done directly about labor standards in developing countries. Policies that foster economic growth will induce higher standards. Economic progress precedes social progress.

Yet labor codes exist in most countries, independent of their level of income. This is the case for legislation on minimum wages, length of the work period, and so on. How to explain this fact? There are, it seems, some international social norms on labor relations, such as the International Labor Organization's conventions, even if they are few in number and compliance is imperfect.

Theoretically, it is not difficult to show that labor standards can be established exogenously in a particular country. For example, Freeman (1994) argues that "any country that wants higher labor standards for itself can have them . . . if it is willing to pay. A country can pay for standards that increase its cost of production in three ways: through exchange-rate devaluation, with all consumers bearing the burden; through lower wages of workers who gain the benefits; or through taxes on the general public" (p. 87).

Labor Standards and Social Progress

I would like to add another argument. Many analysts, including Freeman, view labor standards only as costs. Their economic benefits are rarely taken into account. In a

society with a high degree of economic and social inequality, as is the case in most developing countries, policies that raise labor standards, however slight or imperfect the improvement, may help to reduce social inequality or discrimination. Greater equality will reduce political instability and allow for the growth of a more modern society. The recent literature suggests strongly that countries with extreme inequality do not attract much private investment (Alesina and Perotti 1993; Figueroa 1995).

Not every type of economic growth will bring about social progress and economic development. Most Latin American countries have experienced periods of rapid economic growth, and yet social progress is lagging. Taiwan, China, applied policies—including land reform—designed to reduce high levels of inequality, resulting in "[marked] improvements in both growth and equity . . . a decade later" (Ranis 1996, p. 51). Dreze and Sen (1991) argue that social progress in industrial countries was the result, among other things, of social policies, not just of pure economic growth.

Labor cost differences between industrial and developing countries are enormous. In 1995 labor costs per hour in the manufacturing sector were $25 in Germany; $16 in France, Japan, and the United States; $5 in the Republic of Korea; $2.40 in Mexico; and about $0.50 in China, India, and Indonesia (*The Economist,* 1 October 1995). Why are the countries with such cheap labor unable to compete more strongly in international markets? The recent literature on international trade shows that trade based on factor endowments alone does not bring development. Bourguignon and Morris (1989) found that developing countries exporting goods that make intensive use of their factor endowments—natural resources and unskilled labor—have higher degrees of inequality. No social progress can result from this type of involvement in international trade.

International trade based on maquila-like systems of production, in which multinational firms try to exploit the abundant labor endowments of a country, would lead to the same result. Kennedy and Cardoza (1995) found that working conditions in maquila firms in Honduras were even below Honduran standards: nine hours of work a day, inadequate space for work, rude treatment by supervisors, and high labor turnover. They also found that employees were mostly young rural women, whose submissive personalities were well suited to the strict discipline required in the maquila work. Again, no social progress can result from this type of involvement in world markets.

To shift to a pattern that is more conducive to social progress, countries must participate in international trade on the basis not only of factor endowments (absolute or comparative advantages) but also of competitive advantage. That comes not from cheap labor but from investments in human capital and technology to increase productivity. Thus competitive advantage depends in part on government policies.

When an entire country competes, as in the case of trade in similar commodities, such as manufactures, productivity may also depend on labor standards. Competitive advantage requires social productivity as well as microeconomic productivity. Under comparative advantage, however, a country can compete in com-

modities that are produced in maquila-style production systems—mining, oil, fishing.

In sum, labor standards may be one of several state policy instruments for building a more productive, modern society. In this view economic progress and social progress go together, rather than one after the other. Moreover, this hypothesis, that exogenously determined labor standards help economic growth, is consistent with the observation that richer countries have higher labor standards.

Krueger's article has contributed to our knowledge of the relationship between labor standards and economic development. However, it does not take the next step of answering the big question of whether labor standards are endogenous or exogenous to the economic process.

References

Alesina, Alberto, and Roberto Perotti. 1993. "Income Distribution, Political Instability, and Investment." NBER Working Paper 4486. National Bureau of Economic Research, Cambridge, Mass.

Bourguignon, Francois, and Christian Morrison. 1989. *External Trade and Income Distribution.* Paris: Organization for Economic Cooperation and Development.

Dreze, Jean, and Amartya Sen. 1991. "Public Actions and Social Security: Foundations and Strategy." In Ehtisham Ahmad, Jean Dreze, John Hills, and Amartya Sen, eds., *Social Security in Developing Countries.* New York: Oxford University Press.

Figueroa, Adolfo. 1994. "Labour Market Theories and Labour Standards." In Werner Sengenberger and Duncan Campbell, eds., *International Labour Standards and Economic Interdependence.* Geneva, Switzerland: International Institute for Labor Studies.

———. 1995. "Peru: Social Policies and Economic Adjustment in the 1980s." In Nora Lustig, ed., *Coping with Austerity: Poverty and Inequality in Latin America.* Washington, D.C.: Brookings Institution.

Freeman, Richard. 1994. "A Hard-Headed Look at Labour Standards." In Werner Sengenberger and Duncan Campbell, eds., *International Labour Standards and Economic Interdependence.* Geneva, Switzerland: International Institute for Labor Studies.

Kennedy, Mirta, and Melizza Cardoza. 1995. "Mujeres en Maquila: El Caso de la ZIP Choloma." Centro de Estudios de la Mujer, San Pedro de Sula, Honduras.

Ranis, Gustav. 1996. "Comment on 'Inequality, Poverty, and Growth: Where Do We Stand?' by Albert Fishlow." In Michael Bruno and Boris Pleskovic, eds., *Annual World Bank Conference on Development Economics 1995.* Washington, D.C.: World Bank.

Comment on "International Labor Standards and Trade," by Alan B. Krueger

Rakesh Mohan

Alan Krueger has addressed a topical and contentious issue that is engaging the attention of governments, businesses, and professional economists. Over the past fifteen years or so most countries have moved toward an increasing acceptance of open trade, which is now generally viewed as enhancing social and economic welfare. In many developing countries the opening of trade has also been influenced by structural adjustment programs adopted under the aegis of the World Bank and the International Monetary Fund. Similarly, most developing countries have lowered their trade barriers as part of the long negotiations that led up to the completion of the Uruguay Round.

Just as open trade has become more acceptable and desirable in developing countries, however, a number of protectionist pressures have emerged in industrial countries. At the root of some of these pressures is high and increasing unemployment in most of Europe and declining or stagnant real wages of unskilled workers in the United States. Some observers have argued that these developments are the result of increasing competition from low-wage countries. As a result there is increasing pressure from industrial countries for the imposition of both minimum environmental standards and labor standards in developing countries. Developing countries fear that if such standards are uniformly applied, they will eliminate their cost advantage in the industries in which they are competitive. Thus developing countries have generally resisted the linkage of such standards with trade.

Krueger's Argument

Krueger's brief theoretical review of the case for and against international labor standards draws no strong conclusion on the effect of labor standards on equity and efficiency. Thus he argues that the case for or against international labor standards can be analyzed only on an empirical basis.

Rakesh Mohan is director general of the National Council of Applied Economic Research in New Delhi, India.

Annual World Bank Conference on Development Economics 1996

Krueger contrasts two opposing views. First is the view associated with free trade advocates such as Jagdish Bhagwati and T. N. Srinivasan, who argue that the linkage between trade and environmental and labor standards arises largely from protectionist motives. The other view is that market failures justify the imposition of international labor and environmental standards. In the case of the environment it is argued that because workers, firms, and consumers may not have adequate information on safety risks and other environmental hazards, an expert agency is needed to impose standards. Similarly, labor markets suffer from market failures arising from imperfect and asymmetric information. There are extreme cases of forced labor or of child labor that lead to market failures in the determination of wages. Other labor market failures may arise from different types of discrimination. Thus it is argued that to prevent a race to the bottom—in which labor exploitation in some countries drives down wages in other countries—the imposition of minimum labor standards would be justified.

The article's main empirical analysis tests whether support for minimum labor standards is motivated by protectionist pressures or by other concerns. Krueger examines support in the U.S. Congress for the proposed Child Labor Deterrence Act of 1995, which is designed to prohibit U.S. imports of goods produced with child labor. Krueger hypothesizes that if demand for this act arises from protectionist impulses then support for it should come mainly from areas that have a higher proportion of industries requiring low skills. The variable used to capture this determinant is the proportion of the population age twenty-five or older with less than a high school degree in each U.S. congressional district. The assumption is that districts with a large share of high school dropouts would have a higher proportion of industries requiring low skills (such as textiles). Krueger finds that representatives of districts with a high concentration of high school dropouts are, in fact, less likely to be cosponsors of the Child Labor Deterrence Act. Another finding is that representatives of districts with a high union participation rate are more likely to be cosponsors of this legislation. He also finds that supporters of international labor standards such as the Child Labor Deterrence Act were more likely to have opposed both NAFTA and GATT. That is, representatives who support international labor standards are more likely to favor protectionist policies generally.

Yet Krueger concludes that there is scant support for the hypothesis that advocacy of international labor standards reflects disguised protectionism. Instead, he theorizes that demand for international child labor standards is a normal good and that people of higher socioeconomic status (higher education) elect representatives who are more supportive of placing limits on child labor.

It is rather surprising that Krueger is willing to come to so strong a conclusion on the general issue of labor standards from one test applied to support for one piece of legislation. First, the main variable used, the proportion of people with less than a high school education, is not necessarily a good proxy for people who feel threatened by cheap imports. As Krueger acknowledges, most low-wage workers in the United States, and presumably those with less than a high school education, are generally employed in the nontraded goods sector, where employment would not be

noticeably affected by a moderate increase in the minimum wage. This suggests that the proxy used by Krueger is inappropriate for the hypothesis being tested. People with less than a high school education are less likely to be employed in the traded goods sector and therefore should not feel threatened by cheap imports from developing countries. Krueger's finding of a positive correlation between unionization and support for labor standards—implying protectionist pressures for labor standards—is more credible.

Second, it is worth considering whether representatives from districts with a large share of people with less than a high school education truly represent these constituents. To illustrate, consider the case of the lowest-income U.S. states—Mississippi, Arkansas, West Virginia, New Mexico, and Oklahoma. Presumably, these states also have a large share of poorly educated people, many of them members of minority groups. It is debatable whether representatives from such districts effectively represent the interests of these constituents. The group of poorly educated people in any congressional district would be a relatively small minority. In a democratic system where congressional representatives are elected on a "first past the post" basis, it is not reasonable to assume that elected representatives would effectively represent all minorities.

India offers another illustration of this point. Surveys find that most Indians are thoroughly dissatisfied with the delivery of services such as banking, telecommunications, and power, which are supplied largely by the public sector. One survey found that as many as 70 percent of Indians would favor privatization in order to improve these services. Yet most members of Parliament, whatever their party, do not favor privatization. This calls into question the appropriateness of Krueger's methodology to the task at hand.

Krueger seems to draw rather strong conclusions from rather shaky evidence. Nonetheless, his empirical exercise is of great interest in that it attempts to explain the voting pattern of legislators on an important economic issue. Clearly, however, much more work is needed to understand the motivation behind such legislation. It would also be interesting to analyze the different pressures on representatives and the influence these pressures have on their votes. To what extent do legislators directly represent the interests of their constituents (and which segments of their constituents), and to what extent do they reflect other interests, including ideological views?

Labor Standards As a Normal Good

Not surprisingly, Krueger finds that rich countries tend to have more stringent labor standards and better working conditions than do poor countries. Similarly, the use of child labor declines as per capita income increases. From these observations, however, is it legitimate to conclude that labor standards, specifically the banning of child labor, are a normal good?

If, as people get richer, they do not like the idea of consuming goods that are produced by child labor, then they should be willing to pay to see these practices abol-

ished. Thus, in theory, consumers would be willing to pay for social funds devoted to the education of children in poor countries, the enforcement of children's rights, and so on. Yet surveys of high-income voters in industrial countries find declining support for foreign aid for such purposes. This waning support is evidenced by the increasing difficulty of replenishing funds for the International Development Association (IDA), the World Bank's concessional lending arm. Bilateral funding for such activities is also declining, particularly from the United States. Thus it is difficult to argue that support for labor standards originates from their status as a normal good rather than from protectionist sentiments.

By contrast, it is easier to argue that environmental standards are a normal good. The increasing concern of industrial countries with environmental standards is evident from the many conventions aimed at preserving the environment, among which linkage with trade is only one. The characterization of environmental standards as a normal good can plausibly be supported by the fact that industrial countries have been willing to provide funds for the Global Environment Facility and other facilities designed to help developing countries upgrade their environmental standards. No such evidence is available for the enforcement of labor standards.

Why the Recent Interest in International Labor Standards?

The basic question is whether we really need labor standards. Every country has legislation governing working conditions in factories and other places of work. India alone has at least fifty acts that prescribe different labor standards relating to wages, safety, working hours, and the like. Thus there is clearly a broad consensus that labor standards are needed to protect workers from exploitation and workplace hazards. The issue is whether there is any justification for imposing uniform labor standards across countries with different income levels and different cultures.

Even among industrial countries labor standards differ. For example, the members of the European Union tend to have higher labor standards than those prevailing in the United States. Within the European Union, the United Kingdom has a different and lower set of labor standards than do the other social democracies of Western Europe. Labor participation in management is actively pursued in a number of Western European countries and is proposed as a standard in the European Union. There is, however, little agreement within OECD countries on whether such standards should be uniform across countries. The United States, for example, clearly would not agree to many of the labor standards that are commonplace in the European Union. Thus there seems to be little justification for the imposition of uniform labor standards across countries that have very different incomes, cultures, and political and social traditions.

Those in favor of linking labor standards and trade respond that uniformity is needed only for a few standards that are almost universally regarded as basic. These include freedom of association, the right of collective bargaining, prohibition of forced labor and of child labor, and nondiscrimination. Most countries already ban forced labor and child labor. The problem is with enforcement. The issue, then, is

whether enforcement should become international, by linking standards to trade and imposing trade penalties for noncompliance.

An analogous situation concerns the linking of trade and illicit drug usage in industrial countries like the United States. Because of the inability of U.S. and other Western governments to enforce their drug laws, much illegal activity arises in some Latin American and Asian countries in the production of these drugs, thereby doing great harm to their societies and economies. The illicit trade in drugs generates large amounts of unaccounted for money, adversely affecting these countries' exchange rates and trade volumes, as well as other aspects of their economies. Should there be trade sanctions against the United States because of its inability to curb illegal drug use, since this practice has deleterious effects on international trade as a whole?

Another question that arises is why now? Why has concern with minimum labor standards surfaced in industrial countries now? As noted earlier, free trade has become increasingly accepted around the world, and completion of the Uruguay Round and other trade measures is opening industrial countries to a large volume of consumer goods made with low-wage labor in developing countries. In particular, only in recent years has the Multifiber Arrangement been amended substantially, and over the next ten years it is to be phased out entirely. Increasing unemployment and stagnating wages for low-skill workers have occurred coincidentally with this expansion of imports. Thus it is not surprising that labor organizations, governments, and even consumers in industrial countries view the expansion of such imports as a threat to employment opportunities. The paradox, however, is that most studies—including those sponsored by OECD countries—indicate that increasing imports of developing country goods are not responsible for high unemployment in Europe. Furthermore, since most low-wage employment in industrial countries is in the nontraded sectors, workers in the traded sectors should not feel threatened by imports from developing countries.

Conclusion

These observations suggest the need for a different, more realistic approach to the enforcement of labor standards in developing countries. Essential to this approach will be faster growth of these countries' income levels so that the ills of child labor and forced labor become a thing of the past. The most effective way to do this would be to expand these countries' exports rather than to erect trade barriers against them. What is required is empirical research to convince concerned interests in industrial countries that low-wage goods do not threaten workers in those countries. Only with such evidence and credible communication campaigns will unions in industrial countries be persuaded to drop their opposition to the import of these goods and their support for such protectionist measures as the linkage of trade with labor standards. This objective is not served by the justification of such pressures through the characterization of labor standards as a normal good.

Floor Discussion of "International Labor Standards and Trade," by Alan B. Krueger

In discussing endogeneity, Adolfo Figueroa (discussant) was more at odds with trade economists than he himself was, said Krueger. As Krueger understood it, Figueroa had said that labor standards raise incomes; thus a country's income rises as the percentage of children working falls. Richard Freeman (1993, 1994), on the other hand, has argued that countries can have the standards they want, but that they have to pay for them in the form of lower GDP. Freeman's view is probably the view that most economists have. Krueger said that he sympathizes with Figueroa's view and had tried to cover it by discussing how standards might improve welfare when there were market failures. But he did not agree that standards were necessarily endogenously set to raise income.

Krueger disputed Rakesh Mohan's (discussant) criticism of the regressions dealing with protectionism, which he considered the heart of his article. Krueger finds it useful to evaluate empirical work by asking, before he sees the results, what he expects to find. And when he started writing his article, he expected to find the opposite of what he found. He expected to find that high school dropouts are competing with cheap imports produced by child labor, so their congressional representatives would be more likely to support restrictions on imported goods made with child labor. But the results were contrary to this prediction. Although the share-of-high-school-dropouts variable would not have mattered in the NAFTA regressions if the constituents did not respond to trade issues, it did matter in a protectionist way in these estimates. That asymmetry is important, Krueger thought.

Krueger agreed that there might be more efficient ways to buy the moral satisfaction of knowing that you are helping poor people abroad, as Mohan pointed out—for example, by contributing to funds for children's education abroad. And a lot of that goes on, both privately and through governments. But if raising labor standards is the most efficient way of achieving this goal, then maybe it is the best approach. Krueger did not know whether that was the case, which is why he raised the ques-

This session was chaired by Michael Walton, chief economist, East Asia and Pacific Regional Office, at the World Bank.

Annual World Bank Conference on Development Economics 1996

tion in his article. Both discussants seemed to agree that standards are necessary; they just had doubts about linking them to trade. Krueger also said he believes that it is important to distinguish core standards—such as the use of forced labor or discrimination—from others.

Another point he had emphasized, Krueger said, was enforcement. The NAFTA side agreements do not work by imposing standards that do not exist; rather, they require the enforcement of standards that already exist. Richard Cooter and Avner Greif (presenters in other sessions) had made relevant points along those lines: it does no good to have a set of standards that everyone disobeys; in fact, that could be corrosive, leading to rampant noncompliance.

A visiting scholar from the University of Maryland asked the panelists to assume that most trade economists agree with the hypothesis that countries with higher incomes have a preference for higher labor standards and to suppose that with higher incomes such preferences apply not just to the high-income country but to other countries as well. Even so, the General Agreement on Tariffs and Trade and the World Trade Organization already have a mechanism in place to make those preferences known—safeguards. If the United States does not like a certain labor standard in India, it can take a trade policy action against India. But then it either has to give a compensatory concession or let India take a reciprocal action against U.S. exports. However, the United States might not be in the best position to criticize other countries about child labor. In the United States children often earn money babysitting, and many U.S. newspapers are delivered by eight- or nine-year olds. So in some ways even the United States could be considered noncompliant on child labor law.

A participant from the World Bank's International Trade Division added that the United States is not successful in enforcing a lot of its own standards—for example, among immigrant laborers working in sweatshops in California. From a practical point of view, how can industrial countries get developing countries to enforce standards that they do not want? Krueger agreed, saying that if standards exceeded what a society believed appropriate, they would be ignored. As for enforcement of such laws in the United States, the best way to ensure compliance is not necessarily to have more inspection. Faced with budget cuts for inspectors, the U.S. Department of Labor is starting to publicize the names of companies that use sweatshop labor. This approach seems to have been effective, because demand is fairly elastic and it does not take much to get people to switch to different stores.

Noncompliance is unlikely to be completely eliminated, said Krueger. Minimum wage regulations, for example, generate both compliance and noncompliance. In the United States the penalty for employers caught not paying the minimum wage is to pay employees the back pay they should have received. But most employers who use low-wage workers pay the minimum wage. For the most part compliance appears to be voluntary. The same thing happens with compulsory schooling laws in the United Kingdom. The remedy for noncompliance is not to set standards too high. Rather, standards should be set at a level that the public agrees with (that is, they should be

set endogenously), and what the public thinks is right depends on where the economy is and what people think they can afford.

The participant from the World Bank then asked if Krueger's regressions reflected the right concerns. If people in the United States are truly concerned about child labor abroad, why had this concern emerged only recently, with the regional and global integration of economies? Why had it not surfaced twenty years earlier? Rakesh Mohan also wondered about the belated concern. After forty years of protectionism in textiles and garments, the industrial countries had finally agreed through the Uruguay Round to phase out the Multifiber Arrangement. It seemed to be more than coincidence that as soon as that agreement was reached the West suddenly became concerned about child labor laws in developing countries that export textiles and garments.

Concern about international labor standards is nothing new, Krueger responded. The International Labor Organization, for example, was founded in 1919. And it is difficult to sort out the issues on something like the Child Labor Deterrence Act because such acts typically are attached to legislation about other issues. Thus Krueger thought it was wrong to say that concern about labor conditions had surfaced only now that the world had become more integrated and protections have been stripped away.

A participant from American University, noting that the last two times he had been in a large, growing economy in Asia he had been horrified by the general working conditions, asked to what extent industrial countries should get involved in other countries' labor markets. He was bothered not by conditions for forced labor in that country but by conditions for regular workers—conditions that would not be allowed in most Western countries. In trying to improve labor conditions, should industrial countries be concerned only with tradable goods, which they can do something about? Should labor standards for the production of nontradables not be a concern? This issue affects not just labor standards, but also environmental and other standards.

Mohan agreed that the conditions in which children work in some developing countries are abominable and that there is absolutely no justification for it. And in most countries—certainly in India—there is a ban on child labor. These countries are not promoting child labor. It is a question of compliance and of necessity. Similarly, there had been a good deal of bonded labor in India for a long time, but voluntary groups in India and other countries had, almost as vigilantes, sought out bonded labor and forced child labor and tried to end it.

Krueger thought bilateral and multilateral negotiations could profitably include labor standards. Standards may be a way to fairly divide the surplus created by such agreements. Mohan thought that the issue is more complex than that. When power is not equally shared between different signatories to multilateral agreements, he said, it is questionable how voluntary the final agreement is. Still, Mohan prefers multilateral to bilateral agreements.

Michael Walton (chair) said that some members of the international financial community were interested in putting pressure—in the tradition of ethical investment—on countries and companies with weak standards. With the growth of insti-

tutional investors, there was increasing pressure from labor unions and the like to invest their pension funds in a way that influences the behavior of companies and countries. They are more inclined to support companies like Levi Strauss, which has consistently pursued good policies on child labor, and to avoid companies that do not have ethical policies. Krueger agreed that institutional investors could apply pressure; that was an extension of a point he had made about consumers caring about the way products are made. There are some questions about how much pressure can be applied under current laws, which require fiduciary responsibility for pension funds. But some pressure may be preferable to none. After all, in the early and mid-1800s European countries had not refused to trade with the United States because of slavery. Maybe if they had, the U.S. Civil War could have been avoided.

A number of speakers had argued that labor standards are likely to vary according to a country's stage of development, said a participant from Concordia University in Montreal. The same is true of environmental standards, and yet the international community did not seem to expect them to be the same in all countries. Why should it expect conformity in labor standards?

Krueger said that his article was not advocating uniform international labor standards. On a personal level, his view was that the international community should try to get the International Labor Organization to more vigorously enforce certain core standards: allowing freedom of association and collective bargaining, for example, and prohibiting forced labor and discrimination. Surely doing away with forced labor, in addition to being morally right, would improve efficiency and welfare. He did not view his article as a brief for having a global minimum wage, either, nor did he believe that there should be one. Rather, he believed that a minimum wage is a normal good and that when a country reaches a certain income it starts to think that it is desirable to redistribute money by having or raising a minimum wage.

References

Freeman, Richard. 1993. "Labor Market Institutions and Policies: Help or Hindrance to Economic Development?" In Lawrence H. Summers and Shekhar Shah, eds., *Proceedings of the World Bank Annual Conference on Development Economics 1992*. Washington, D.C.: World Bank.

———. 1994. "A Hard-Headed Look at Labour Standards." In Werner Sengenberger and Duncan Campbell, eds., *International Labor Standards and Economic Interdependence*. Geneva, Switzerland: International Institute for Labor Studies.

Environmental Standards and International Trade

Kym Anderson

Concerns about natural resource use and the environment on one side and about the trade effects of environmental policies on the other are becoming increasingly prominent in trade and trade policy discussions, including those involving the recently formed World Trade Organization. Many developing countries perceive the entwining of environmental and trade issues as a threat to their sovereignty and their economies, while influential groups in industrial countries consider it unfair, ecologically unsound, and even immoral to trade with and invest in countries adopting much lower environmental standards than theirs. This article examines why these issues are becoming more prominent, whether the World Trade Organization is an appropriate forum in which to discuss them, how they affect developing countries, and what those countries should do about it. It concludes that the direct effect on developing economies is likely to be small and may even be positive for some through improved terms of trade or compensatory transfer payments. But great care is needed to avoid inappropriate uses of trade measures to pursue environmental objectives. Otherwise there is considerable risk of an adverse indirect effect on developing and other economies through the erosion of the rules-based multilateral trading system.

During the past decade or so international economic integration has accelerated as a result of economic policy reforms and the communications revolution. Integration has brought with it greater scrutiny of domestic policies—including environmental policies—that affect the competitiveness of industries in the global marketplace. At the same time concerns about resource depletion and environmental degradation have been growing at the regional, national, and global levels, leading to calls to slow resource exploitation and enforce stricter environmental standards—including at the international level. Together, these developments have caused an entwining of policies relating to trade, foreign investment, and the environment. That entwining has the potential to bring about good outcomes for the economy and the

Kym Anderson is professor of economics at the University of Adelaide. The author is grateful to Maureen Cropper, Robert Hudec, and Will Martin for helpful comments on earlier drafts.

Annual World Bank Conference on Development Economics 1996

natural environment, but unless it is carefully managed there is a considerable risk that both the economy and the environment will suffer.

Why are economic development and the deepening of economic integration raising the demand for international cooperation or coercion on policies affecting natural resource use and the environment? Under what circumstances (if any) is trade policy an appropriate instrument for addressing such concerns? What does this imply for the global trading system? How will developing countries be affected, and what should they do about this phenomenon?

Why Are Environmental Policies Coming under Closer International Scrutiny?

It should not be surprising that there are vast differences across countries in environmental policies, just as there are vast differences in policies affecting such issues as worker rights and standards, human rights more generally, education, health, and national culture. These policy differences are a natural consequence of differences in national incomes, endowments, urbanization, location, preferences, available technologies, and information. Differences in per capita income matter because as countries become richer, they increase their demand for normal goods, including environmental standards. Differences in per capita endowments of natural resources and environmental amenities matter because, other things being equal, they tend to be of greatest concern in countries where they are scarcest. Environmental policies also differ because of international differences in tastes and preferences. Indeed, one of the defining historical features of many countries was the bringing together of a group of people whose preferences were more similar to each others' than to those of neighboring groups (Alesina and Spolaore 1995).

As international economic integration proceeds, though, pressure increases to reduce differences in domestic policies that have significant consequences for trade. This pressure is driven not just by the desire to reduce administrative and conformance costs but also by concerns in regions with high environmental standards that higher production costs for their firms and industries are making them less competitive than firms and industries in regions with lower standards. Such differences become even more important as traditional barriers to trade and investment between regions fall (Bhagwati 1996). Harmonization of social policies and standards can go either way, however, with winners and losers in each region trying to influence the outcome. There is no reason to presume that national or global economic and social welfare will improve simply because such policies are harmonized.

At the same time the list of environmental concerns with international or global dimensions has grown rapidly in recent years. Some groups in rich countries (where environmental standards are high) are concerned that these problems will be exacerbated as economic growth proceeds in newly industrialized countries (where environmental standards are more lax). Since personal values strongly influence debate on these issues, there is the potential for considerable friction between countries with different preferences, resource endowments, available technologies, incomes,

and knowledge about how different activities and policies affect the environment—and thus with different perceptions about optimal levels of national and global environmental and resource policy intervention.

Although concerns about resource depletion and the environment might fluctuate with the business cycle, they are likely to keep growing. One reason is that, even though uncertainties remain, the scientific basis for many of these concerns is more solid now than it was twenty years ago. Another reason is that the world's population and its real per capita income continue to grow rapidly by historical standards, while supplies of most natural resources and environmental services are limited and markets for many of them are incomplete or absent.

Most industrial economies have established institutional structures to help reach social consensus on policies for environmental and sustainable development, on the allocation of property rights, and on enforcement. The same is true of some traditional societies before they begin to modernize and their resources come under pressure. But these institutions are less common in newly industrializing economies, where growth in the world's population and consumption is expected to be concentrated during the next few decades. And at the multilateral level cooperative intergovernmental mechanisms for environmental policy have only recently begun to be formed and will take time to become effective, especially where free-rider problems are rife.

So with adequate forums for multilateral environmental dialogue not yet fully developed and with an increasing sense of urgency about environmental problems, environmental groups—especially in industrial countries—have become interested in using trade restrictions, the one policy instrument available to their governments to influence environmental outcomes.

Environmental groups see trade policy as useful in two respects: as a way of raising environmental standards at home and abroad and of inducing countries to become signatories to and abide by international environmental agreements. Imposing import restrictions on products from countries with lower environmental standards can reduce opposition by local firms to higher standards at home by offsetting the loss of competitiveness and can increase the incentive for foreign firms and their governments to adopt higher standards.

Another major attraction is that trade measures are easy to use as sticks or carrots and fairly immediate in their impact. Even the threat of trade sanctions (particularly if it is broadened to include unrelated products) can have a rapid effect on raising national standards or encouraging a country to sign an international environmental agreement and abide by its rules. But while these features make trade policy attractive to environmentalists, such discriminatory and protectionist uses of trade policy are unattractive to supporters of liberal world trade.

How Do Environmental Policies Fit into Standard Theories of Comparative Advantage?

Although the standard theory of changing comparative advantages in a growing world economy was developed without consideration of environmental concerns,

it can readily be modified to incorporate at least some of those concerns. As espoused by Krueger (1977) and Leamer (1987), this theory suggests that when a developing country opens up to international trade, its exports initially will consist mainly of primary products since its stocks of produced capital relative to natural resources are comparatively low. If non-natural capital stocks per worker (including human skills) expand more rapidly for the country than for the rest of the world, the country's comparative advantage will shift to more capital- and skill-intensive activities. For a country that is relatively land abundant, some of that produced capital and new capital-intensive technology may be used to extract minerals or farm the land. But in most cases the new capital will encourage the expansion of nonprimary sectors and shift comparative advantage away from primary products.

Thus countries that are lacking in natural resources or that are densely populated will tend to industrialize at an earlier stage of economic development than countries that are resource rich or people poor. Moreover, their nonprimary exports initially will tend to be more intensive in the use of unskilled labor. In manufactures the gradual upgrading to more capital-intensive production leaves room in international markets for later-industrializing, resource-poor countries to also begin with labor-intensive, export-oriented manufacturing.

National Environmental Policies and Comparative Advantage

If there were no international environmental spillovers and no global commons, these determinants of comparative advantage would need to be complicated only slightly to incorporate nonmarketed environmental services and pollution by-products. This complication is required to accommodate the fact that as a country's per capita income and industrial output grow, the value its citizens place on the environment increases. As this value increases, so too do demands for proper valuation of resource depletion and environmental degradation, for better policing of property rights, and for implementation of domestic pollution abatement policies—at least after certain thresholds of income or pollution have been reached.[1] Beyond those thresholds the intensity of such abatement policies is likely to be positively correlated with per capita income, population density, and urbanization.

If all economies were growing at the same pace but from different bases, the progressive introduction of national environmental taxes and regulations would tend to cause pollution-intensive production to gradually relocate from wealthier or more densely populated countries to developing or more sparsely populated countries. Those policies also would slow or reverse the growth in demand for products whose consumption is pollutive, especially in wealthier or more densely populated countries, where taxes on such products would tend to be high. If more-advanced economies are net importers of products whose production is pollutive—as is likely—or net exporters of products whose consumption is pollutive the imposition of optimal environmental policies would worsen these countries' terms of trade to the benefit of poorer economies (Siebert and others 1980; Anderson 1992b). Thus

as standards rise in advanced economies, developing countries benefit from an improvement in their terms of trade.

Through foreign trade and investment even countries without (or with unchanged) environmental policies are affected by the changes in environmental policies that accompany growth in other countries.[2] That is, one country's environmental policy choices are not independent of those of other countries. The imposition of higher standards or pollution charges makes domestic pollution-intensive industries less internationally competitive. Unless they had been developing new, environmentally friendlier technologies, these industries would tend to lobby against higher standards, particularly if their foreign competitors were not being subjected to similar cost-raising policies. And while producers in the less-polluting industries at home would benefit from the raising of a particular environmental standard, such industries are more diffuse and so are not likely to add much support to the lobbying efforts of environmentalists.

It was because of this flagging support that trade policy entered the environmental picture during the late 1960s, when the first wave of widespread concern for the environment began in industrial countries. Environmental groups saw the potential for offsetting the reduced competitiveness of pollution-intensive industries by restricting imports from lower-standard countries. Thus such industries would be less opposed to higher standards at home, and foreign firms and their governments would have an incentive to adopt higher standards to avoid being labeled a pollution haven and subjected to anti–"eco dumping" duties.

Demand for unilateral use of trade policy in this manner has grown in two ways with the internationalization of the global economy. First, the decline in traditional trade barriers (tariffs, transport and communications costs, and so on) has made any given environmental charge relatively more important as a determinant of international competitiveness. Second, the deregulation of financial markets and foreign direct investment during the 1980s made it much easier for firms to pull their investments out of countries with high environmental standards and relocate their factories to countries with lower standards. Environmental groups fear that governments seeking to attract or retain investments and jobs will delay introducing or enforcing stricter environmental policies and may even compete in lowering standards in a "race to the bottom." Environmental groups also worry that greater relocation opportunities will reduce the incentive for firms to develop more environmentally friendly technologies.

The extent of international relocation of productive activities due to the raising and enforcing of environmental standards should not be exaggerated, however. Recent studies suggest that the effect of such policies on comparative costs may be quite small (Leonard 1988; Low 1992; and Jaffe and others 1995). Moreover, multinational corporations tend to use cleaner technologies than local firms in developing countries (Harrison and Eskeland 1996). Japanese multinational corporations, for example, are required to obey the same standards abroad as they do in Japan. And there is little evidence of a change in patterns of trade specialization in response to the imposition of environmental regulations since the 1960s (Tobey 1990). Still, the

absence of observed changes in trade patterns may be the result of import barriers that were raised to offset any decline in the competitiveness of affected industries (Hoekman and Leidy 1992). Technological changes induced by the raising of environmental standards and environmental standards raised by the introduction of technological changes also would reduce the likelihood of a correlation between the raising of standards and the international relocation of production.

International Environmental Spillovers and Comparative Advantage

The story becomes more complicated when international environmental spillovers are taken into account. These spillovers may be both physical and—for want of a better term—psychological. An example of a psychological spillover is that I may grieve if another country's activities threaten a particular animal or plant species in its jurisdiction. Or I may grieve if I believe that the desires of another country's citizens for higher environmental standards at home are not adequately recognized by their national government (a political market failure). Controversial though such views are, many people see a need for multilateral action to reduce these spillovers—and that is where trade policy again enters the debate.[3] Environmentalists view trade measures as powerful carrots and sticks for attracting signatories to multilateral environmental agreements, for penalizing nonsignatories, and for encouraging other countries to adopt better national environmental policies.

As with using trade policy to achieve national environmental policies, the use of trade policy to increase the acceptance of multilateral environmental agreements raises potential conflicts of interest between rich and poorer economies. There is even dispute over what constitutes the global commons: some would argue that countries or regions should not have to bow to international pressure to preserve endangered species in their territory (or at least not without adequate compensation), while others would argue that such countries are merely the custodians of those resources for the benefit of humankind generally.

There are many reasons people in developing countries are upset by the use of trade measures to achieve international environmental objectives. Consider the extreme example of the ban on ivory trade under the Convention on International Trade in Endangered Species. When that ban was introduced in 1989, the strong comparative advantage that southern African nations had had in elephant products virtually disappeared. Another example is the recent ban, adopted under the Basel Convention relating to hazardous waste, on exports of so-called hazardous recyclables from industrial to developing countries. That ban threatens the growth prospects for recycling industries in developing countries. A third example is the proposed limitation on imports of tropical hardwoods into high-income countries, in order to discourage deforestation. An import ban of this kind would reduce the growth in exports of logs and perhaps sawn timber from developing countries that are still well endowed with hardwood forests while improving the terms of trade of net importers of hardwood such as Japan, the Republic of Korea, and Taiwan (China). The Montreal Protocol on phasing out the use of ozone-depleting chloro-

fluorocarbons (CFCs) incorporates discriminatory trade provisions designed to limit the relocation from signatory to nonsignatory countries of industries producing or using CFCs, as well as to encourage nonsignatories to accede to the protocol. Finally, consider the infamous example of the U.S. ban on the importation of Mexican tuna that U.S. authorities deem to have been caught in dolphin-unfriendly nets. If implemented alone, the U.S. regulation on the use of dolphin-unfriendly nets on U.S. registered fishing vessels would have boosted Mexican competitiveness in tuna fishing; the subsequent ban on tuna imports reduced it instead.

Should Trade Policy Be Used to Achieve Environmental Objectives?

As is clear from these examples, trade policy measures can be motivated by a mixture of concerns: for national competitiveness and for the global commons and animal welfare. The second set of concerns is typically felt more strongly by people in industrial countries than by people in developing countries. This disparity stems not just from the lower incomes in developing countries, which mean that people are less able to afford high standards, but also from the fear that such trade interventions threaten their export earnings.

The increasing use of discriminatory trade measures to address environmental issues has led to calls from Western Europe and elsewhere to amend Article XX (the exceptions clause) of the General Agreement on Tariffs and Trade (GATT) to allow the use of trade barriers for environmental protection. Apart from the adverse effect this move might have on the export earnings of developing countries, this change should concern all countries—but especially developing countries—for at least three other reasons. First, trade policy measures usually are not the best instruments for achieving environmental objectives because trade sanctions (or the threat of sanctions) do not directly address the root cause of the environmental problem. Their use in place of more efficient instruments increases the use of global resources[4] and unnecessarily reduces the level and growth of global economic welfare as conventionally measured—and may even add to rather than reduce global environmental degradation.[5]

The second reason for concern is that producer interest groups and environmental groups are finding it mutually advantageous to use environmental arguments to support their claims for unilateral import restrictions, particularly following the imposition of costly environmental standards on domestic producers (Hillman and Ursprung 1992; Hoekman and Leidy 1992). Environmental concerns can provide a convenient—and socially respectable—excuse for raising trade barriers (Steil 1994). But such protectionist actions reduce real incomes both at home and abroad, especially in developing and natural resource–abundant countries.

Third, insofar as this activity can lead to an escalation in trade disputes—as seems inevitable, given the divide between developing and industrial countries and the fact that environmental uses of trade policy are inherently discriminatory—they could be followed by retaliatory and counter-retaliatory action. Such actions would ultimately undermine the global trading system on which the dynamism of developing economies depends.

Does Liberalizing Trade and Investment Harm the Environment?

Another important way in which elements of the environmental movement are threatening the global trading system is through opposition to liberalization of trade and investment. Before considering the reasons for this opposition, it is useful to review what economic theory has to say on the matter.

The standard theory of distortions and welfare tells us that if there is only one distortion in the global economy (for example, trade restrictions), then reducing that distortion will be welfare enhancing for the world as a whole. Not all groups need gain, but the gain to those who benefit is more than enough to compensate those who lose. But theory also tells us that if the world has more than one distortion (for example, environmental externalities that have not been addressed by optimal policies from a global viewpoint), then reducing only one distortion (trade restrictions) may not increase global welfare. For example, liberalizing trade between industrial and developing countries could lead to excessive felling of tropical forests if there is inadequate protection of forest property rights in developing countries (Chichilnisky 1994). To achieve unequivocal global welfare improvement in such cases, all other distortions must be reduced while trade is being liberalized. Even then, some countries may be made worse off and the environment may still be harmed (Copland and Taylor 1995).[6]

Thus it is not possible to claim with certainty that liberalizing trade will improve the environment and increase welfare when there are significant environmental externalities that, because of some political failure, have not been addressed with appropriate environmental policies. Even so, the first-best action is to overcome that political market failure so that trade reform *can* help boost welfare (Bhagwati and Srinivasan 1996).

With that in mind, it is possible to examine why environmental groups oppose trade and investment liberalization. They oppose the GATT, the World Trade Organization (WTO), and regional attempts to reduce barriers for at least three reasons: freer trade means more output and income, which they presume means more resource depletion and degradation of the natural environment; freer trade and investment increase transport activity and encourage the relocation of environmentally degrading industries to countries with lower environmental standards or more fragile natural environments, contributing to further environmental damage; and freer foreign investment reduces the incentive to develop environmentally friendlier technologies.

None of these assertions is unambiguously supported by empirical evidence. The first, that higher incomes mean greater damage to the natural environment, may be true initially for some developing countries (in which case any additional environmental damage has to be weighed against the marginal economic benefits of higher incomes for poor people). But once middle-income status is reached, people tend to alter their behavior in ways that reduce pressures on the environment. A key change is in family size: higher incomes eventually lead to lower population growth. This change would likely have a major effect on reducing the rate of environmental

degradation caused by population pressures in developing countries. In rural areas it means fewer people felling trees and denuding hillsides to eke out a subsistence income; in urban areas it means fewer unemployed or underemployed squatters in shanty towns with poor water and sanitation services.

Another common behavioral change as economies open up and incomes rise is that the demand for education expands. With more income and education come more skillful management of all resources, including the environment, and more forceful demands on governments to improve private property rights and environmental policies. In addition, the political cost of such policy reforms is reduced because businesses have more opportunities to meet stricter standards by acquiring more and cheaper environmentally benign production processes and products from abroad. Thus one might expect that as trade and investment liberalization lead to upward convergence in incomes around the world, there would be an upward harmonization of environmental standards (Casella 1995). That realization points to the inappropriateness of the blanket call by some environmental groups for upward harmonization of standards to precede trade liberalization, since liberalization may in fact induce harmonization.

As to the assertion by some groups that the global environment is necessarily harmed by the relocation of production following trade and investment liberalization, only empirical studies can say. We know from the law of comparative advantage that not all industries will relocate from rich to poor countries when trade barriers in rich countries are lowered. Some industries in industrial countries will expand at the expense of industries in developing countries, and some industries in developing countries will expand at the expense of industries in industrial countries. In any case it should not automatically be assumed that relocating production to developing countries necessarily harms the environment.

Recent examinations of the likely environmental effects of reducing government assistance to two of the most protected industries in industrial countries—coal and food—reveal that in both cases the global environment may benefit from trade liberalization, especially if complementary environmental policies are in place. That outcome is possible partly because production of those goods tends to be more pollutive in industrial countries than elsewhere. Moreover, reducing coal producer subsidies in Europe would raise the international price of coal, thereby discouraging its use and so lowering global carbon emissions (Anderson 1992a; Steenblik and Coroyannakis 1995). What is not yet known is whether broadly based liberalizations such as the Uruguay Round are more environmentally friendly than liberalization in individual product markets. Even if quantitative estimates of the key environmental effects were available, a formidable task would remain in valuing them and comparing the net value with the conventionally measured economic welfare gain from trade liberalization.

An increase in transport activity resulting from trade reform also need not increase the risk of environmental damage. The lowering of import barriers to processed primary products, for example, would allow more raw materials to be processed in resource-rich countries, reducing the bulkiness of shipments. And if

there are negative externalities associated with shipping (such as the risk of oil spills), it would be more efficient to ensure that shippers pay more of the full cost of their activity—say, through an international agreement requiring a minimum standard of double hulls on oil tankers—than to reduce trade generally.

Finally, there is the assertion that the opportunity for capital outflow offered by liberalization breeds pollution havens abroad and thereby reduces the development of environmentally friendly production technologies in countries with higher environmental standards. Some observers have argued that the potential benefits of such innovations are so great that raising environmental standards could boost rather than retard a country's economy (Porter and van der Linde 1996). But that argument begs the question of why such investments would not have been made in an open economy in the absence of stricter standards (Palmer, Oates, and Portney 1996). In any case there is little empirical evidence to suggest that raising standards stimulates innovation, just as there is little theoretical or empirical support for the notion that raising standards has a significant impact on the competitiveness of firms in industrial countries or on their decisions to invest in developing countries (Jaffe and others 1995; Wilson 1996).[7]

The GATT, the World Trade Organization, and the Environment

How "green" are the GATT's rules, how have they been adapted over time, and should they be altered further? From the outset the GATT has been a conservationist institution in the sense that its purpose is to reduce trade barriers and thereby inefficiency in the use of the world's resources. The heart of the GATT, agreed to by twenty-three original contracting parties in 1947 and since then by another hundred or so countries, is the nondiscrimination requirements of Articles I and III. These articles obligate parties to treat imports from any GATT contracting party no less favorably than other imports (the most favored nation requirement) and no less favorably, after border taxes are paid, than similar domestic products (the national treatment requirement).

Article XX provides exceptions to these general rules, however, including provisions for some environmental regulations. Specifically, parts (b) and (g) of Article XX allow trade restrictions "necessary to protect human, animal, or plant life or health" and "relating to the conservation of exhaustible natural resources if such measures are made effective in conjunction with restrictions on domestic production or consumption," subject to the requirement that such restrictions "are not applied in a manner which would constitute a means of arbitrary or unjustifiable discrimination between countries where the same conditions prevail, or a disguised restriction on international trade." This article has been interpreted to mean that the measure must be *primarily* for a conservation purpose (rather than for a mixture of motives) and must be *necessary* in the sense of being the *least GATT-inconsistent* measure available.

These provisos have ensured that the article has been rather narrowly interpreted, which is one reason some environmental groups believe that the GATT needs fur-

ther greening (Charnovitz 1991; Esty 1994). But there is nothing in the GATT that prevents a country from adopting the most efficient measures to offset environmental externalities, which typically are associated with production, consumption, or disposal activities. Since trade itself is almost never claimed to be the root cause of environmental problems, GATT supporters traditionally have seen little reason to consider trade measures as part of the solution to environmental problems.

When widespread public interest in trade and environmental issues surfaced in industrial countries in the late 1960s and early 1970s, industrial pollution within and between neighboring advanced economies was the main concern. The foreign trade and investment issues raised at that time focused on how the imposition of strict pollution standards might damage the international competitiveness of domestic firms and on how to avoid such damage through border protection measures.

Where the environmental damage caused by production is purely local, there is no economic logic to calls by firms for trade restrictions or subsidies to offset the decline in their international competitiveness caused by new environmental standards: such assistance would tend to offset the desired effect of limiting by-product pollution.[8] Nor is it reasonable to conclude that countries with laxer environmental standards are engaging in eco-dumping if their standards are consistent with their preferences and natural resource endowments (for example, because the countries are poorer, less densely populated, or less urbanized). Even so, claims for protection against eco-dumping have political appeal and may have pushed import barriers or export subsidies in advanced economies higher than they would otherwise have been.

As part of its preparations for the United Nations Conference on the Human Environment, held in Stockholm in June 1972, the GATT Secretariat established a Working Group on Environmental Measures and International Trade and produced a background paper on the issue (GATT 1971). But there were no significant changes to the GATT during the Tokyo Round as a result of these concerns, and it took two decades for the working group just to meet for the first time.

Trade policy actions are more likely to occur—and to be more difficult to dismiss as inappropriate—when environmentalists view particular kinds of environmental damage as unacceptable regardless of the country in which it occurs. This case is even more problematic if the damage is not just psychological (as with animal rights) but also physical, because then the relocation of production to a country with laxer environmental standards may worsen animal welfare, or the environment at home, in addition to reducing the profitability of domestic firms. The U.S.-Mexico dispute over the use of dolphin-unfriendly nets for tuna fishing again comes to mind. The GATT dispute panel ruled against the U.S. ban on imports of tuna from Mexico, partly because the ban did not discriminate according to the type of net used. That kind of discrimination is difficult to achieve efficiently because what is considered objectionable is an aspect of the production process rather than the final traded product itself.

Had the GATT panel ruled in favor of the tuna import ban by the United States, it would have set a major precedent. It would have opened a potentially huge loop-

hole in the GATT for any country to apply trade restrictions unilaterally as a means of imposing its environmental standards on other countries. Such a loophole would work against the main objective of the multilateral trading system, which is to provide stable and predictable nondiscriminatory market access opportunities through agreed rules and disciplines and bound tariffs on imports. This is yet another reason to resist calls to amend Article XX of the GATT to include environmental protection as an acceptable exception to the nondiscrimination principles of Articles I and III.

Environmental Provisions under the Uruguay Round

The current wave of public concern for the environment—given international expression at the United Nations Conference on Environment and Development held in Rio de Janeiro, Brazil in June 1992—is much more intense, widespread, and likely to be sustained and to affect a broad range of countries and products than was the case until the late 1980s. The agenda for the Uruguay Round had already been set before the current wave of concern had built up, so the trade-environment issue was not a separate item for negotiation. Nor was an environmental impact assessment conducted for the Round as a whole. But the GATT's Working Group on Environmental Measures and International Trade, formed in 1971, finally met for the first time in 1991. It has met frequently since then. In addition, several Uruguay Round agreements contain provisions that relate to the environment and build on articles in the GATT.

The most fundamental environmental provision in the Uruguay Round is in the preamble to the agreement establishing the World Trade Organization. It defines the WTO's objective as providing contracting parties with maximum opportunities for:

> Expanding the production and trade in goods and services, while allowing for the optimal use of the world's resources in accordance with the objective of sustainable development, seeking both to protect and preserve the environment and enhance the means for doing so in a manner consistent with their respective needs and concerns at different levels of economic development.

To help promote that objective, ministers meeting in Marrakesh in April 1994 to sign the final act of the Uruguay Round agreed to establish a Committee on Trade and Environment to report to their first biennial meeting (in December 1996 in Singapore). Other Uruguay Round agreements with environmental provisions relate to technical barriers to trade, sanitary and phytosanitary measures, subsidies and countervailing duties, and trade-related intellectual property rights (see Anderson 1995 for details). Overall, the trade liberalization resulting from the Uruguay Round will almost certainly conserve global resources by allowing countries to specialize in producing more of what they do best. Whether it will also reduce environmental degradation is an empirical question that has yet to be fully answered. That outcome becomes more likely, however, as more countries establish enforceable property rights and implement optimal environmental policies.

Multilateral Environmental Agreements

The other way in which environmental groups are calling on trade policy to help achieve environmental objectives is to use it as a carrot or stick to entice or prod countries to sign and abide by multilateral environmental agreements. This argument may have some validity, but great care is needed in drawing out its policy implications. Efforts to combat global environmental problems such as ozone depletion or climate change, for example, are likely to encounter the free-rider problem. One of the more obvious and possibly more cost-effective ways of reducing the free-rider problem is to write trade provisions into the agreements, as was done in the 1987 Montreal Protocol on reducing the use of CFCs and halons to slow ozone depletion. No GATT contracting party has formally objected to that use of trade policy. Nor has any contracting party objected to the bans on trade in ivory, rhinoceros horn, and tiger products that are part of the Convention on International Trade in Endangered Species, or to the provisions in the Basel Convention on trade in hazardous wastes. Future conflicts may well arise, however, if trade provisions are drafted into more contentious multilateral environmental agreements (say, an attempt to impose a global carbon tax). That is why this matter figures prominently on the agenda of the WTO Committee on Trade and Environment. Discussions of trade provisions and multilateral environmental agreements in the GATT and the WTO have focused on the idea of providing waivers on a case-by-case basis. Recently, the idea of providing an "environmental window" for multilateral environmental agreements within the GATT exceptions clause (Article XX) also has been advanced.

To help assess the appropriate role for trade policy in multilateral environmental agreements, it is helpful to recall that supporters of trade liberalization and of environmental protection share a common goal: improving social welfare. They also share a common problem: the need to foster multilateral cooperation to achieve that objective, because in each sphere (the economy and the environment) there is considerable and increasing interdependence among countries. But the two groups differ in that supporters of liberal world trade have understood its virtues for two centuries and have been active for more than fifty years in building institutions such as the GATT and the WTO to help achieve their goal. Widespread concerns about the environment, on the other hand, are relatively new, and only recently have supporters of environmental protection become significant players in international policy arenas.

Understandably, supporters of liberal trade, the GATT, and the WTO resent the encroachment of these "new kids on the block" on what they perceive as their hard-won territory—especially when they believe that reducing trade barriers is likely to be environmentally friendly and consistent with sustainable development because it allows the world to use its resources more efficiently. But advocates of greater environmental protection are equally frustrated that international agreements as important as those resulting from the Uruguay Round can be implemented without being subject to environmental impact assessments or environmental safeguards.

There is clearly scope for greater understanding and altered strategies on both sides. More than that, there is the distinct possibility that both groups can advance their objectives by working together—a win-win outcome. Some observers believe that it ultimately may require a world environment organization to set rules, consolidate existing international environmental agreements and negotiate new ones, monitor compliance, and settle disputes over environmental policies—in the same way that the GATT has presided over trade rules and policies for the past five decades (Esty 1994). And just as the GATT and the WTO strengthen the capacity of governments to resist the demands of domestic interest groups seeking higher import taxes, a world environment organization could help governments resist interest group demands to lower environmental standards (Deardorff 1995).

The advantage of a world environment organization for liberal traders, Esty argues, is that it could redirect environmentalists' attention away from trade measures and toward more appropriate policy instruments for achieving environmental objectives. This shift would allow both sets of policies to contribute, in mutually supportive ways, to the common goals of sustainable development and improvement in the quality of life.

A world environment organization is unlikely to be created in the near future, however, partly because many governments are under pressure to downsize. Moreover, in some countries there is growing mistrust and even hostility toward international bureaucracies. In any case a world environment organization—like the International Labor Organization, International Standards Organization, World Intellectual Property Organization, and other standards-setting international organizations—would lack the teeth to enforce agreements. Hence the search for more workable institutions for international environmental cooperation (Barrett 1995). Moreover, the issue of whether the rulings of the WTO or a world environment organization would take precedence would need to be resolved. In cases where the objectives of liberal traders and environmentalists are in conflict, achieving the welfare-maximizing outcome would require both groups to compromise (Corden 1996).

Thus the trade policy community needs to be involved in the negotiation of multilateral environmental agreements that are likely to include trade provisions and in the development of criteria that WTO members can use to assess in advance the acceptability of trade restrictions within such agreements. There are several such criteria, some of which were enunciated at the United Nations Conference on Environment and Development. First, trade provisions must be strictly necessary, in the sense that there are no alternative, more effective instruments, and they must be effective in achieving their environmental objectives. Second, where trade restrictions are required—because less costly policy measures are unavailable—they should be proportionate to the size of the associated environmental problem and the least restrictive measure available. Finally, trade measures should be transparent and not protectionist, and where possible they should be consistent with GATT principles of nondiscrimination (most favored nation and national treatment) and key environmental principles (polluter pays, precautionary principles). If those conditions are met, WTO members would be unlikely to object to the use of trade measures in multilateral environmental

agreements (witness the absence of objections by GATT contracting parties to the trade provisions in the Montreal Protocol and the Convention on International Trade in Endangered Species). Hence even the possible use of trade provisions in multilateral environmental agreements does not provide sufficient reason to amend GATT Article XX to allow trade measures to be used for environmental protection.

What Should Developing Countries Do?

Demands for greater harmonization of domestic policies for reasons of competitiveness, coupled with the greening of world politics (and a growing interest in worker rights and other human rights beyond national borders), are likely to put the WTO and trade policy under pressure to perform tasks for which they were not designed and are not well suited. These developments come at a time when the WTO needs first to consolidate its role in the world and ensure the implementation of the Uruguay Round before moving into these more thorny issues that are only peripherally connected with trade.[9]

The pressure on the WTO to become more entwined with environmental issues should be of considerable concern to developing countries. The reason is not so much that developing countries may be required to impose higher environmental standards to avoid seeing trade barriers raised against their exports. In fact, the competitiveness of industries in middle-income countries with mid-level environmental standards may well be enhanced if low-income countries with low standards are required to raise their standards to minimum acceptable levels. Even the negative direct effect for low-income countries of having to raise their standards could be offset somewhat by terms of trade improvements if many such countries were to raise their standards simultaneously. The benefits would be even greater if developing countries taxed the export (or better still, the production) of environmentally sensitive products rather than leaving industrial countries to tax their imports. Still, people in developing countries are suspicious of high-income countries and object to what they perceive as social imperialism and a denial of their national sovereignty.

Although developing countries are not being singled out for environmental issues, environmental standards do tend to be lower in developing countries simply because they are poorer. That, together with the fact that their comparative advantages often are in natural resource– and pollution-intensive industries, means that those countries are vulnerable to being pressured to enforce stricter standards or to facing less market access for their exports and less foreign investment from countries with stricter standards. Furthermore, if efforts to use trade policy to harmonize standards lead to retaliation and counter-retaliation, the result could be a weakening of the multilateral trading system on which developing countries increasingly depend as they liberalize their economies. One possible consequence is that developing countries could seek refuge from anti–eco-dumping duties through association with or accession to the European Union or the North American Free Trade Agreement, where they are more likely to be compensated for raising their standards over time. In such cases any net gain they might enjoy could well be at the expense of excluded developing countries.

Arguments against Using Trade-Restrictive Measures

Since the entwining of environmental issues with trade and investment policy is more likely to tighten than to disentangle in the foreseeable future, how should developing countries respond? One response is to point out that industrial countries had lower standards at earlier stages of their development. Also, since developing countries contribute a disproportionately small amount per capita to global environmental problems, they should be compensated for contributing to their solutions rather than threatened with trade sanctions if they fail to contribute. Compensation would be even more justified in cases where industrial countries are demanding that other countries reduce the international psychological spillovers of environmental concerns mentioned earlier.

Developing countries could also respond by more widely disseminating the sound arguments for not using trade-restrictive measures to achieve environmental objectives (and hence for not amending Article XX of the GATT to allow trade discrimination for environmental purposes). Among those arguments:

- Differences in standards are a legitimate source of comparative advantage insofar as they reflect differences in countries' resource endowments, preferences, and abilities.
- Standards rise with per capita income, and liberal trade promotes income growth.
- There is little reason from theory or empirical evidence to expect that the raising of standards in industrial countries will contribute significantly to production costs and hence to trade and investment patterns, or that downward harmonization of standards (a race to the bottom) is occurring.
- If freer trade were to worsen welfare because of inappropriate environmental policies in some countries, nontrade measures such as labeling ("dolphin-friendly tuna") would be more cost-effective than trade policies because they allow consumers to exercise their preferences through the market.
- Outside pressure on developing countries to raise their environmental standards could be used by domestic protectionist groups to argue against their government's export-oriented development strategy.
- The GATT rules-based multilateral trading system is threatened by the risk that environmental groups will be captured by traditional protectionist groups in countries with high environmental standards. The system will be at even further risk if Article XX of the GATT is amended to include in its list of exceptions the use of unilateral trade measures for environmental purposes.

More Empirical Analysis

More empirical analysis to support some of these arguments is sorely needed. The Uruguay Round and the Intergovernmental Panel on Climate Change have made clear that empirical studies are far more powerful than abstract arguments in focusing attention on the need for policy reform and the shape it should take. Those

quantitative exercises have provided the world with a suite of multisector, multicountry models that can be modified to estimate the linkages between trade and resource depletion and environmental degradation. Efforts are now being made in that direction, but there is great scope for further, high-payoff research.

Such forward-looking modeling requires the inclusion of endogenous behavioral relationships for private households and firms and for governments, to capture the demographic transition and the transitions in trade and environmental policies that typically accompany growth in per capita income. Government behavior needs to be included in the base case not to suggest that policy choices are inevitable but rather to represent what would happen if no further action were taken. Various alternative cases can be compared against that base case, with a view to convincing governments to choose a different set of policies.

There are also numerous avenues for microempirical case study analysis. The effect of higher environmental standards on the production costs of firms in industrial countries, and hence on their trade, has been limited (Jaffe and others 1995). To what extent has that outcome occurred because the standards were raised only after a more environmentally friendly technology was about to become available? How do those small cost increases compare with the costs that would be incurred by producers in developing countries if they were required to reach similar (or even lower) standards? And to what extent is the finding of little effect on trade patterns from the raising of environmental standards (reported in Tobey 1990, for example) due to the fact that the country with higher standards raised its trade barriers to offset any decline in its firms' international competitiveness? What explains the inverted U–shaped relationship between per capita income and emissions (for example, changing composition of production, stricter environmental standards, cleaner new technologies, relocation of pollutive activities to poorer countries)? At what level of per capita income will that curve peak for any particular country, and why? Is there any evidence of competitiveness-driven downward harmonization of environmental standards (a race to the bottom)?

More Dialogue and Compromise

Helpful though such analysis would be, more dialogue and compromise between high-income and developing countries will still be needed. One approach would be for developing countries to commit themselves to enforcing minimum standards and to raising those standards over time according to a specified schedule. In return, the remaining barriers to their exports to high-income markets would be gradually reduced, and interest groups in high-income countries would be less able to deny that improvements in environmental standards are positively related to income and trade growth. Such an approach would use trade policy as a carrot rather than as a stick. Likewise, if developing countries were seen to be enforcing their standards on their foreign investors, concerns about capital outflows to pollution havens would be less justified.

Alternatively or additionally, developing countries could transfer the onus back to countries with high standards by insisting that their firms adopt the same high

standards when they invest in developing countries as they do in other industrial countries (thereby ensuring the importation of more environmentally friendly technologies). Anxiety about deforestation, for example, could be reduced if developing countries were to demonstrate that they are prepared to police restrictions on felling in return for, say, greater access to industrial country markets or aid (for example, through the Global Environment Facility, which compensates countries for cleanup costs whose benefits are external).

A more controversial suggestion is Rodrik's (1994) concerning trade policy and social standards. He makes a case for countries with high standards to take action against a trading partner if trade with that country violates a widely held social standard (that is, one that is accepted by export and consumer interests in the high-standard countries as well as by import-competing producers and environmental or labor groups). This argument rests on the premise that an erosion of confidence in the fairness of the trading system may ultimately be more costly to the world economy than the action against the offending trading partner. Rodrik suggests that the Safeguards Agreement of the Uruguay Round be broadened to incorporate a "social safeguards" clause that allows the aggrieved country to restrict the offending imports and compensate the trading partner. Rodrik recognizes that this strategy could do more harm than good (not least because it would formalize a link between trade policy and social standards). Even so, he argues that its merits need to be weighed against the other options available to developing countries to minimize the damage from the encroachment of social issues into the trade policy domain. However, the sobering history of abuse of the GATT's other safeguard clauses offers little reason for optimism about this proposal (Finger 1995).

Finally, what principles should govern the design of trade policies and trade-related environmental policies to ensure equitable and sustainable development? Several have been mentioned in passing. Even if developing countries were simply to discuss a list of such principles with industrial countries, the dialogue itself could diffuse some of the concerns expressed by environmental groups. The Asia-Pacific Economic Cooperation (APEC) group, with its diverse but relatively small membership, provides an obvious forum for such discussion before the much larger WTO membership debates the issues. In the same spirit, APEC could start monitoring trade-related environmental measures as part of its overall compilation of trade impediments in Asia and the Pacific. In addition, as a priority in its trade facilitation and liberalization initiatives launched at Bogor in November 1994, APEC might actively seek the removal of trade policies that incidentally harm the environment—again, providing a regional win-win example of what eventually might be achievable globally through the WTO.

Notes

1. The empirical finding that many forms of pollution first increase and then decrease as per capita income rises has been reported by numerous authors (Grossman 1995; Grossman and Krueger 1993, 1995; Hettige, Lucas, and Wheeler 1992; Radetzki 1992; Seldon and Song 1994; and Shafik and Bandyopadhyay 1992). This finding suggests that the demand for implementing and enforcing pollution

abatement policies is income-elastic and that an inverted U-shaped environmental transition may occur as stricter environmental policies are implemented and less-polluting technologies are introduced. See also Deacon and Shapiro (1975) on the correlation between income levels and voter attitudes toward environmental priorities. Studies aimed at explaining this transition (sometimes called an environmental Kuznets curve) are now beginning to emerge. Beltratti (1995) seeks to explain it in terms of transitional dynamics of endogenous growth models, while Jones and Manuelli (1995) provide a positive political economy model.

2. Similarly, if as they grow economies were to institutionally shorten working hours per week, raise wages for time worked outside those hours, or otherwise increase the cost of labor time in attempting to raise labor standards, that would speed the transformation of those economies' comparative advantages away from labor-intensive activities. If those institutional changes mainly affected unskilled labor, the competitiveness of developing countries in unskilled labor-intensive products would strengthen even faster (Anderson forthcoming).

3. Some observers would argue that psychological spillovers are less worthy of consideration than physical spillovers, not least because they are less measurable, less objective, and hence offer more scope for environmentalists to be "captured" by traditional protectionists. Others would counter that there is so much uncertainty about the extent and effects of physical spillovers that they too are subjective and hence are qualitatively no different from psychological spillovers. Nor is there any reason a priori to presume that one spillover is more important than the other in some willingness-to-pay sense.

4. Even the threat of trade restrictions can be environmentally counterproductive. The possibility of European import bans on tropical hardwood logs (together with tariff escalation on timber product imports) encouraged Indonesia to ban log exports. But since felling and timber product exports have been allowed to continue, this policy has simply lowered the domestic price of logs and thereby raised effective assistance to Indonesia's furniture and other timber-using industries to extremely high levels (GATT 1991, p. 127). With lower log prices and lower-quality saw-milling techniques than in importing countries, less of each tree is now used and little reduction in logging has been observed since the log export ban was introduced.

5. The ban on ivory trade again provides a case in point. By lowering the value of elephant products, the ban reduces the incentive for rural Africans to tolerate elephants trampling their crops and so ultimately could result in more rather than less culling of elephants in some areas. In other areas with poor meat storage and transport facilities, the ivory trade ban has reduced the value of the animal so much that it is no longer profitable to cull the herd. An unfortunate consequence is that bushland in national parks is being devastated by the increased number of elephants, which is endangering other species (Barbier and others 1990; Barbier and Schulz 1995).

6. The theory can be made even more complex by allowing for imperfect competition and strategic environmental policymaking (Ulph 1994).

7. That does not mean, however, that it would be inexpensive for developing country firms to conform to the high environmental standards of industrial countries. The reason is that local firms in developing countries often use older, less environmentally friendly technology that is more costly to adapt. Indeed, the cost of conformance in industrial countries may be relatively low simply because the raising of legislated standards to some extent follows rather than precedes the development of cleaner technologies. That has been cited as one reason for the relative ease with which agreement was reached on reducing the use of ozone-depleting substances under the Montreal Protocol (Enders and Porges 1992).

8. See, for example, Baumol (1971) and Siebert (1974). Such protection from import competition cannot be justified on economic efficiency grounds (or, for that matter, on environmental grounds) because the environmental policy is aiming to eliminate an unjustifiable (implicit) subsidy arising through undervaluation of environmental resources, rather than to add an unjustifiable tax (Snape 1992).

9. The suggestion has been made, for example, that the WTO become active in monitoring and enforcing agreed minimum social standards. That presumably would involve environmental and labor standards being reviewed as part of the WTO's Trade Policy Review Mechanism. Since that mechanism is already stretched to its limit in covering even the major trade policies of contracting parties, such an addition to its workload would require a substantial boost to its resources—not to mention the extra burden on those employed in national capitals when the reviews are under way. An even greater potential increase in workload would result for the WTO's dispute settlement mechanism.

References

Alesina, Alberto, and Enrico Spolaore. 1995. "On the Number and Size of Nations." NBER Working Paper 5050. National Bureau of Economic Research, Cambridge Mass.

Anderson, Kym. 1992a. "Effects on the Environment and Welfare of Liberalizing World Trade: The Cases of Coal and Food." In Kym Anderson and Richard Blackhurst, eds., *The Greening of World Trade Issues.* Ann Arbor, Mich.: University of Michigan Press and London: Harvester Wheatsheaf.

———. 1992b. "The Standard Welfare Economics of Policies Affecting Trade and the Environment." In Kym Anderson and Richard Blackhurst, eds., *The Greening of World Trade Issues.* Ann Arbor, Mich.: University of Michigan Press and London: Harvester Wheatsheaf.

———. 1995. "The Entwining of Trade Policy With Environmental and Labor Standards." In Will Martin and L. Alan Winters, eds., *Implications of the Uruguay Round for Developing Countries.* Washington, D.C.: World Bank.

———. Forthcoming. "Social Policy Dimensions of Economic Integration: Environmental and Labor Standards." In Takatoshi Ito and Anne O. Krueger, eds., *Regionalism versus Multilateral Trade Arrangements.* Chicago: University of Chicago Press for the National Bureau of Economic Research.

Barbier, E. B., and C. E. Schulz. 1995. "Wildlife, Biodiversity and Trade." Nota Di Lavoro 52.95. Fondazione Eni Enrico Mattei, Milan.

Barbier, E. B., J. C. Burgess, T. M. Swanson, and D. W. Pearce. 1990. *Elephants, Economics and Ivory.* London: Earthscan.

Barrett, S. 1995. "Towards a Theory of International Environmental Cooperation." Nota Di Lavoro 60.95. Fondazione Eni Enrico Mattei, Milan.

Baumol, William. 1971. *Environmental Protection, International Spillovers, and Trade.* Stockholm: Almqvist and Wiksell.

Beckerman, Wilfred. 1992. "Economic Growth and the Environment: Whose Growth? Whose Environment?" *World Development* 20(4): 481–96.

Beltratti, Andrea. 1995. "Can a Growth Model With Defensive Expenditures Generate an Environmental Kuznets Curve?" Paper prepared for a workshop on Designing Economic Policy for Management of Natural Resources and the Environment, May 12–13, Venice.

Bhagwati, J. N. 1996. "The Demand to Reduce Domestic Diversity Among Trading Nations." In J. N. Bhagwati and R. E. Hudec, eds., *Fair Trade and Harmonization: Prerequisites for Free Trade?* Cambridge, Mass.: MIT Press.

Bhagwati, J. N., and T. N. Srinivasan. 1996. "Trade and the Environment: Does Environmental Diversity Detract from the Case for Free Trade?" In J. N. Bhagwati and R. E. Hudec, eds., *Fair Trade and Harmonization: Prerequisites for Free Trade?* Cambridge, Mass.: MIT Press.

Casella, Alessandra. 1995. "Free Trade and Evolving Standards." CEPR Discussion Paper 1204. Centre for Economic Policy Research, London.

Charnovitz, Steve. 1991. "Exploring the Environmental Exceptions in GATT Article XX." *Journal of World Trade* 25(5): 37–55.

Chichilnisky, Graciela. 1994. "North-South Trade and the Global Environment." *American Economic Review* 84(4): 851–74.

Copland, B. R., and M. S. Taylor. 1995. "Trade and Transboundary Pollution." *American Economic Review* 85(4): 716–37.

Corden, W. M. 1996. "The Environment and Trade Policy." In W. M. Corden, ed., *Trade Policy and Economic Welfare,* rev. ed. Oxford: Clarendon Press.

Deacon, R., and P. Shapiro. 1975. "Private Preference for Collective Goods Revealed Through Voting on Referenda." *American Economic Review* 65: 943–55.

Deardorff, A. V. 1995. "International Externalities in the Use of Pollution Policies." Discussion Paper 383. University of Michigan, School of Public Policy, Ann Arbor.

Enders, Alice, and Amelia Porges. 1992. "Successful Conventions and Conventional Success: Saving the Ozone Layer." In Kym Anderson and Richard Blackhurst, eds., *The Greening of World Trade Issues.* Ann Arbor, Mich.: University of Michigan Press and London: Harvester Wheatsheaf.

Esty, D. C. 1994. *Greening the GATT: Trade, Environment, and the Future.* Washington, D.C.: Institute for International Economics.

Finger, J. Michael. 1995. "Legalized Backsliding: Safeguard Provision in the GATT." In Will Martin and L. Alan Winters, eds., *The Uruguay Round and the Developing Economies.* Washington, D.C.: World Bank.

GATT (General Agreement on Tariffs and Trade). 1971. *Industrial Pollution Control and International Trade.* GATT Studies in International Trade 1. Geneva: GATT Secretariat.

———. 1991. *Trade Policy Review: Indonesia.* Geneva: GATT Secretariat.

———. 1994. *International Trade Statistics 1994.* Geneva: GATT Secretariat.

Grossman, G. M. 1995. "Pollution and Growth: What Do We Know?" In Ian Goldin and L. Alan Winters, eds., *The Economics of Sustainable Development.* Cambridge: Cambridge University Press.

Grossman, G. M., and A. B. Krueger. 1993. "Environmental Impacts of a North American Free Trade Agreement." In P. M. Garber, ed., *The Mexico-U.S. Free Trade Agreement.* Cambridge, Mass.: MIT Press.

———. 1995. "Economic Growth and the Environment." *Quarterly Journal of Economics* 110(2): 353–78.

Harrison, A. E., and G. S. Eskeland. 1996. "Playing Dirty? Multinationals and the Pollution Haven Hypothesis." Columbia University, Department of Economics, New York.

Hettige, Hemamala, Robert E. B. Lucas, and David Wheeler. 1992. "The Toxic Intensity of Industrial Production: Global Patterns, Trends, and Trade Policy." *American Economic Review* 82(2): 478–81.

Hillman, A. L., and H. N. Ursprung. 1992. "The Influence of Environmental Concerns on the Political Determination of Trade Policy." In Kym Anderson and Richard Blackhurst, eds., *The Greening of World Trade Issues.* Ann Arbor, Mich.: University of Michigan Press and London: Harvester Wheatsheaf.

Hoekman, Bernard, and Michael Leidy. 1992. "Environmental Policy Formation in a Trading Economy: A Public Choice Perspective." In Kym Anderson and Richard Blackhurst, eds., *The Greening of World Trade Issues.* Ann Arbor, Mich.: University of Michigan Press and London: Harvester Wheatsheaf.

Jaffe, A. B., S. R. Peterson, P. R. Portney, and R. N. Stavins. 1995. "Environmental Regulation and the Competitiveness of U.S. Manufacturing: What Does the Evidence Tell Us?" *Journal of Economic Literature* 33(1): 132–63.

Jones, L. A. and R. E. Manuelli. 1995. "A Positive Model of Growth and Pollution Controls." NBER Working Paper 5205. National Bureau of Economic Research, Cambridge, Mass.

Krueger, A. O. 1977. *Growth, Distortions, and Patterns of Trade among Many Countries.* Princeton, N.J.: International Finance Section.

Lawrence, R. Z. 1995. *Regionalism, Multilateralism, and Deeper Integration.* Washington, D.C.: Brookings Institution.

Leamer, E. E. 1987. "Paths of Development in the Three Factor, n-Good General Equilibrium Model." *Journal of Political Economy* 95(5): 961–99.

Leonard, N. J. 1988. *Pollution and the Struggle for World Product: Multinational Corporations, Environment and International Comparative Advantage.* Cambridge: Cambridge University Press.

Levinson, Arik. 1996. "Environmental Regulation and Industry Location: International and Domestic Evidence." In J. N. Bhagwati and R. E. Hudec, eds., *Fair Trade and Harmonization: Prerequisites for Free Trade?* Cambridge, Mass.: MIT Press.

Low, Patrick. 1992. "Trade Measures and Environmental Quality: The Implications for Mexico's Exports." In Patrick Low, ed., *International Trade and the Environment.* World Bank Discussion Paper 159. Washington, D.C.

Meadows, D. H., D. L. Meadows, J. Randers, and W. W. Behrens III. 1972. *The Limits to Growth.* New York: Universe Books.

Palmer, Karen, Wallace E. Oates, and Paul R. Portney. 1996. "Tightening Environmental Standards: The Benefit-Cost or the No-Cost Paradigm?" *Journal of Economic Perspectives* 9(4): 119–32.

Porter, M. E., and C. van der Linde. 1996. "Towards a New Conception of the Environment-Competitiveness Relationship." *Journal of Economic Perspectives* 9(4): 97–118.

Radetzki, M. 1992. "Economic Growth and Environment." In Patrick Low, ed., *International Trade and the Environment.* World Bank Discussion Paper 159. Washington, D.C.

Rodrik, Dani. 1994. "Developing Countries after the Uruguay Round." CEPR Discussion Paper 1084. Centre for Economic Policy Research, London.

Seldon, T. M., and D. Song. 1994. "Environmental Quality and Development: Is There a Kuznets Curve for Air Pollution Emissions?" *Journal of Environmental Economics and Management* 27(2): 147–62.

Shafik, Nemat, and Sushenjit Bandyopadhyay. 1992. "Economic Growth and Environmental Quality: Time Series and Cross-Country Evidence." Policy Research Working Paper 904. World Bank, Washington, D.C.

Siebert, H. 1974. "Environmental Protection and International Specialization." *Welwirtschaftliches Archiv* 110: 494–508.

Siebert, H., J. Eichberger, R. Gronych, and R. Pethig. 1980. *Trade and Environment: A Theoretical Enquiry.* Amsterdam: Elsevier.

Snape, R. H. 1992. "The Environment, International Trade and Competitiveness." In Kym Anderson and Richard Blackhurst, eds., *The Greening of World Trade Issues.* Ann Arbor, Mich.: University of Michigan Press and London: Harvester Wheatsheaf.

Steenblik, R. P., and P. Coroyannakis. 1995. "Reform of Coal Policies in Western and Central Europe: Implications for the Environment." *Energy Policy* 23(6): 537–54.

Steil, Benn. 1994. "'Social Correctness' is the New Protectionism." *Foreign Affairs* 73(1): 14–20.

Tobey, J. A. 1990. "The Effects of Domestic Environmental Policies on Patterns of World Trade: An Empirical Test." *Kyklos* 43(2): 191–209.

Ulph, Alistair. 1994. "Environmental Policy and International Trade—A Survey of Recent Economic Analysis." Paper presented at the Workshop on Designing Economic Policy for the Management of Natural Resources and the Environment, September 7–9, Crete.

Wilson, J. D. 1996. "Capital Mobility and Environmental Standards: Is There a Theoretical Basis for a Race to the Bottom?" In J. N. Bhagwati and R. E. Hudec, eds., *Fair Trade and Harmonization: Prerequisites for Free Trade?* Cambridge, Mass.: MIT Press.

Comment on "Environmental Standards and International Trade," by Kym Anderson

Maureen L. Cropper

Kym Anderson makes two important points about trade policy and environmental standards. First, when it comes to domestic environmental problems countries should be allowed to set their own standards. Low-income countries may choose lower air and water quality standards than high-income countries and, as a consequence, develop a comparative advantage in pollution-intensive industries. Differences in competitiveness that result from differences in domestic environmental standards are not, however, a source of inefficiency and do not call for a trade policy response.

Second, when a country's production or consumption decisions impose environmental externalities on other countries, there may be a theoretical case for using trade policy to correct these externalities. Many international environmental problems—acid rain, global warming, biodiversity destruction—require multilateral cooperation to achieve a first-best solution. But if the first-best solution cannot be achieved, trade policy may produce a second-best outcome. Markusen (1975) and Baumol and Oates (1975, 1988) have demonstrated that, in the case of transboundary pollution, a tariff on a polluting good may improve welfare. The threat of trade sanctions also may provide an incentive for countries to abide by multilateral environmental agreements. Even in these cases, however, trade policy must be used with caution: as a second-best solution it is likely to be an extremely blunt instrument for correcting environmental externalities.

Because these points are not made often enough, I would be remiss if I emphasized situations in which environmentally motivated trade policies might improve welfare. Closer analysis of these cases, however, makes it possible to show how weak the argument is for trade policy solutions to environmental problems. Here I consider four instances where trade policy might be used to achieve environmental objectives: when domestic environmental standards do not reflect social preferences, when nothing else is being done to correct transboundary pollution problems, when

Maureen L. Cropper is principal economist in the Environment, Infrastructure, and Agriculture Division of the Policy Research Department at the World Bank.

Annual World Bank Conference on Development Economics 1996

endangered species are being overharvested, and when used as an enforcement mechanism for international environmental agreements.

Using Trade Policy to Correct Domestic Environmental Problems

If one accepts the premise that countries should be allowed to set domestic environmental standards, then these standards, much like differences in natural resource endowments, help determine a country's comparative advantage. As such, they are not cause for environmental tariffs, at least not on efficiency grounds. Under autocratic regimes, however, domestic environmental standards may fail to reflect social preferences. For example, dictators might accept bribes to allow shipments of hazardous waste to enter their country, while members of society bear the risks associated with disposing of the waste. This is also a case in which the employment benefits of less stringent environmental standards are likely to be slim.

Is this enough of a problem that it requires a trade policy response? Research by Rajamani (1995) suggests that it may be, and thus provides support for agreements such as the Basel Convention, designed to monitor international trade in hazardous waste. Rajamani examined imports of nonferrous metal-bearing waste (for example, ash or sludge containing metals) by ninety countries during 1980–90. She found that such imports tend to flow from high-income to low-income countries and from high population-density to low population-density countries. Imports are also more likely to flow from more democratic to less democratic countries.[1] This finding suggests a possible problem requiring further investigation: if disposal practices in less democratic countries are less safe than the public would prefer, then other countries might be warranted in imposing minimum safe disposal standards, or at least in monitoring such disposal (as called for by the Basel Convention).

Using Trade Policy to Correct International Environmental Problems

The case for using trade policy is stronger in the case of international environmental problems. Transboundary pollution and global environmental problems—global warming, biodiversity destruction—require multilateral cooperation to achieve a first-best solution, such as a tax on carbon emissions. If the first-best solution cannot be achieved, however, there may be a role for trade policy as a second-best instrument.

Trade policy might be welfare enhancing, for example, in cases where a country imposes environmental externalities on other countries. Suppose that country A produces steel and sulfur dioxide emissions. People in country A suffer from this pollution, as do their neighbors in country B. If country A does nothing to regulate air pollution, the welfare of both countries can be improved if country B levies an import duty on steel (Baumol and Oates 1975, 1988). There are three important caveats to this outcome, however. First, the tariff is a second-best solution. It will never achieve the same allocation as a tax on sulfur dioxide emissions equal to marginal social damages, the first-best policy instrument. Second, the optimal tariff

(from the perspective of both countries) is lower than the tariff that country B would likely choose (the one that maximizes country B's welfare). Third, for the tariff to improve welfare, country B must import steel.

Given these caveats, the scope for improving welfare through environmentally motivated tariffs is likely to be small. More realistic cases—as when many countries are affected by country A's pollution—only complicate the conditions under which a tariff on polluting goods improves welfare.

Another case in which trade policy may solve international environmental problems concerns the protection of endangered species. The Convention on International Trade in Endangered Species (CITES) prohibits trade in ivory, tiger skins, and other products culled from endangered species. Does such an agreement enhance welfare? The answer depends on the source of the endangerment. If endangered species are threatened by overharvesting (say, by poachers who are killing pandas for their pelts) and if the countries that are home to the species do nothing to regulate harvesting, then the treaty may improve welfare. But the treaty does nothing to protect species that are threatened by habitat encroachment; for example, by farmers who slash and burn the animals' habitats for agriculture and have no interest in harvesting the species. In this case the treaty may actually be counterproductive, since allowing trade in endangered species could provide the necessary incentives to protect animal habitats.

The case of elephants in Zimbabwe is one instance in which CITES may have reduced incentives to preserve species. Zimbabwe's Campfire program is designed to provide people in villages surrounding habitat preserves with incentives to preserve elephants and other big game. The program is one of the few instances in which local peoples have been induced to protect habitats, and its success derives from the fact that villagers who would otherwise use animal habitats for agriculture actually receive money for preserving them. Most of the monetary rewards come from fees paid by outfitters for the right to hunt a sustainable yield of big game. Until the CITES treaty, ivory sales provided $5 million a year in funds for the program (Child forthcoming). Thus the treaty has reduced incentives for habitat preservation.

The important question for trade policy is which cause of biodiversity destruction is more important—habitat destruction or overharvesting of species. Most experts believe that habitat destruction is by far the more important cause of biodiversity loss (Swanson 1994), which suggests that the CITES treaty may not be as effective a means of preserving endangered species as some of its proponents believe.

A third instance in which trade policy might be used to correct international environmental problems is when trade sanctions are used as a means of enforcing agreements to deal with global environmental externalities. Such sanctions exist to enforce the Montreal Protocol and, as Anderson notes, no one has objected to them in this context. The Montreal Protocol is somewhat unusual, however. Murdoch and Sandler (forthcoming) argue that the Montreal Protocol codifies what is, in reality, the noncooperative solution to chlorofluorocarbon (CFC) reduction; that is, the treaty calls for countries to do what they would have done without the treaty. Many

countries had reduced CFC emissions before the treaty went into effect, to levels that exceeded what the agreement called for.

The situation is, however, likely to be very different in the case of an agreement to reduce carbon dioxide emissions. In the case of global warming, the noncooperative and cooperative solutions to reducing greenhouse gas emissions call for carbon tax rates that differ by an order of magnitude (Nordhaus and Yang 1996). Efforts to use trade sanctions to enforce a cooperative solution would likely meet with resistance from many countries.

Conclusion

The arguments for using trade policy to correct environmental externalities are weak and are not sufficient cause for amending the GATT's rules. There are, moreover, many instances in which welfare-reducing trade policies have been instituted in the name of the environment. Environmental problems are not caused by international trade; they are caused by market failures and the absence of property rights. We should not look to trade policy to correct them.

Note

1. Formally, Rajamani regresses the size of imports to country *i* from country *j* on the difference in the per capita incomes of the two countries, the difference in their population densities, and the difference in the democracy scores assigned to their governments. Imports increase the larger is country *j*'s income, population density, and democracy score than country *i*'s.

References

Baumol, William J., and Wallace E. Oates. 1975. *The Theory of Environmental Policy,* 1st ed. New York: Prentice Hall.

———. 1988. *The Theory of Environmental Policy,* 2nd ed. New York: Cambridge University Press.

Child, Brian. Forthcoming. "Zimbabwe." In Ernst Lutz and Julian Caldecott, eds., *Decentralization and Biodiversity Conservation.* Washington, D.C.: World Bank.

Markusen, James R. 1975. "International Externalities and Optimal Tax Structure." *Journal of International Economics* 5: 15–29.

Murdoch, James C., and Todd Sandler. Forthcoming. "The Voluntary Provision of a Pure Public Good: The Case of Reduced CFC Emissions and the Montreal Protocol." *Journal of Public Economics.*

Nordhaus, William D., and Zili Yang. 1996. "RICE: A Regional Dynamic General Equilibrium of Optimal Climate-Change Policy." Yale University, New Haven, Conn.

Rajamani, Kamala. 1995. "International Trade in Hazardous Wastes: An Empirical Analysis." Doctoral dissertation. University of Maryland, Department of Economics, College Park.

Swanson, Timothy M. 1994. *The International Regulation of Extinction.* New York: New York University Press.

Comment on "Environmental Standards and International Trade," by Kym Anderson

Robert E. Hudec

Kym Anderson's article examines proposals to change GATT rules (or at least clarify them) to permit the greater use of trade measures for environmental purposes. Anderson divides these proposals into two major categories. First, he discusses the proposals that GATT/WTO rules be amended to permit governments to unilaterally impose trade measures against other governments whose environmental practices are considered harmful. U.S. restrictions against imports of tuna caught by dolphin-unsafe fishing methods, ruled GATT-illegal by separate GATT panels in 1991 and 1994, are perhaps the best example of such measures (GATT 1991, 1994). Other examples include U.S. restrictions against Taiwan (China) for failing to discourage trade in rhinoceros horns and tiger parts (BNA 1994) and the European Union's 1983 prohibition of imports of baby seal fur, or its current effort to prohibit the import of fur of animals from countries that permit the use of leghold traps (GATT 1985, EC 1991).

Second, Anderson discusses proposals that GATT/WTO rules be amended to permit trade measures authorized by certain international environmental agreements, such as the restrictions on trade in endangered species provided for by the Convention on International Trade in Endangered Species. Because of the potentially greater legitimacy attached to widely adopted international agreements, this second class of internationally authorized trade measures tends to be viewed in a much more favorable light than the first group. Most discussions within the WTO's Trade and Environment Committee about possible amendments to Article XX of the GATT have focused on the possibility of authorizing trade measures in this second category.

Although I agree with Anderson's general skepticism about the utility of trade measures to advance environmental objectives, I believe it is useful to present a more detailed description of the context in which each of these proposals arises and in which each is likely to be dealt with.

Robert Hudec is professor of law at the University of Minnesota.

Annual World Bank Conference on Development Economics 1996

Unilateral Trade Measures

GATT rules on unilateral trade measures are currently in a state of some confusion. The two GATT panel rulings on U.S. tuna restrictions, known as Tuna/Dolphin I and Tuna/Dolphin II, reached somewhat different conclusions. The Tuna/Dolphin I ruling interpreted GATT rules as imposing a rather clear prohibition on unilateral trade restrictions against the environmental practices of other governments. Although the ruling was supported by thirty-nine of the forty GATT countries that took a position on the issue—the fortieth being the United States—it caused a storm of protest from environmentalists. Perhaps because of that protest, the ruling in Tuna/Dolphin II was considerably more equivocal. Although the second ruling did find some of the more extreme aspects of U.S. trade measures to be in violation of the GATT, it also suggested that in other cases unilateral trade measures might be permitted. Since neither ruling was formally adopted by the GATT, however, the definitive interpretation of GATT rules pertaining to unilateral trade measures has not yet been made.

The most likely venue for further efforts to resolve these issues will be yet another panel ruling, triggered by yet another dispute settlement complaint by a WTO member affected by another such unilateral trade measure. Although this legal issue falls within the mandate of the WTO's Trade and Environment Committee, national positions are still too different to permit a negotiated solution. The new WTO dispute settlement procedure could provide the sort of impetus needed to break this deadlock because, for the first time, the procedure promises a legally binding ruling verified by appellate review. If this occurs, then developing countries hoping to defend themselves against unilateral trade restrictions would do well to invest in the legal resources required to participate effectively in the forthcoming GATT litigation.

Anticipating the eventual resolution of this issue, Anderson's article rightly points out that, despite the criticism levied against the ruling in Tuna/Dolphin I, there is still considerable support within the GATT for the broad prohibition of unilateral measures called for by that ruling. The more closely one looks at national positions on this issue, however, the more fragile that general consensus appears. The behavior of the principal practitioners of unilateral trade measures—the United States and the European Union—provides ample evidence of a political house badly divided.

To date, the U.S. government has rejected the legal prohibitions of the Tuna/Dolphin rulings. The U.S. Congress responded to the first ruling by passing an even more rigorous trade restriction against dolphin-unsafe fishing methods generally. More recently, congressional opposition blocked an effort to resolve the tuna/dolphin controversy by means of an international agreement requiring fishing methods that severely limit (but do not eliminate) harm to dolphins. Although the new agreement has been accepted by the U.S. executive branch, several leading environmental groups, and several key legislators, the new agreement has been blocked by several more-demanding environmental groups, their legislative allies, and, possibly, the U.S. tuna fleet (which benefits from the restrictions).

Despite its vigorous commitment to its own unilateral trade measures, however, the United States has not hesitated to attack unilateral trade measures imposed by

others. One example is a recent EU directive calling for trade restrictions on fur imports from countries that allow the use of leghold traps. The United States joined with Canada in protesting the leghold trap restrictions as a violation of the GATT, and their opposition persuaded the European Union to postpone implementation of the trade restrictions until 1997. Anyone listening to the United States complain about the EU measures would have no reason to suspect that the United States employs the very same kind of trade measures itself.

To be sure, the European Union displays the same ability to straddle both sides of the issue. The European Union, after all, led the legal and political assault on the unilateral U.S. trade measures involved in the two Tuna/Dolphin cases—the same European Union that recently promulgated unilateral trade restrictions against countries that permit the use of leghold traps.

A colorful interpretation of these inconsistencies might lead the casual observer to venture a disdainful comment about governmental hypocrisy. Examined more closely, however, these inconsistent positions reveal an important fact about the domestic politics of the trade-environment issue—that is, national governments are badly divided over trade-environment issues. On the one side are the trade ministries that staff WTO delegations and that determine the positions of their government inside the WTO. This group tends to favor the rather strict GATT prohibition on unilateral trade measures, whether for environmental purposes or otherwise. On the other side are the interest groups who, advocating forceful use of unilateral measures, have assembled a strong political coalition that is often capable of achieving legislative majorities on particular issues. Thus, while trade ministries tend to follow their pro-trade policies in GATT affairs, they frequently are unable to make their own governments toe that line. Although it might seem that governments would have trouble facing two directions at once, such a position is actually second nature to most of them.

What are the implications of this political situation for developing countries seeking to prevent unilateral trade measures? The first lesson is not to view GATT legal proceedings as the end of the story. Before GATT rules can exert any influence over the countries that threaten such restrictions, the GATT point of view will have to gain the upper hand in the internal politics of those countries. Moreover, developing countries must realize that their principal allies in the political struggle will be the industrial country trade officials with whom they deal in the WTO. The ultimate objective of WTO proceedings, therefore, is not to win victories over industrial country delegations. It is to strengthen the hand of industrial country trade ministries in their domestic political battles.

Trade Measures in International Environmental Agreements

The legal status of trade measures provided for in international environmental agreements breaks down into two quite different problems. First there is the question of the legal rights under the GATT of governments that have signed the agreement in question. This is largely a nonproblem from a legal point of view. Most legal

experts that have considered the problem believe that a government that has signed an international environmental agreement authorizing trade restrictions in particular situations has waived any rights it may have to object to the application of such trade measures against itself. (There may be legal problems in determining what the agreement means, but these are secondary.)

The main legal problem in this area involves the legal rights under the GATT of governments that have not signed the international environmental agreement. The free-rider problem that figures prominently in Anderson's discussion of international agreements exists solely with regard to such nonsignatories. The question is whether other governments, by virtue of having formed such an agreement, should have the power to override the legal rights under the GATT of nonsignatory governments in order to compel them to observe the disciplines of such an agreement. The claim that governments should have such power is usually based on a best-case scenario involving two major factors: an environmental problem that is serious enough to require some sacrifice of interest and an agreement addressing that problem that represents the will of a broadly representative group of governments. International environmental agreements that most persuasively meet these criteria are the Convention on International Trade in Endangered Species and the Montreal Protocol to protect the ozone layer, each involving more than one-hundred signatory governments and a problem of irreversible environmental harm.

As Anderson points out, none of the trade provisions in international environmental agreements has ever been challenged in the GATT. The reasons are clear. Signatory governments are unlikely to challenge measures that they have agreed to, and nonsignatory governments have so far recognized the futility of asking the GATT to rule such measures illegal when the vast majority of its membership has signed the agreement authorizing them. From a practical point of view one could argue, as Anderson does, that there is really no need to make any formal changes in the GATT, because the GATT does not really threaten such agreements. Indeed, GATT tradition has been to ignore such theoretical problems if they are not causing actual problems. Burying the problem in this way usually has had the advantage of avoiding the need to answer questions that cannot be answered.

There could be considerable political benefit to a GATT amendment affirming the legitimacy of such agreements, however. The earlier, more ambiguous methods used under the GATT worked well when they were dealing with problems between trade ministries, because the ministries knew that they shared common goals. The current trade-environment debate is different, because it involves conflicts between trade ministries and outside groups that do not have established relationships. For the GATT to establish an effective working relationship with these outside groups, it may have to give stronger evidence of its desire to accommodate environmental interests. In any case a legal provision giving de jure recognition to environmental measures that almost certainly have de facto recognition is unlikely to cause much harm. On balance, therefore, I disagree with Anderson's opposition to such an amendment.

Anderson correctly points out that the GATT's legal response to existing agreements, whatever it is, will be less important than its ability to influence the terms of future agreements. He suggests that developing countries give primary attention to making sure that the GATT exercises its influence to limit the trade measures embodied in environmental agreements to their least intrusive form. In carrying out that advice, it is worth remembering the point made earlier about the domestic politics of the trade-environment debate. In the end, the content of international environmental agreements will be determined not by the international negotiating conference that writes the agreement, but by the domestic negotiating conferences where trade interests and environmental interests hammer out their national positions. Thus the best way for the GATT to influence these agreements is by strengthening the hand of its member trade ministries in these national debates.

References

BNA (Bureau of National Affairs). 1994. "President Clinton Announces Sanctions on Taiwan for Rhino, Tiger Parts Trade." *International Trade Reporter* 11: 576.

EC (European Commission). 1991. "Council Regulation Banning Leghold Traps (L 308)." *Official Journal of the European Communities* 34 (April 13).

GATT (General Agreement on Tariffs and Trade). 1985. "European Community: Ban on Importation of Skin of Certain Seal Pups and Related Products." GATTDOC L/5940 1985. Geneva.

———. 1993. "United States: Restrictions on Imports of Tuna." In *Basic Instruments and Selected Documents*. Supplement 39. DS21/R. 1991. Geneva.

———. 1994. "United States: Restrictions on Imports of Tuna." DS29/R. 1994. Geneva.

Floor Discussion of "Environmental Standards and International Trade," by Kym Anderson

A participant from the World Bank's Policy Research Department asked the panelists how they felt about countries like Japan requiring their multinational corporations to use much cleaner technologies in developing countries than those countries require. Anderson responded that the idea is a good one, and that the OECD could do the same sort of thing among its members by establishing higher standards for firms operating in developing countries. The same thing could be done for labor standards.

Maureen Cropper (discussant) agreed that such standards are desirable but argued that imposing them is not always the best approach. Certainly, if a multinational corporation enters a developing country to produce exports, it has an obligation to produce them in an environmentally friendly way. But that situation differs from one in which India receives, say, a loan or grant to reduce mine tailings. If, instead, the residents were given the funds directly, so that everyone received an extra hundred dollars in income, they would be unlikely to use that money to clean up the mine waste. It would be difficult for economists who believe in choice to defend such a requirement, however.

The participant from the World Bank then asked the panelists what they thought about the Bank's policies, which set high environmental standards for Bank projects and pass the mitigation costs on to Bank clients, regardless of their own standards. Anderson believed that the Bank's high standards are a good starting point, although he was not sure how effective they are in practice.

Robert Hudec (discussant) pointed out that neither of the participant's questions was related to the main subject of Anderson's article, the use of trade restrictions in support of environmental policy. What is happening, Hudec said, is that people everywhere are becoming more concerned about environmental policy—as reflected in corporate policies, national legislation, the activities of consumer and other public interest groups, and international gatherings like the United Nations 1992

This session was chaired by Ismail Serageldin, vice president, Environmentally Sustainable Development, at the World Bank.

Annual World Bank Conference on Development Economics 1996

Conference on Environment and Development in Rio de Janeiro. With all this activity, changes in trade policy are just a small part of the developing consensus on environmental policies. Although increasing concern with the environment seems both inevitable and beneficial, the fact remains that trade measures are a costly and inefficient way of achieving worthwhile environmental objectives.

A participant from American University asked how the discussants felt about a passage in Anderson's article that had not been mentioned. Discussing how developing countries would respond to industrial country efforts to raise environmental standards, Anderson pointed out that industrial countries had had lower standards at earlier stages of their development and that developing countries, which have contributed much less to global environmental problems, should be compensated for contributing to their solutions rather than have that contribution demanded of them under threat of trade sanctions. The participant suggested that World Bank policies be adapted to include such a proposal.

Ismail Serageldin (chair) responded by saying that the Global Environment Facility, of which the World Bank is trustee, was created to address just that point. The facility helps cover the incremental costs of projects in developing countries—those activities for which benefits are global and costs are local. For example, if the costs of improving local air standards are more than offset by the benefits that accrue to local residents, the Bank assumes that it is in a country's interests to reduce air pollution, regardless of whether it helps reduce global warming. Only in projects where benefits are global and costs are local does the facility provide an outright grant, which industrial countries pay for.

Serageldin further clarified the World Bank's position on environmental standards. First, the Bank is concerned about people's welfare, and pollution and environmental degradation are antithetical to welfare maximization. Second, many of the defenses countries offer for weak environmental standards are based on incorrect measurements. Once externalities have been internalized and proper costing has been applied, the standards the Bank advocates are in fact least-cost solutions. It can be difficult to measure the costs of environmental degradation, Serageldin admitted. But even though there are disagreements, it is wrong to assign these costs a zero value simply because no other measurement is available. Small-step approximations are better than nothing.

Because the Bank's borrowers are by definition among the world's poorest and weakest countries, Serageldin continued, it would be wrong for the Bank to impose standards that are not supported by international treaties, followed by other developing countries, or offset by a mechanism like the Global Environment Facility. Current debates about environmental standards reminded Serageldin of the debates about redistribution and growth in the early 1970s. At that time the main issue was whether efforts to reduce poverty were inimical to growth and would detract from the economic focus needed to maximize growth. Then Chenery and others (1974) showed that redistribution and growth are not incompatible. More recently, the World Bank's *World Development Report 1991* concluded that both redistribution and growth are essential, or poverty will hinder development efforts. In ten or fif-

teen years, Serageldin predicted, current debates about environmental standards will be viewed in the same way, because the standards being proposed are quite modest.

Finally, Serageldin said, the World Bank is sometimes criticized by environmental groups for taking a very economic approach to its pollution standards. But in the Bank's view it does not make much sense to invest $100 million in a new plant in, say, Poland that meets the world's highest environmental standards if other plants in the region continue to pollute massively. Instead, efforts should focus on reducing air pollution overall. For example, reducing particulate emissions should be a priority, because doing so has a considerable impact for a limited cost. Reducing emissions of sulfur oxides and nitrogen oxides is much harder, costs a lot more, and has far more limited benefits. But in some cases where the Bank has decided to focus on reducing particulate emissions, it has been criticized for not adhering to U.S. or European standards for other pollutants. The Bank, however, is unwilling to impose standards on developing countries that other countries are unwilling to meet. Rather, we strive to elucidate the issues by developing better methodologies for calculating true costs—for internalizing externalities—and we allow for a range of ways for dealing with pollution concerns where standards do exist.

References

Chenery, Hollis, Montek S. Ahulwalia, Clive Bell, John H. Duloy, and Richard Jolly. 1974. *Redistribution with Growth.* New York: Oxford University Press.

World Bank. 1991. *World Development Report 1991: The Challenge of Development.* New York: Oxford University Press.

Distributors of World Bank Publications

Prices and credit terms vary from country to country. Consult your local distributor before placing an order.

ALBANIA
Adrion Ltd.
Perlat Rexhepi Str.
Pall. 9, Shk. 1, Ap. 4
Tirana
Tel: (42) 274 19; 221 72
Fax: (42) 274 19

ARGENTINA
Oficina del Libro Internacional
Av. Cordoba 1877
1120 Buenos Aires
Tel: (1) 815-8156
Fax: (1) 815-8354

AUSTRALIA, FIJI, PAPUA NEW GUINEA, SOLOMON ISLANDS, VANUATU, AND WESTERN SAMOA
D.A. Information Services
648 Whitehorse Road
Mitcham 3132
Victoria
Tel: (61) 3 9210 7777
Fax: (61) 3 9210 7788
http://www.dadirect.com.au

AUSTRIA
Gerold and Co.
Graben 31
A-1011 Wien
Tel: (1) 533-50-14-0
Fax: (1) 512-47-31-29
http://www.gerold.co.at/online
E-mail: buch@gerold.telecom.at

BANGLADESH
Micro Industries Development Assistance Society (MIDAS)
House 5, Road 16
Dhanmondi R/Area
Dhaka 1209
Tel: (2) 326427
Fax: (2) 811188

BELGIUM
Jean De Lannoy
Av. du Roi 202
1060 Brussels
Tel: (2) 538-5169
Fax: (2) 538-0841

BRAZIL
Publicacões Tecnicas Internacionais Ltda.
Rua Peixoto Gomide, 209
01409 Sao Paulo, SP.
Tel: (11) 259-6644
Fax: (11) 258-6990

CANADA
Renouf Publishing Co. Ltd.
1294 Algoma Road
Ottawa, Ontario K1B 3W8
Tel: 613-741-4333
Fax: 613-741-5439
http://fox.nstn.ca/~renouf
E-mail: renouf@fox.nstn.ca

CHINA
China Financial & Economic Publishing House
8, Da Fo Si Dong Jie
Beijing
Tel: (1) 333-8257
Fax: (1) 401-7365

COLOMBIA
Infoenlace Ltda.
Apartado Aereo 34270
Bogotá D.E.
Tel: (1) 285-2798
Fax: (1) 285-2798

COTE D'IVOIRE
Centre d'Edition et de Diffusion Africaines (CEDA)
04 B.P. 541
Abidjan 04 Plateau
Tel: 225-24-6510
Fax: 225-25-0567

CYPRUS
Center of Applied Research
Cyprus College
6, Diogenes Street, Engomi
P.O. Box 2006
Nicosia
Tel: 244-1730
Fax: 246-2051

CZECH REPUBLIC
National Information Center
prodejna, Konviktska 5
CS – 113 57 Prague 1
Tel: (2) 2422-9433
Fax: (2) 2422-1484
http://www.nis.cz/

DENMARK
SamfundsLitteratur
Rosenoerns Allé 11
DK-1970 Frederiksberg C
Tel: (31)-351942
Fax: (31)-357822

EGYPT, ARAB REPUBLIC OF
Al Ahram
Al Galaa Street
Cairo
Tel: (2) 578-6083
Fax: (2) 578-6833

The Middle East Observer
41, Sherif Street
Cairo
Tel: (2) 393-9732
Fax: (2) 393-9732

FINLAND
Akateeminen Kirjakauppa
P.O. Box 23
FIN-00371 Helsinki
Tel: (0) 12141
Fax: (0) 121-4441
URL: http://booknet.cultnet.fi/aka/

FRANCE
World Bank Publications
66, avenue d'Iéna
75116 Paris
Tel: (1) 40-69-30-56/57
Fax: (1) 40-69-30-68

GERMANY
UNO-Verlag
Poppelsdorfer Allee 55
53115 Bonn
Tel: (228) 212940
Fax: (228) 217492

GREECE
Papasotiriou S.A.
35, Stournara Str.
106 82 Athens
Tel: (1) 364-1826
Fax: (1) 364-8254

HONG KONG, MACAO
Asia 2000 Ltd.
Sales & Circulation Department
Seabird House, unit 1101-02
22-28 Wyndham Street, Central
Hong Kong
Tel: 852 2530-1409
Fax: 852 2526-1107
http://
www.sales@asia2000.com.hk

HUNGARY
Foundation for Market Economy
Dombovari Ut 17-19
H-1117 Budapest
Tel: 36 1 204 2951 or
36 1 204 2948
Fax: 36 1 204 2953

INDIA
Allied Publishers Ltd.
751 Mount Road
Madras - 600 002
Tel: (44) 852-3938
Fax: (44) 852-0649

INDONESIA
Pt. Indira Limited
Jalan Borobudur 20
P.O. Box 181
Jakarta 10320
Tel: (21) 390-4290
Fax: (21) 421-4289

IRAN
Kowkab Publishers
P.O. Box 19575-511
Tehran
Tel: (21) 258-3723
Fax: 98 (21) 258-3723

Ketab Sara Co. Publishers
Khaled Eslamboli Ave.,
6th Street
Kusheh Delafrooz No. 8
Tehran
Tel: 8717819 or 8716104
Fax: 8862479
E-mail: ketab-sara@neda.net.ir

IRELAND
Government Supplies Agency
Oifig an tSoláthair
4-5 Harcourt Road
Dublin 2
Tel: (1) 461-3111
Fax: (1) 475-2670

ISRAEL
Yozmot Literature Ltd.
P.O. Box 56055
Tel Aviv 61560
Tel: (3) 5285-397
Fax: (3) 5285-397

R.O.Y. International
PO Box 13056
Tel Aviv 61130
Tel: (3) 5461423
Fax: (3) 5461442

Palestinian Authority/Middle East
Index Information Services
P.O.B. 19502 Jerusalem
Tel: (2) 271219

ITALY
Licosa Commissionaria Sansoni SPA
Via Duca Di Calabria, 1/1
Casella Postale 552
50125 Firenze
Tel: (55) 645-415
Fax: (55) 641-257

JAMAICA
Ian Randle Publishers Ltd.
206 Old Hope Road
Kingston 6
Tel: 809-927-2085
Fax: 809-977-0243

JAPAN
Eastern Book Service
Hongo 3-Chome,
Bunkyo-ku 113
Tokyo
Tel: (03) 3818-0861
Fax: (03) 3818-0864
http://www.bekkoame.or.jp/~svt-ebs

KENYA
Africa Book Service (E.A.) Ltd.
Quaran House, Mfangano Street
P.O. Box 45245
Nairobi
Tel: (2) 23641
Fax: (2) 330272

KOREA, REPUBLIC OF
Daejon Trading Co. Ltd.
P.O. Box 34
Yeoeida, Seoul
Tel: (2) 785-1631/4
Fax: (2) 784-0315

MALAYSIA
University of Malaya Cooperative
Bookshop, Limited
P.O. Box 1127
Jalan Pantai Baru
59700 Kuala Lumpur
Tel: (3) 756-5000
Fax: (3) 755-4424

MEXICO
INFOTEC
Apartado Postal 22-860
14060 Tlalpan,
Mexico D.F.
Tel: (5) 606-0011
Fax: (5) 624-2822

NETHERLANDS
De Lindeboom/InOr-Publikaties
P.O. Box 202
7480 AE Haaksbergen
Tel: (53) 574-0004
Fax: (53) 572-9296

NEW ZEALAND
EBSCO NZ Ltd.
Private Mail Bag 99914
New Market
Auckland
Tel: (9) 524-8119
Fax: (9) 524-8067

NIGERIA
University Press Limited
Three Crowns Building Jericho
Private Mail Bag 5095
Ibadan
Tel: (22) 41-1356
Fax: (22) 41-2056

NORWAY
Narvesen Information Center
Book Department
P.O. Box 6125 Etterstad
N-0602 Oslo 6
Tel: (22) 57-3300
Fax: (22) 68-1901

PAKISTAN
Mirza Book Agency
65, Shahrah-e-Quaid-e-Azam
P.O. Box No. 729
Lahore 54000
Tel: (42) 7353601
Fax: (42) 7585283

Oxford University Press
5 Bangalore Town
Sharae Faisal
PO Box 13033
Karachi-75350
Tel: (21) 446307
Fax: (21) 454-7640
E-mail:
oup@oup.hi.erum.com.pk

PERU
Editorial Desarrollo SA
Apartado 3824
Lima 1
Tel: (14) 285380
Fax: (14) 286628

PHILIPPINES
International Booksource Center Inc.
Suite 720, Cityland 10
Condominium Tower 2
H.V dela Costa, corner
Valero St.
Makati, Metro Manila
Tel: (2) 817-9676
Fax: (2) 817-1741

POLAND
International Publishing Service
Ul. Piekna 31/37
00-577 Warzawa
Tel: (2) 628-6089
Fax: (2) 621-7255

PORTUGAL
Livraria Portugal
Rua Do Carmo 70-74
1200 Lisbon
Tel: (1) 347-4982
Fax: (1) 347-0264

ROMANIA
Compani De Librarii Bucuresti
S.A.
Str. Lipscani no. 26, sector 3
Bucharest
Tel: (1) 613 9645
Fax: (1) 312 4000

RUSSIAN FEDERATION
Isdatelstvo <Ves Mir>
9a, Kolpachniy Pereulok
Moscow 101831
Tel: (95) 917 87 49
Fax: (95) 917 92 59

SAUDI ARABIA, QATAR
Jarir Book Store
P.O. Box 3196
Riyadh 11471
Tel: (1) 477-3140
Fax: (1) 477-2940

SINGAPORE, TAIWAN, MYANMAR, BRUNEI
Asahgate Publishing Asia
Pacific Pte. Ltd.
41 Kallang Pudding Road #04-03
Golden Wheel Building
Singapore 349316
Tel: (65) 741-5166
Fax: (65) 742-9356

SLOVAK REPUBLIC
Slovart G.T.G. Ltd.
Krupinska 4
PO Box 152
852 99 Bratislava 5
Tel: (7) 839472
Fax: (7) 839485

SOUTH AFRICA, BOTSWANA
For single titles:
Oxford University Press
Southern Africa
P.O. Box 1141
Cape Town 8000
Tel: (21) 45-7266
Fax: (21) 45-7265

For subscription orders:
International Subscription
Service
P.O. Box 41095
Craighall
Johannesburg 2024
Tel: (11) 880-1448
Fax: (11) 880-6248

SPAIN
Mundi-Prensa Libros, S.A.
Castello 37
28001 Madrid
Tel: (1) 431-3399
Fax: (1) 575-3998
http://www.tsai.es/mprensa

Mundi-Prensa Barcelona
Consell de Cent, 391
08009 Barcelona
Tel: (3) 488-3009
Fax: (3) 487-7659

SRI LANKA, THE MALDIVES
Lake House Bookshop
P.O. Box 244
100, Sir Chittampalam A.
Gardiner Mawatha
Colombo 2
Tel: (1) 32105
Fax: (1) 432104

SWEDEN
Fritzes Customer Service
Regeringsgaton 12
S-106 47 Stockholm
Tel: (8) 690 90 90
Fax: (8) 21 47 77

Wennergren-Williams AB
P. O. Box 1305
S-171 25 Solna
Tel: (8) 705-97-50
Fax: (8) 27-00-71

SWITZERLAND
Librairie Payot
Service Institutionnel
Côtes-de-Montbenon 30
1002 Lausanne
Tel: (021)-341-3229
Fax: (021)-341-3235

Van Diermen Editions
Techniques
Ch. de Lacuez 41
CH1807 Blonay
Tel: (021) 943 2673
Fax: (021) 943 3605

TANZANIA
Oxford University Press
Maktaba Street
PO Box 5299
Dar es Salaam
Tel: (51) 29209
Fax (51) 46822

THAILAND
Central Books Distribution
306 Silom Road
Bangkok
Tel: (2) 235-5400
Fax: (2) 237-8321

TRINIDAD & TOBAGO, JAMAICA
Systematics Studies Unit
#9 Watts Street
Curepe
Trinidad, West Indies
Tel: 809-662-5654
Fax: 809-662-5654

UGANDA
Gustro Ltd.
Madhvani Building
PO Box 9997
Plot 16/4 Jinja Rd.
Kampala
Tel/Fax: (41) 254763

UNITED KINGDOM
Microinfo Ltd.
P.O. Box 3
Alton, Hampshire GU34 2PG
England
Tel: (1420) 86848
Fax: (1420) 89889

ZAMBIA
University Bookshop
Great East Road Campus
P.O. Box 32379
Lusaka
Tel: (1) 213221 Ext. 482

ZIMBABWE
Longman Zimbabwe (Pte.)Ltd.
Tourle Road, Ardbennie
P.O. Box ST125
Southerton
Harare
Tel: (4) 6216617
Fax: (4) 621670

Themes for the 1997 Annual World Bank Conference on Development Economics (ABCDE)

May 5–6, 1997

Corruption: Catalysts and Constraints

Incentives and Performance in Public Organizations

Poverty and Environment

Leaders in Growth: Can Others Follow?

Also in this volume:

Comments on "Understanding Financial Crises: A Developing Country Perspective"
Masahiko Aoki
Edward J. Kane

Comments on "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?"
Donald J. Mathieson
Roberto Zahler

Comments on "Political Economy of Alleviating Poverty: Theory and Institutions"
Arsenio M. Balisacan
Karla Hoff

Comments on "Rural Finance in Africa: Institutional Developments and Access for the Poor"
Jean-Philippe Platteau
Dominique van de Walle

Comments on "The Rule of State Law and the Rule-of-Law State: Economic Analysis of the Legal Foundations of Development"
Cheryl W. Gray
Patrick Juillard
Michael J. Trebilcock

Comments on "Contracting, Enforcement, and Efficiency: Economics beyond the Law"
Robert C. Ellickson
Sally Falk Moore

Comments on "International Labor Standards and Trade"
Adolfo Figueroa
Rakesh Mohan

Comments on "Environmental Standards and International Trade"
Maureen L. Cropper
Robert E. Hudec

Cover design by American Writing Corporation

ISBN 0-8213-3786-6