
Regulating Infrastructure

*Monopoly, Contracts,
and Discretion*

José A. Gómez-Ibáñez

Harvard University Press

Cambridge, Massachusetts, and London, England | 2003

Copyright © 2003 by the President and Fellows of Harvard College
All rights reserved
Printed in the United States of America

Library of Congress Cataloging in Publication Data

Gómez-Ibáñez, José A., 1948-
Regulating infrastructure : monopoly, contracts, and discretion /
José Gómez-Ibáñez.

p. cm.

Includes bibliographical references and index.

ISBN 0-674-01177-5 (alk. paper)

1. Contracting out.
2. Privatization.
3. Public contracts.
4. Government ownership. I. Title.

HD3860 G66 2003

363.6—dc21 2003041745

To Nan

Contents

Acknowledgments ix

1. Monopoly as a Contracting Problem 1
 2. The Choice of Regulatory Strategy 18
- ### I Regulatory Politics and Dynamics
3. The Behavior of Regulatory Agencies 39
 4. Capture and Instability: Sri Lanka's Buses and U.S. Telephones 55
 5. Incompleteness and Its Consequences: Argentina's Railroads 84
 6. Forestalling Expropriation: Electricity in the Americas 109

II Contract versus Discretionary Regulation

7. The Evolution of Concession Contracts: Municipal Franchises in North America 157
8. The Rediscovery of Private Contracts: U.S. Railroad and Airline Deregulation *with John R. Meyer* 188
9. Price-Cap Regulation: The British Water Industry 217

III Vertical Unbundling and Regulation

10. The Trade-off in Unbundling: Competition versus Coordination 247
11. Regulating Coordination: British Railroads 264

- 12. Designing Capacity Markets: Electricity in Argentina with *Martín Rodríguez-Pardina* 298
- 13. The Prospects for Unbundling 326
- 14. The Future of Regulation 341

Notes 361

Index 419

Acknowledgments

Many people have taught me about regulation, but I am especially indebted to John R. Meyer, Raymond Vernon, and the faculty and participants in the Infrastructure in a Market Economy executive program at Harvard's Kennedy School of Government. John Meyer has made many contributions to the study of regulation, including being one of the first to document the problems caused by U.S. railroad and airline regulation in the late 1950s. I was a member of the research teams he assembled to study the effects of airline and railroad deregulation in the 1980s, and we coauthored a book on the privatization of transport. Meyer taught me, among other things, to think of regulation as a last resort, given the inherent difficulties of the task. He was among the first to appreciate the potential importance of private contracts as a substitute for government regulation, an idea that plays a key role in this book.

Ray Vernon was more willing than most economists to entertain the idea that there might not be a lasting solution to problems, and that history might offer cautionary lessons for the present. This perspective was illustrated by his argument, first advanced in 1970, that the presence of multinational companies in developing countries was an "obsolescing bargain." The countries initially welcomed the foreigners for the capital, technical expertise, and access to overseas markets they offered. But these advantages declined over time as the locals gained experience. Eventually the natural desire of the countries to control their economic resources and destiny reasserted itself. When the movement to privatize infrastructure swept the developing world in the 1990s, Vernon was one of the first to worry that it would prove to be another example of an obsolescing bargain. He was no fan of public enterprises, having worked for many years to improve their performance without much success. But he encouraged me to read the histories of the firms that were being privatized to understand why they had once been nationalized, and to think about the risk that nationalization might happen again. Chapters 6 and 7 particularly are a product of this prodding.

My interest in regulation would not have grown into a book without the challenge of developing programs on infrastructure regulation for senior executives. This project grew out of concern among faculty at Harvard and staff at the World Bank and the Inter-American Development Bank (IDB) that countries were privatizing very fast, without adequate thought given to the regulatory institutions that were critical to long-term success. The programs brought together faculty interested in the topic and, equally important, senior officials from around the world who were involved in privatizing, regulating, financing, and managing infrastructure. The sessions were taught by the case method, and the discussions served as a laboratory to test whether the faculty's concepts and arguments were persuasive and helpful to men and women with practical responsibility in the field. Many of the examples presented in this book began as teaching cases in these executive programs.

At the Kennedy School I owe a particular debt to Henry Lee and Kathy Eckroad for being my partners in developing the Infrastructure in a Market Economy executive program, and to Ashley Brown, Jack Donahue, Alexander Dyck, Willis Emons, Bill Hogan, David Luberoff, John Meyer, Ray Vernon, Jay Walder, Lou Wells, and the other members of the 1997-98 faculty seminar on infrastructure privatization and regulation for their stimulating discussions of the concepts and teaching cases we used in the program. Officers at the Inter-American Development Bank, particularly Terry Powers, Faith Wheeler, and Jim Wylde, provided critical early support by funding the development of cases and by sending some of the best IDB staff to attend and contribute. Later the IDB, led by Roberto Manrique, supported the creation of the Latin American University Regulation and Infrastructure Network (LAURIN), an association of eight universities (including Harvard) involved in research and executive training on private utilities. At the World Bank I owe a special debt to Antonio Estache, who designed the bank's executive program for transport regulators, invited me to teach in it, supported the development of cases used at both the bank and the Kennedy School, and has been a friend and advisor ever since.

The cases would not have been possible without assistance from many people. At the risk of some unfairness, I would like to single out Ian Byatt, who explained the pressures on him as Britain's water regulator; John Diandas, with whom I enjoyed spirited arguments about Sri Lanka's buses; Stephen Glaister, who guided me through the British railway industry; Jorge Kogan, the architect of transport reform in Argentina; and Martín Rodríguez-Pardina, who helped me understand electricity in Latin America. Alan Altshuler, my colleague at Harvard, was especially helpful in advising me on the relevant political science literature. Thanks also to the anonymous referees enlisted by Harvard University Press and to Michael Aronson and Eliza-

beth Gilbert of the Press for the many improvements they made to the manuscript. Doctoral students are the lifeblood of academia, and I learned a great deal from my conversations with Christoph Meier, Christopher Shugart, and Sashi Verma while they were working on their dissertations on regulation. I began this book in the fall of 1998 while on sabbatical at the University of California Transportation Center at Berkeley, where Martin Wachs and Melvin Webber were my gracious hosts. Thanks also to the Brookings Institution for permission to publish Chapter 8, which appeared in an earlier form (coauthored with John R. Meyer) as "Government and Markets in Transport: The U.S. Experience with Deregulation," in *Governance amidst Bigger, Better Markets*, edited by John D. Donahue and Joseph S. Nye, Jr. (Washington, D.C.: Brookings Institution, 2001). Obviously none of those named is responsible for the errors I have made.

Finally, I would like to thank my wife, Nanette Wilson, who has always supported this effort, even though I was often distracted by it. My mother-in-law, observing the painful process, concluded that it was much better to read books than to write them. I plan to take her advice, at least for a while.

Regulating Infrastructure

Monopoly as a Contracting Problem

The Search for Commitment

Interest in the regulation of private infrastructure increased during the last two decades of the twentieth century, when many countries turned to private companies to build and operate infrastructure and utility services. In the 1980s Britain was a leader, selling off its telephone, electricity, gas, water, and railway companies in the hopes that the private sector could provide better service at a lower cost. In the 1990s many other countries followed suit, particularly in Latin America but also in Southeast Asia, Austral-Asia, and Europe. For example, almost all of the new high-performance expressways and many of the new power plants opened in developing countries in the 1990s were built by private concessionaires. By the end of the century, most of the major railways and telephone companies and many of the electric companies in Latin America had been sold or offered as concessions to private operators as well.

One fear is that this latest round of privatization will not last. By the first years of the twenty-first century a backlash of sorts had already developed. Investors in many of the newly privatized utilities were voicing disappointment about the returns that they were earning. Meanwhile, consumers were increasingly skeptical that the tariffs they were being charged were fair and the service adequate. Several well-publicized failures—such as the collapse of California's private wholesale electricity market in 2000 and the bankruptcy of Britain's private rail infrastructure company in 2001—suggested to many that privatization had gone too far.

History cautions that this disenchantment may grow to the point where governments begin to take back the companies. Private provision of infrastructure was the norm throughout the world during the first half of the twentieth century, but this era ended with most private infrastructure companies being bought out or expropriated by government. Many of the utilities that Britain privatized in the 1980s had been nationalized only in the 1940s

and 1950s, for example, while many Latin American companies privatized in the 1990s had been nationalized as recently as the 1960s and 1970s.

The United States managed to retain a significant private presence in infrastructure throughout the twentieth century, which suggests that privatization can be a long-term solution in some circumstances. France also maintained private provision of certain municipal utilities, most notably water and solid waste, although it nationalized its electricity, telephone, and railway companies. But the U.S. and French experiences with private infrastructure have not been wholly satisfactory either. Concern about corruption in the award of private infrastructure concessions increased in France during the 1990s, for example, after French municipalities were granted more autonomy. And in the 1970s and 1980s, while other countries were beginning to shift to private but government-regulated utilities, the United States became so dissatisfied with that strategy that it deregulated its private railroads, airlines, long-distance telephone services, and natural gas pipelines.

Fifty years from now we may look back to view private provision of infrastructure as the norm, and public provision as a failed experiment of the mid-twentieth century. Indeed, the U.S. movement to deregulate suggests that the ultimate goal in infrastructure may be to dispense not just with public provision but, where possible, with public regulation as well. Nevertheless, it is striking that the United States was the only country that was able to maintain private ownership of most of its utilities throughout the twentieth century. In most other cases, the private providers were eventually taken over by government.

Many proponents of private infrastructure view the problem as one of establishing a commitment to a fair and stable set of rules governing the relationship between the government and private infrastructure providers. The usual concern is that the government will renege on commitments to private infrastructure rather than vice versa. Private companies are vulnerable because infrastructure requires expensive, durable, and immobile investments in roadways, power plants, local telephone or water lines, or other facilities that private investors can't withdraw if the government changes the rules.

In this view, the government usually feels compelled to regulate the prices and quality of infrastructure services because the services are essential to modern life and because infrastructure has elements of monopoly, so that consumers are at the mercy of a single provider. To induce private investment in infrastructure, the government also must commit to a schedule of tariffs and other terms that give the investors a reasonable return. Once the investments are in place, however, the government will be tempted to yield to popular pressures to lower tariffs or renege on other promises to the companies, knowing that they can't retaliate by withdrawing their investments. A gov-

ernment that engages in such opportunistic behavior is likely to have trouble convincing investors to renew or expand their facilities. If the facilities are long lived, however, the government may hope that some other solution will emerge in the interim, such as a crop of new investors naive enough to overlook a history of broken promises.

This account is one-sided, however, in that it focuses on the company's vulnerability to opportunism by government and downplays government and consumer vulnerability to opportunism by the company. The government and consumers are vulnerable because the incumbent local infrastructure company usually enjoys a monopoly, and that monopoly stems largely from the durability and immobility of both the company's and the consumers' investments. The company's durable and immobile investments discourage competitors from entering the market to challenge the incumbent. Potential challengers realize that an incumbent, fearful of losing the market, could rationally drop prices so that they just cover its short-run variable costs, leaving little chance for the challenger to recover its investments. And consumers also typically make durable and immobile investments in their local communities that make it difficult for them to move elsewhere in search of cheaper infrastructure services. In these circumstances, the customers of the infrastructure company, or the government acting on the customers' behalf, will want some commitment that the local infrastructure company will not take advantage of its position by raising prices well above costs. In short, the expensive, durable and immobile investments help make all parties—the company, its customers, and the government—vulnerable to opportunism and desirous of stability and commitment.

The prospects for commitment are complicated, however, by the long lives of most infrastructure investments. In the case of a highway, for example, the paving may last ten or twenty years, the bridges and the sub-base for forty years or more, and the basic grading and right of way indefinitely. How realistic is the desire for commitments that last as long as infrastructure assets, particularly in a world where the infrastructure companies, their customers, and governments have complex needs that may change in ways that are difficult to anticipate? What are the basic forms that commitment can take? And how can one design schemes that balance the desire for commitment with the need for flexibility to accommodate unexpected developments?

The basic perspective of this book is that the problem of infrastructure monopoly is similar to any other long-term contracting problem, and particularly analogous to contracting in private sector procurement. Infrastructure is not the only sector of the economy that employs assets so durable and specialized that their suppliers and users are vulnerable to each other. The developer of a shopping center and the lead tenant are often vulnerable to each

other, for example, as are the power plant and the coal mine that supplies it. The usual remedy in other sectors is to sign a long-term contract before such specialized investments are made. One hundred percent of a new shopping center need not be preleased, but a significant portion is usually required before financing is forthcoming. And most lead tenants would never move in without the protection of a multiyear lease.

From this perspective, monopoly does not necessarily require government intervention. Just as private long-term contracts can protect suppliers and users of other durable assets, they may be able to protect utility companies and their customers as well. Whether government regulation is necessary and the form it should take will depend fundamentally on how difficult it is to negotiate and enforce an explicit long-term contract. Before we explore these issues, it is helpful to define infrastructure and review the various reasons, including monopoly, why government is so often involved in its provision.

Motives for Government Involvement in Infrastructure

Infrastructure has special characteristics that have traditionally justified or encouraged government involvement. Infrastructure means beneath (*infra*) the building (*structure*), and thus usually encompasses services or facilities that are underground, such as piped water and sewerage, or that lie on the surface, such as roads and railways. Electric power and telecommunications are often included as well, even though they are frequently provided by lines strung on poles or towers rather than in underground conduits. All of these industries involve networks that distribute products or services over geographic space, and in most cases the networks are capital extensive and the investments are durable and immobile. Infrastructure industries are often called public utilities, and the two terms will be used interchangeably here.

One motivation for government involvement, and the primary focus of this book, is the tendency toward monopoly in infrastructure industries. This tendency arises because many infrastructure networks have the characteristics of a so-called natural monopoly, which are a combination of durable and immobile investments and strong economies of scale or traffic density. The economies of scale mean that the cheapest way to serve a community is with a single company, particularly if the local network has a relatively low density of traffic. And the durability and immobility of the investments increase the risk for new entrants who seek to challenge the incumbent. Concern over monopoly often leads the government either to provide infrastructure service itself or to regulate the prices and quality of service of private infrastructure companies.

A second motive for government involvement—and one that often pre-

dates and enhances concerns about monopoly—is the difficulty of assembling the right of way required for an infrastructure network. Railroads, highways, and power, water, and telephone lines all require long, linear, and contiguous rights-of-way that would be difficult to assemble without the government's power to expropriate private property through the process of eminent domain. Absent the threat of eminent domain, private landowners along the alignment could extort high prices for key or missing parcels. The government often exercises the power of eminent domain on behalf of infrastructure companies, or allows the companies to place their pipes or lines in local streets or other publicly owned rights of way. Governments are hesitant to delegate eminent domain powers to private companies or to grant companies unrestricted access to local streets, however, for fear these privileges will be abused. Moreover, the understandable reluctance of the government to expropriate property for new infrastructure rights of way can contribute to monopoly by making it harder for a new company to enter the business and challenge the incumbent.

A third rationale for government involvement is that some types of infrastructure generate benefits beyond those that accrue to its immediate users or subscribers. For example, clean drinking water and sanitary waste disposal protect the general public from the spread of disease and the contamination of the environment. Similarly, a lamp outside a private residence or business reduces the risk of accidents and crime for neighboring properties and passersby. If important benefits of infrastructure services accrue to nonsubscribers, then it may be difficult to persuade subscribers to pay voluntarily for the level of service that is socially desirable. Nonuser benefits have stimulated governments to promote infrastructure provision in a variety of ways. Most city governments contract directly with electricity companies for street lighting, for example, and compel all households and businesses to subscribe to piped water and sewerage services.

Economic development and equity considerations are two additional and related motives for government intervention, and they also often enhance monopoly concerns. Infrastructure is viewed as an important ingredient to local economic development, and thus governments are often concerned about the infrastructure endowments of lagging or underdeveloped regions of their countries. Similarly, ensuring universal access to a basic level of infrastructure services is often thought to be important to the protection of equal opportunity for individual citizens, much as universal access to basic education and basic health care is. Infrastructure may not be deemed as essential to development and equal opportunity as education or health, but it is often just behind.¹

These developmental and equity considerations have led governments to

encourage the development of more extensive infrastructure networks than can be financed with the tariffs that users are willing to pay. Many governments have created special programs to support rural electrification, telephones, and roads, for example, in the belief that such important services ought to be available throughout the country. Similarly, many countries try to keep the tariffs for basic levels of service low so that even poor households can afford piped water, electricity, and a telephone. Developmental and equity considerations also have led some governments to promote monopoly and limit competition, so as to allow the infrastructure company to charge some customers prices in excess of costs and use the proceeds to cross-subsidize low tariffs in rural areas or for poor households.

A final motivation for government involvement in infrastructure is to reduce safety and environmental problems. Railroads, highways, and power lines present safety hazards to both users and nonusers, for example, while power plants, locomotives, and motor vehicles pollute the environment. To the extent that these risks fall on nonusers, the government often feels justified in regulating the harms on the public's behalf. And even if the safety and health risks fall on users, government intervention may be warranted if users are not well enough informed to judge the hazards that they are being exposed to. In most countries, the regulators in charge of safety and environmental concerns are separate from those responsible for controlling monopoly. The separation is designed to avoid any potential conflict of interest between setting tariffs and setting health and safety standards.

The Nature of Monopoly

Sources of Market Power

When one is assessing the degree of competition that a firm faces, the concept of market power is more helpful than that of monopoly. Monopoly is defined as a single seller serving a market. Market power is usually defined as the degree to which a company can raise the prices for its products above its costs without losing too many sales. Monopoly can be misleading, because the presence of only a single seller is neither a necessary nor a sufficient condition for effective market power. Even if there is only one seller in a market, for example, that firm may not be able to charge prices above costs if it believes that doing so will simply invite many other firms to enter the market and compete with it. Similarly, even if there are several sellers in a market, they may find ways to collude so as to effectively inhibit competition.

Two conditions are necessary for a firm to have market power. The first is

the presence of some type of barrier that prevents other firms from entering the market to provide competing services. Barriers to entry can be either created by governments or firms or inherent in the technology of the industry. Examples of created barriers are the patent protections that governments award to inventors and the brand loyalty that some firms attempt to develop through extensive advertising. The primary example of an inherent barrier is the combination of large economies of scale and durable and immobile investments commonly referred to as a natural monopoly.

The second condition is that there must be few close substitutes to the good or service in question. Even if there are barriers to entry, the firm serving the market will not have much power over its customers if they can find close substitutes to the services it provides. These substitutes might be similar goods or services or alternative locations where the identical good is produced or sold. The idea is to avoid defining the market too narrowly by overlooking the competition provided by alternative products and sources of supply.

To illustrate these two conditions, consider local buses, which provide an example of a service that is likely to be highly competitive. Buses are mobile and not very durable, and studies show that economies of scale in local bus services are exhausted with fleets of twenty-five to fifty buses. Thus it is usually possible to have several firms serve a single city, or a single corridor within a city. Moreover, the private automobile usually competes with the bus in high-income countries, while walking, bicycles, and motorcycles often provide wide competition in low-income countries and where climates are mild.

At the other extreme, the distribution of piped water to residential neighborhoods is usually a classic natural monopoly. Underground pipes last fifty years or more, and there are strong economies of scale because it is more economical to serve all the households on a street from a single pipe than from two or three competing parallel pipes. One pipe is cheaper because the cost of digging and back-filling the trench for the pipe and the cost of the pipe itself do not increase proportionately with the pipe's capacity. Moreover, the alternatives to piped water—such as private wells, tanker trucks, or bottled water—are usually more expensive and less convenient. As a result, a local piped water company often faces little effective competition and could price its services well above costs.

Within any given infrastructure industry, market power often varies according to the types of customers involved or the specific circumstances of the firm. Freight railroads typically have less market power over shippers of high-value, manufactured commodities than they do over shippers of low-value, bulk commodities, for example, since trucks are a more viable alterna-

tive for manufactured than for bulk commodities, especially over short distances. Similarly, a railroad may have limited market power even over shippers of bulk commodities if it competes with a navigable waterway.

The degree of market power can also vary among the different stages or components of an infrastructure service. Many countries recently have broken up or unbundled their electric utilities into separate companies for generation, long-distance transmission, and local distribution, because electricity generation is potentially competitive while transmission and distribution are not. Similarly, wireless, long-distance, and local telephone services are now often provided by separate companies, because competition is possible in long-distance and wireless services but more difficult in local hard-wire services.

In addition, market power varies over time as markets and technologies evolve. The potential for competition in electricity generation increased in the last several decades after technological advances reduced the minimum size of a cost-efficient power plant. Competition in long-distance telephony was made possible by the development of microwave, satellite, and fiber optic transmission technologies. Competition in local telephony may become more effective in the future if wireless technologies become more cost-competitive with the hard-wire local telephone line.

Natural Monopoly and Long-Term Contracts

Natural monopoly is the most important form of barrier to entry in infrastructure. Created or artificial barriers are common as well, particularly those established by governments reluctant to exercise powers of eminent domain to create new infrastructure rights of way or desirous of protecting cross-subsidies that support low tariffs for poor customers or remote sections of the network. But natural monopoly is the most common barrier, and harder to avoid since it is inherent in the technology.

Most descriptions of natural monopoly stress the importance of large economies of scale and underplay the role of durable and immobile investments in establishing the barrier to entry. Economies of scale occur when average or unit costs of a firm fall as volume increases. If the economies extend over volumes sufficient to serve the entire market, then it will be cheaper for a single firm to serve the market rather than two or more. And since a larger firm always has a cost advantage over a smaller one, only one firm is likely to serve the market at any given time.

In telecommunications and transportation industries, economies of scale are sometimes confused with network economies. In network economies, an infrastructure network is more valuable to each of its subscribers the more

other subscribers it allows them to connect with. This implies that larger networks will have an advantage over smaller networks in recruiting new subscribers, and that the most valuable network of all is one that allows the entire universe of subscribers to communicate with one another. Network economies mean that a universal interconnected network is likely to emerge because of the advantages it offers subscribers, but it does not mean that the network inevitably will be operated by a single firm. That occurs only if there are economies of scale in the supply of networks such that it is cheaper for one firm to provide a universal network than for several independent or interconnected firms to do so.²

Despite the emphasis traditionally placed on them, economies of scale are arguably less important than durable and immobile investments in establishing the barrier to entry in natural monopoly. Economies of scale mean that only one firm will serve the market at any given time, but they don't prevent "hit and run" competition. If investments are short-lived or mobile, then the threat of entry by a challenger is still credible, because both the incumbent and the challenger have the option of exiting the market and taking their investments with them. If investments are durable and immobile, by contrast, then neither firm can exit the market without losing its investments. The most likely result of entry is a punishing price war that ends with either the challenger or the incumbent losing its investments. And if capital costs account for a large portion of total costs and the facilities are long-lived, an incumbent is likely to drop prices sharply and keep them low for a long time until the challenger is driven off.³ Indeed, some economists argue that durable and immobile investments, not economies of scale, are the defining characteristic of infrastructure monopolies.⁴

An effective monopoly in local infrastructure depends on the customers, as well as the company, making durable and immobile investments. The customers make their durable and immobile investments when they establish their residences and businesses in the territory served by the infrastructure company. These investments include the time a family must spend to find a suitable local home, job, and schools for the children, for example, or the resources a business devotes to developing a local workforce or customer base. The households and businesses could conceivably move to another community if their local infrastructure company increased prices. The costs of replacing their local and immobile investments are so high, however, that relocation is seldom a realistic option.

Seen in this way, the problem of an infrastructure monopoly is a variant of a larger and more common problem of relationship-specific investments in procurement.⁵ Often a supplier and a customer can reduce their costs by making investments specific to their relationship. A parts manufacturer and

an automobile company it supplies might be able to cut costs by, for example, investing in specialized machinery or training or by moving their plants closer together. If the investments are durable and specific to the relationship, however, then the party that makes them becomes vulnerable to opportunistic behavior by the other. Once the parts manufacturer invests in specialized machinery, the auto company can demand lower prices, knowing that the parts manufacturer has no other use for the machines. There are a variety of methods to allow suppliers and customers to make relationship-specific investments by protecting them from opportunism, but one of the most common is the long-term contract. In our auto parts example, neither the parts manufacturer nor the automobile company is likely to invest in specialized machinery without a contract guaranteeing certain prices and quantities.

The infrastructure natural monopoly is a variant of the procurement problem in which relationship-specific investments are not optional but are an inevitable consequence of the technology. It is hard for an infrastructure company to avoid investments in durable and immobile facilities or for its customers to avoid investments specific to their communities. The company's facilities cannot be transported to other locations if the local customers, or the government representing those customers, insist on price reductions. And the customers can't easily relocate if the infrastructure company decides to raise prices.

As in the more general procurement problem, one way to protect against opportunism is to sign a long-term contract before making the relationship-specific investments. And as in the general procurement problem, these contracts need not involve the government: they can be private contracts between the infrastructure company and individual customers. Private contracts are not far-fetched, particularly for customers that purchase such large quantities of infrastructure services that it is worthwhile to spend the time and resources needed to negotiate a contract. A manufacturer might find it worthwhile to negotiate long-term rates with the local railroad, power, or water company, for example, before it invested in a plant in their community. Similarly, a local railroad or water company might insist on a long-term contract before it built a special spur or a major water main to the manufacturer's plant.

Government regulation can be viewed as a substitute for private contracts, used when it is too costly or difficult for the companies and customers to reach individual agreements. In effect, the government contracts with the infrastructure company on the customers' behalf. In some cases the agreement between the government and the companies takes the form of an explicit contract, as in the franchise or concession contracts that governments often

grant infrastructure companies. In other cases the understanding is less detailed and more flexible.

The Range of Solutions to Monopoly

The solutions to monopoly can be arrayed along a continuum according to the relative roles that markets and politics play in determining infrastructure prices and service quality, as shown in Figure 1.1. At one extreme prices and quality are determined largely or entirely by markets, at the other extreme largely by politics, and in between by a mixture of the two. Along the continuum from markets to politics the basis for commitment gradually shifts from commercial contracts and courts, to specialized regulatory institutions, and finally to the broader social and political institutions and interests that shape public policy.

There are many variants and hybrids along the continuum, but most can be assigned to one of four main groups or categories. *Private contracts* between

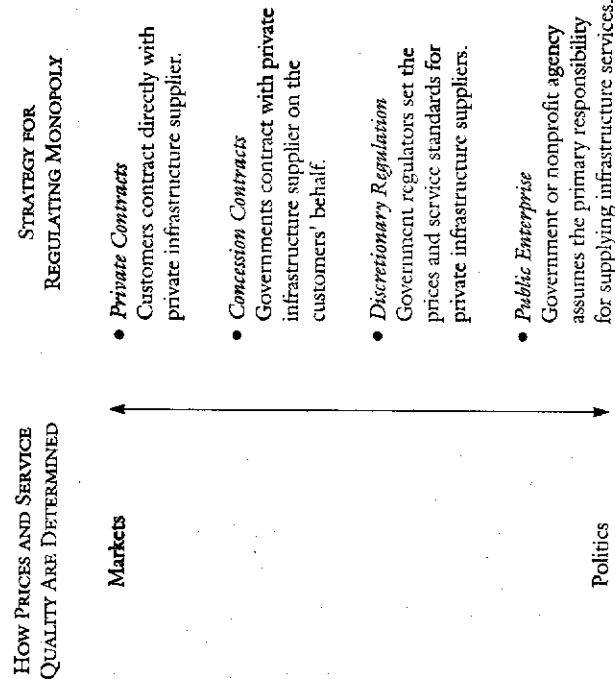


Figure 1.1 The range of solutions to monopoly.

infrastructure companies and customers is the category in which the market plays the largest role and politics the smallest. Politics is involved because the legislature and the courts enact and enforce the laws that govern private commercial contracts. But within that general legal framework, infrastructure suppliers and their customers negotiate voluntary agreements about the prices to be paid and the quantities and qualities of services to be provided.

The next alternative is *concession contracts*, which were popular in many parts of the world in the nineteenth century and were rediscovered and improved in the late twentieth century. Under this scheme the government awards a private firm a concession or franchise to provide a specific infrastructure service for a limited period of time, typically ten or twenty years. The services the concessionaire must provide and the maximum tariffs it can charge are specified in the concession contract. The contract is usually awarded competitively, often to the bidder proposing the lowest tariff for a specified level of service. The government monitors the concessionaire's performance to make sure it is in compliance with the contract, but neither it nor the concessionaire is supposed to unilaterally change the contract after it is awarded.

Concession contracts increase the role of politics by substituting contracts between the private suppliers and governments for contracts between the suppliers and individual customers. In effect the government represents consumers in deciding what combination of price and service quality would be best. Market forces still play a major role, however, because the concessionaire signs the contract voluntarily and the contract is not supposed to be changed subsequently without his agreement. If the contract is awarded competitively, moreover, consumers have some assurance that the concessionaire's expected profits will not be excessive relative to those of firms in competitive markets.

The third alternative is *discretionary regulation*, which is used extensively in the United States and Britain and has been applied selectively in many developing countries. The discretionary approach involves creating a government regulatory agency with the power to unilaterally establish the infrastructure firm's tariffs and service standards. The legislation establishing the regulatory agency usually sets out the principles the agency must consider when making its decisions. And the agency's decisions can usually be appealed to the courts or some other tribunal on the grounds that it has not followed its statutory guidance. The principles and guidelines are stated in fairly broad terms, however, in contrast to the highly specific language of a concession contract. The broad language gives the regulator substantial discretion and flexibility to respond to novel or unforeseen circumstances.

The discretionary approach increases the role of politics further by aban-

doning the effort to describe all the commitments between the government and the infrastructure company in an explicit contract. The market is still importantly involved, however, because the regulated firm depends entirely or in part on private capital markets to finance its investments. If the regulatory agency's decisions are too harsh, the company will not be able to raise capital to replace worn-out facilities or to accommodate growth in demand. But to the extent that the assets are durable and provide adequate capacity, the consequences of harsh decisions may not be apparent for many years.

The final option is to give the responsibility for providing the infrastructure services to a *public or nonprofit enterprise*. Private firms might still be involved as contractors to the public agency, but the contracts would be more limited in scope or duration than a concession. And markets are still influential, since the public agency's revenues are affected by the prices consumers are willing to pay for its services while the agency's costs are affected by the prices it has to pay in the labor, equipment, materials, and other input markets. The main difference is that the enterprise responsible for the service is not owned and controlled by private investors, so that it may have less incentive to take advantage of its monopoly position by charging prices above costs.

The Plan of This Book

This volume explores the advantages and disadvantages of these regulatory options and various hybrids in different industries and countries. Almost all the attention is devoted to the first three options: private contracts, concession contracts, and discretionary regulation. While the choice between private and public provision overshadows the entire discussion, it is a complex topic that has been studied by many others.⁶ This volume assumes that private provision of infrastructure is generally desirable, particularly if the problems of regulating monopoly can be solved in a politically acceptable and economically sensible way. The main question is what those acceptable and sensible solutions might be.

The book draws on the existing literature on regulatory economics and politics to set out the basic regulatory options and then explores how these options work in the real world by examining specific cases in which they have been applied. The case approach is critical, because the key questions are empirical. Each regulatory strategy has its strengths and limitations in theory, but how powerful are these strengths and limitations in practice? Cases are also valuable because they allow one to examine how the interplay of economics, politics, and institutions affects regulatory commitment and performance. Some researchers have tried to examine this interplay using large data

sets, and their results are reported here. Few such data sets exist, however, and they capture only a portion of the many variables that might be involved. To appreciate how regulation works in practice, there is simply no substitute for examining the evolution of a specific concession contract or the history of a particular regulatory agency.

The undeniable danger of using cases is that the selection may be so unrepresentative that the generalizations developed from it are misleading. The cases used here are drawn from a wide variety of infrastructure industries, including electricity, railways, telecommunications, and water. The cases also cross a considerable time span: from the middle of the nineteenth century to the beginning of the twenty-first century. The selection of countries is considerably narrower, however: most cases are based in the United States, Britain, or Latin America. The United States is of special interest because many of its private utilities were never bought out or expropriated by the public sector. U.S. cases provide the opportunity to consider why this country was an exception and how regulatory systems evolved over long periods of time. Britain and Latin America are heavily represented because they were pioneers in the latest wave of privatization, and the sources of many innovations in regulation.

The book is divided into three parts. It begins with two introductory chapters, this one and the next, which set out the principal regulatory options and their pros and cons, drawing on the literature on transaction costs and institutional economics to provide a basic framework. The next three parts are designed to be read together but can be read selectively as well, according to the reader's interests.

Part I, Chapters 3–6, focuses on the politics and dynamics of regulation. The options and framework introduced in this chapter and Chapter 2 are basically static, but regulatory systems change over time and some may be inherently unstable. Chapter 3 examines the behavior of regulatory agencies, focusing particularly on the economic, political, and institutional factors thought to influence the risk that a regulatory agency will be “captured” by special interests. Chapter 4 examines how capture can undermine a discretionary regulatory system. Two common forms of unstable capture are illustrated: one with the case of bus regulation in Sri Lanka and the second with telephone regulation in the United States. Chapter 5 examines the dynamics of concession contracts, using Argentina’s railways as the case. Finally, Chapter 6 considers the circumstances that discourage government expropriation of private utilities by examining the history of the electricity industry in North and South America. The obvious question is why almost all of the private electric companies in South America and two-thirds of those in Canada were expropriated or bought out by 1970, while very few private electric companies in the United States suffered the same fate.

Part II examines the pros and cons of the three main regulatory options in more detail. Chapter 7 considers the strengths and limitations of concession contracts by considering the rise, fall, and revival of municipal franchises in the United States and Canada in the nineteenth and twentieth centuries. Chapter 8 considers the potential for private contracts to substitute for public regulation by examining the deregulation of the U.S. railroad and airline industries. Private contracts were a deliberate and central part of the plan to deregulate the railroads and an unplanned but surprisingly important development in the deregulation of the airlines. Chapter 9 examines price-cap regulation, the modern form of discretionary regulation pioneered in Britain in the 1980s and adopted by many other countries since. The advantages of price cap over older forms of discretionary regulation are explored through the experience of Britain’s water industry.

Part III focuses on the important experiments with the vertical unbundling or restructuring of utilities. The idea of unbundling is to separate the activities that are potentially competitive from those that are not, so that regulation can be confined to those activities where it is absolutely necessary. Vertical unbundling threatens to reduce the coordination between the formerly integrated activities, however, since this coordination now must be achieved through contracts between separate firms. The coordination is further complicated because these contracts may have to be supervised by a regulator, if one of the parties enjoys a monopoly. Chapter 10 considers the options for achieving this coordination, including the possibility of private contracts negotiated without the supervision of a regulator. Chapter 11 examines the problems regulators face in setting tariffs for the remaining monopoly suppliers in the industry, using Britain’s railroads as the case study. Chapter 12 explores the possibility that markets for infrastructure capacity could be developed to alleviate coordination problems by considering the experience of Argentina’s electricity industry. The part concludes with Chapter 13, which compares the potential for vertical unbundling across the main infrastructure industries.

The book’s concluding chapter speculates about the future of regulation, including the ways in which the private contract, concession contract, and discretionary approaches might evolve.

Commitment and Flexibility

Three main themes emerge from this survey. The first is that the market-oriented and contractual solutions to monopoly are preferable where they are practical. All things being equal (which they seldom are), private contracts are better than concession contracts and concession contracts are better than discretionary regulation. One reason for this ranking is that the stronger the

exposure to market forces, the greater the incentives to improve services and reduce costs. Private contracts are better than concession contracts, for example, because the private contract directly involves the consumer in the design of her service while the concession contract requires that the government design the service offerings on the consumer's behalf. Another reason for the ranking is that contracts enforced through the normal commercial courts usually provide a clearer and stronger form of commitment than specialized regulatory institutions. Our understanding of how to design specialized regulatory institutions is limited, and experience strongly suggests that they eventually fall prey to tendencies and forces inimical to the long-term interests of the regulated industry and its customers.

It is striking how often the market-oriented or contractual solutions can be applied. The potential for private contracts to alleviate monopoly is suggested by how common relationship-specific investments seem to be throughout private industry. Private contracts may be more practical where the quantities purchased are large enough to make the negotiation and enforcement of a contract worthwhile. But the experience of the American airline industry recounted in Chapter 8 suggests that private contracts sometimes can be developed for customers purchasing relatively small quantities.

The second theme, which qualifies the first, is the importance of being realistic about the level of commitment that is possible, even through contracts. The world often changes in unexpected ways to undermine commitments. And commitments that impose large and unforeseen costs on either party are hard to sustain, no matter how sincere the parties were at the outset. Less specific and more flexible arrangements pose their own problems, of course, by exposing the firm, its customers, and the government to the risk of opportunistic behavior. But if it is impossible to draft a commitment that has a reasonable chance of surviving, then the parties will be exposed to opportunism anyway when the agreement collapses.

The degree of commitment possible varies from one type of industry and situation to another. It is easier to write a complete and workable contract, for example, where the investments are not very durable and somewhat mobile, or where the technological, economic, and political environments are relatively stable. The limitations of commitment also make hybrid strategies—combining some of the specificity of a contract with some of the flexibility of discretion—more interesting. It may be desirable to build some elements of discretionary regulation into a contract, for example, so as to limit the scope for opportunism if the contract has to be renegotiated. The limitations of commitment also make extra-contractual strategies more attractive, such as promoting widespread stock ownership to increase popular support for the firm.

The final theme, which is a corollary of the second, is that regulatory schemes are bound to change, especially in the long run. Very specific commitments, such as contracts, are vulnerable to unanticipated changes. More flexible commitments, as in discretionary regulation, may be inherently unstable because of the tendency of discretionary regulators to be captured by special interests or to ossify and lose their flexibility over time. Many utility policy reforms, like vertical unbundling, are experimental in the sense that they are adopted before their implications are fully understood. And the choice among private contracts, concessions, and discretionary regulation may change as technology changes and legal, social, and political institutions evolve.

The inevitability of change means that it may be reasonable to adopt a scheme that is unlikely to last—if it is still the best of those currently feasible. In the space of 150 years, the American railroad industry was regulated first through a regime of private contracts, then through concession systems, later discretionary regulation, and finally private contracts again. During any particular regime, moreover, the rules of the game typically changed in relatively important ways every decade or so. Each scheme was arguably the most appropriate at the time it was adopted. The changes had their costs as well as their benefits, but if there was an error it was probably a tendency to change too slowly, the most obvious example being the maintenance of discretionary regulation for several decades after it had ceased to be useful.

Nevertheless, it is important to choose the regulatory scheme carefully if private infrastructure is to survive. This means relying on private and contractual solutions where practical, since they generally increase the level of commitment and the chances that consumers will get the infrastructure services they value. But it also means being realistic about when private or contractual solutions will work, adopting discretionary schemes where necessary, or not privatizing at all where no regulatory scheme seems workable.

The Choice of Regulatory Strategy

Markets, Politics, and Values

The choice of whether and how the government should regulate monopoly depends in part, if not fundamentally, on the values or goals one considers important. What criteria should one use for judging the performance of alternative remedies to monopoly?

One obvious response is to use the values embodied in the competitive market as the standard. Monopoly is defined as a lack of competition, and the corrective implied is to make the market behave as if it were competitive. The values at the heart of the market are individualistic and utilitarian. Free markets allow individuals to engage in voluntary exchanges of goods and services. Subject to certain conditions to be discussed later, the voluntary nature of the exchange ensures that the parties regard themselves as better off than they were before. Otherwise, the parties would not agree to the exchange. From that perspective, the role of society is to encourage individual happiness by facilitating voluntary transactions, and smoothly functioning, competitive markets are a primary vehicle for doing so.

The individualistic values embodied in the market enjoy wide popular support in most developed and market-oriented societies, but there are other values that are widely shared as well. In particular, support for individualism is often combined with a concern that the inequities in society should not be too large. People disagree about the definition of equity and the best ways to protect it, but many believe that society should ensure that individuals start out on a relatively equal footing by providing them universal access to basic education and health care. Such equity concerns are particularly salient for infrastructure, moreover, since, as noted in the previous chapter, many people would add water, electricity, and a telephone for emergencies to the list of basic services that should be universally available.

Moving from strategies that rely primarily on markets, such as private con-

tracts, to those that rely more heavily on politics, such as concession contracts or discretionary regulation, changes values in two ways. First, the involvement of government and politics almost inevitably opens the door to a wider set of values than the individualism of the market. If the government becomes involved in regulating telephone companies, for example, then it is very likely to insist on low "lifeline" tariffs for basic service to poor households, or on similar tariffs for basic services in rural and urban areas. The government might have adopted a scheme to improve access to telephones by poor households or in rural areas even if it were not regulating telephone service as a monopoly. Once the government takes responsibility for regulating telephone tariffs, however, then the pressure to include regulations that respond to other popular concerns, such as equity, is often overwhelming.

Whether one regards the introduction of other values as desirable depends upon one's sympathy for them. A long-standing concern of regulatory scholars is that government agencies may be vulnerable to "capture" by "special interests," including the regulated firms, their employees, customers, and others. The adjective "special" is used to indicate that they do not have the interests of the general public at heart, and thus that their influence over the regulatory agency should be seen as undesirable. Interests that seem special to one person may seem public to another, of course. It's all in the eye of the beholder. Nevertheless, experience demonstrates there is a significant risk, discussed at greater length in Chapters 3 and 4, that a regulatory agency will be captured by interests that conflict with widely shared values or goals, such as the promotion of efficient and affordable services.

A second way that the shift from more market- to more government-based strategies affects values is by increasing the concern with procedural fairness, both perceived and real. Procedural concerns arise because the government has a power that markets do not: the power to coerce individuals to do things that, from their own selfish perspectives, make them worse off. The most fundamental reason for using government to help solve a problem like monopoly is because coercion is thought to be needed. Problems that can be solved by voluntary agreement among self-interested individuals will be solved automatically by markets, without the government.

If the government's unique coercive powers are to be used against members of society, then the public is understandably anxious that coercion will be used fairly and only for important public purposes. The well-warranted price of using coercion in a democracy is usually elaborate procedural safeguards to prevent its abuse. These procedures are educational in the broadest sense, both in alerting policymakers to unanticipated problems and opposition and in building popular understanding and support for the actions ulti-

mately taken. And for that reason the perception of fairness is often as important as the reality of fairness. In the public sector, the process of reaching decisions is often as important as the decisions themselves.¹

The understandable concern with procedures and perceptions may conflict with other goals, including the promotion of efficient markets. Elaborate procedural safeguards may delay important decisions, for example, and are often thought to give regulatory processes a strong bias in favor of the status quo. If regulatory powers can only be adopted after lengthy and careful procedures, then regulatory change is likely to be slow. Similarly, the concern over perceptions of fairness may result in broad classes of consumers or firms being treated the same way despite substantial differences among them.

Much of the analysis in this book assumes that the primary goal of public policy toward monopoly is to encourage efficient markets. But it recognizes that other values and goals are also relevant and widely supported, particularly the desire to promote equity through universal access to certain basic services. And it recognizes, further, that the understandable consequence of invoking government powers to solve monopoly is both the introduction of wider values and an enhanced concern over procedures and perceptions.

Transaction Costs and Government Intervention

Given the perspective adopted, the choice of whether and how to regulate monopoly can be understood by using the concept of transaction costs to analyze the performance of markets and contracts. The concept of transaction costs was developed by Ronald Coase, Oliver Williamson, Douglass North, and the many researchers they inspired.² The central idea is that all economic activities can be broken down into a series of transactions, such as buying a good or service, and that executing these transactions imposes costs on the parties involved.

Transaction costs occur for two fundamental reasons. First, it is impossible for the parties to a transaction to know every relevant fact or anticipate every relevant contingency. Gathering information is time consuming and costly, and there are limits to our cognitive abilities to process information and forecast the future. Second, a party will behave opportunistically by withholding relevant information if it can gain advantages by doing so. In a world with perfect information and foresight and where parties were completely honest with each other, even complicated transactions could be arranged easily. But in the real world, the parties usually must spend time and money gathering and interpreting information and drafting and enforcing complex contracts to protect themselves against contingencies and the opportunistic behavior of the others.

Different institutional arrangements can affect the costs of a transaction, and some institutional arrangements may be more efficient for a given type of transaction than others. In an influential book, for example, Douglass North and Robert Thomas argued that the development of a legal system to enforce contracts was critical to Western economic development, because it dramatically reduced the costs of the many commercial transactions essential to a modern specialized economy.³ It was easier for the government, rather than private individuals, to organize an effective legal system, because only the government has the power to coerce individuals to comply.

As the example of the legal system suggests, economists increasingly use transaction costs as a framework to help decide when and how government should intervene in an economy.⁴ Ronald Coase demonstrated many years ago that there would be far less reason for government to intervene in markets in a world with no transaction costs, including fully defined property rights.⁵ Coase singled out property rights as a special source of transaction costs because ambiguous rights made it easier for one party to harm others without gaining their consent. Zero transaction costs would make it possible for individuals to engage in a variety of different voluntary transactions, and fully defined property rights would protect third parties not participating directly in the transaction. Government might intervene for equity reasons, for example, to help ensure that all individuals start out on a relatively equal footing by guaranteeing access to basic education and health care. The market would be efficient, however, in the sense that (1) all the transactions that could make at least one person better off and no one else worse off would be completed and (2) no transaction that failed this test would occur.

In this framework, government intervention is justified if it either clarifies ambiguous property rights or otherwise reduces transaction costs. Government regulation of pollution, for example, can be seen as a response to problems of property rights and other sources of transaction costs. If other transaction costs are low, pollution can be resolved by simply clarifying property rights. Imagine that a stream forms the boundary between two neighbors and that the rights to use the stream for various purposes are not clearly defined. Imagine further that one neighbor finds it convenient to dump his trash in the stream, harming the second neighbor, who uses the stream for fishing and as a source of drinking water. In the absence of clearly defined rights for the use of the stream, the two neighbors may find it difficult to reach an agreement as to whether the trash can be dumped. If the government establishes either a clear right to dump trash in a stream or a clear right to a fishable and drinkable stream, however, then the two neighbors should be able to negotiate an efficient use of the stream.

One of Coase's famous contributions was to argue that the solution would

be efficient no matter how the property right was defined. Either way, the first neighbor would dump his trash only if the benefit he received was greater than the harm to his neighbor. If trash dumping was a right, then the second neighbor would pay the first not to dump only if the loss of fishing and drinking from the stream exceeded the convenience of dumping. If a fishable and drinkable stream was a right, then the first neighbor would pay the second for permission to dump only if the convenience of dumping exceeded the value of fishing and drinking.

The clarification of property rights alone may not solve the pollution problem if there are other sources of serious transaction costs. Suppose that our stream is now a river that is hundreds of miles long and passes by thousands of households and businesses that want to use it for trash disposal, fishing, and drinking. Suppose further that the issue is no longer whether to dump a fixed amount of trash at one site but rather what quantities of a dozen different pollutants should be dumped at hundreds of different locations. Even in the presence of clearly defined property rights, the transaction costs of negotiating, monitoring, and enforcing a voluntary agreement among the thousands of river users are likely to be enormous.

In the presence of such high transaction costs, it may be more practical for the government to regulate the quality of the water and the amount of dumping permitted. In effect, the government tries to estimate the agreement that the thousands of parties would have reached voluntarily if the transaction costs were zero, and then uses its unique power of coercion to enforce compliance. The government is likely to make errors in its estimate, however, and may introduce other problems by its involvement. Even in the presence of high transaction costs, government intervention is desirable only if it moves society closer to the solution that the parties would have agreed to in a world free of transaction costs. In other words, the cure must be better than the disease.⁶

The case for government intervention in infrastructure monopolies is similar to that for intervention in pollution. The main difference with monopoly is that ambiguous property rights are not a major source of high transaction costs. As explained in Chapter 1, both the infrastructure company and its customers make durable and immobile investments that leave them vulnerable to opportunism. The logical solution is for the parties to sign contracts to protect themselves before they make their investments. As in the case of pollution, however, the transaction costs of negotiating and enforcing these contracts may be high, particularly if many small customers are involved or if their infrastructure requirements are complex and hard to predict. And as in the case of pollution, there are a variety of ways the government might intervene. The appropriate intervention depends on how high the transaction

costs are and how effective the government is at reducing them without introducing new problems.

In the case of monopoly, the scope of government intervention increases gradually as one moves from private contracts to concession contracts, discretionary regulation, and public enterprise. Government intervention can only approximate the results of voluntary agreements in a world with zero transaction costs, since the government can never know as much about the potential costs and benefits of the transaction as the individual parties do. The greater the range of issues the government must decide, the cruder the overall approximation is likely to be.

From this perspective, the most intrusive forms of government involvement are justified only when transaction costs are relatively high, so that even a crude approximation is an improvement. Private contracts are appropriate if transaction costs are relatively low, concession contracts might be justified if transaction costs are moderate, and discretionary regulation or public enterprise could be justified only if transaction costs are extremely high and less intrusive remedies impractical.

Contractual Completeness and Contract Types

Options in Procurement

Many solutions to monopoly rely on a contract of one form or another, the most obvious being private and concession contracts. Some of the practical issues that arise in the contractual approach to monopoly are analogous to those found in a firm's decision about how vertically integrated it should be. Indeed, Oliver Williamson's analysis of procurement and vertical integration was one of the seminal studies of transaction costs, and Williamson would later apply many of these same insights to the analysis of concession contract regulation.⁷

Imagine a firm that produces a final product to sell to consumers by assembling various intermediate products or parts. The firm could produce the parts itself or buy them from independent suppliers. The more inputs the firm produces itself and the farther back those inputs reach into the chain of potential suppliers, the more vertically integrated the firm is said to be. In deciding whether to make or buy a particular part, the firm faces a range of possibilities, as shown in Figure 2.1. At one extreme the firm could produce the part itself, and at the other extreme it could buy the part on the spot market, where goods are offered for immediate delivery. In between are a variety of other options that involve long-term contracting between the firm and an independent parts manufacturer.

Buying on the spot market is often the preferred solution, since competition in spot markets provides built-in incentives for the suppliers to keep the costs of the part low. If the part is produced internally, by contrast, the firm must develop an internal management system to monitor the performance of the employees producing the part and to motivate them to produce it efficiently. The spot market is also typically more reliable, because there are many alternative suppliers instead of just one. The spot market has limitations, however, among the most important being that it may discourage the buyer or its suppliers from taking advantage of the potential economies of tailoring their facilities, employees, and processes to one another's capabilities or needs. As with monopoly, the parties making such durable, relationship-specific investments become vulnerable to opportunistic behavior unless they can protect themselves with some form of long-term contract.

A central limitation of long-term contracts is the possibility that the contract may prove to be incomplete or become obsolete if circumstances change. To protect against opportunism, the contract must last as long as the lives of the relationship-specific investments that it is designed to protect. The longer the contract, however, the more the circumstances of the parties are likely to change over the contract's life. One can include contingencies in the contract, of course. But identifying all the relevant risks and negotiating appropriate contingencies is time consuming, costly, and impractical, however, especially if the contract is to last for a long time and the needs of either party are highly uncertain and complex. And if the contract proves to be incomplete before it is scheduled to expire, then the parties will face the choice of either living with unsatisfactory terms for the remaining life of the contract or exposing themselves to opportunism by renegotiating the contract.

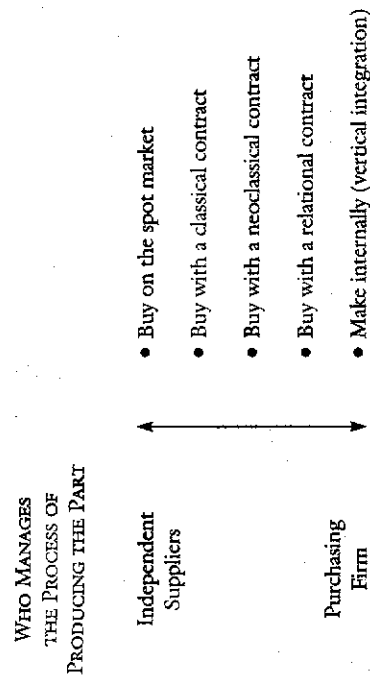


Figure 2.1 The continuum of strategies for procurement by a firm.

The problem of contractual completeness has spawned a variety of contract types, of which Williamson distinguishes three.⁸ The first is the classical contract, which presumes to describe, in advance, all relevant future contingencies and to provide specific remedies for them. This is likely to be an unrealistic ideal for all but the simplest and shortest relationships.

The second type is the neoclassical contract, which acknowledges the possibility of incompleteness and includes provisions for a third party, usually an arbitrator of some sort, to resolve certain classes of disputes. Arbitration is often called for because it can be more efficient at discovering information and resolving disagreements about facts or states of the world than normal litigation. However, the circumstances under which arbitration is allowed, the factors the arbitrator can consider, and the scope of the remedies that he can order are often carefully prescribed.

The third and most flexible type is the relational contract. The relational contract anticipates that the relationship between the parties will have to adapt over time in unforeseeable ways. It relies on renegotiation between the parties as the remedy, and provides a general framework for those negotiations rather than specific remedies for identified contingencies. Relational contracts leave the parties vulnerable to holdups over specialized investments, because the guidelines for renegotiations are general. As a result, relational contracts are likely to be used only where there are additional protections outside the contract against opportunism. One form of extra-contractual protection is when reputation matters because the parties expect to engage in repeat contracts with the same or similar firms. The prospect of repeat contracts is important because the parties' willingness to behave opportunistically in any one situation is reduced by threat of retaliation in the future.

Empirical research confirms that long-term contracts are common in industries where relationship-specific investments are important.⁹ Classical or neoclassical contracts are by far the most common contract types. Relational contracts are used only if the requirements of the parties are difficult to specify in advance and there are extra-contractual protections against opportunism. Finally, the firm is likely to produce the part internally rather than buy it from others if relationship-specific investments and flexibility are important but extra-contractual protections are not available.

The Implications for Monopoly

The range of solutions in procurement is similar to the range of solutions for infrastructure monopoly presented in the previous chapter (compare Figure 2.1 with Figure 1.1).¹⁰ The firm trying to buy parts plays the same role as the consumers trying to buy infrastructure services, except that in some cases the

government acts on the infrastructure consumers' behalf. At the two extremes, buying on the spot market is analogous to buying from a private and unregulated infrastructure supplier; producing the part internally is analogous to having a public or nonprofit enterprise provide the infrastructure service. In between the extremes are a variety of types of contract, some more classical and some more relational. In particular, most private and concession contracts could be classified as classical or neoclassical contracts. Similarly, discretionary regulation can be understood as a form of relational contracting, because the regulatory statute sets out the obligations between the government and the enterprise in only general terms. In discretionary regulation, as in relational contracts, both parties usually rely on other sources of protection against opportunism besides the statute.

There are two important differences between the problem of infrastructure monopoly and the more general problem of procurement. The first is that relationship-specific investments—in the form of durable and immobile investments—are the hallmarks of an infrastructure monopoly, whereas they are not found in every procurement situation. This means that buying on the spot market is a much less common and attractive option in monopoly than it is in procurement generally. In cases of monopoly, some form of contract is usually superior to the spot market.

The second difference is that the monopoly problem often involves many consumers buying from a single infrastructure supplier, whereas the typical procurement problem involves a single firm buying from one or more suppliers. The presence of many buyers can complicate the use of contracts, particularly if each buyer is purchasing a relatively small amount so that the transaction costs of negotiating and enforcing individual contracts are large relative to the total value of each transaction. The high transaction costs of contracting with many small buyers may make it desirable for the government to act on the consumers' behalf through a concession contract or discretionary regulation.

These differences aside, the procurement analogy offers two important lessons for the use of contracts as a solution to monopoly. One lesson is the potential vulnerability of contracts to incompleteness or obsolescence. The contract protects against opportunism best if it anticipates all the important events and provides appropriate contingencies for them. This means that an approach based on a classical contract (whether a private contract or a government concession) is less attractive in situations where it is difficult to anticipate what might happen and how to respond. Classical contracts tend to be difficult for infrastructure monopolies because the lives of the investments tend to be long, and the longer the contract the harder it is to anticipate what will happen.

The second important lesson from procurement is that there are a variety of contract types, ranging from the classical contract to the relational contract, with hybrids in between. The relational contract acknowledges that drafting a complete contract is an impossible goal in some situations; the protection against opportunism rests partly on the specific provisions of the contract and partly on extra-contractual circumstances and institutions. In the case of procurement between private firms, the extra-contractual protections are often the prospect of repeat business and the social ties and trust that have developed between the purchasing agent and the supplier over time. Repeat business and reputation can be important in the case of discretionary regulation as well. The government's incentives to be opportunistic are more limited, for example, if it knows the infrastructure company must return regularly to the private capital market to replace worn-out assets or to expand capacity to accommodate growing demand. Similarly, the infrastructure company is less likely to behave opportunistically if it hopes to win permission to provide other infrastructure services from the same or similar governments.

Reputation and repeat business do not seem to provide as effective protection in discretionary regulation as they do in procurement between private firms. One sign that reputational concerns are not enough is that governments and infrastructure companies often take deliberate steps to develop extra sources of protection. As we shall see in later chapters, governments often attempt to protect themselves by threatening to franchise another private firm or establish a publicly owned competitor, for example. Similarly, the infrastructure suppliers try to protect themselves by buying control of key assets or by selling stock to the general public to broaden the popular support for their cause. The extra-contractual protections are not confined to the specific defenses that the government and the regulated firm erect. The broader legal, social, or political institutions limit the behavior of governments and companies in important ways as well. As discussed in Chapter 6, for example, constitutional restrictions on the government's taking private property without compensation have significantly limited opportunism by U.S. regulatory agencies.

The Pros and Cons of the Principal Options

Private Contracts

In assessing the basic advantages and disadvantages of the different regulatory strategies, it is helpful to think of private contracts, concession contracts, and discretionary regulation as three distinct types. In practice many schemes mix elements of the approaches, but the strengths and weaknesses of the

hybrids can be understood as products of their components. The strengths and weaknesses of the three principal regulatory strategies are summarized in Table 2.1.

One key strength of private contracts is that they can establish clear commitments. The commitments depend on the completeness of the contract and on the integrity of the courts that enforce commercial or contract law. Assuming one has a reasonably complete contract and a fair legal system, however, the contract describes all the obligations of the two parties in advance.

The second strength of private contracts is that they allow the greatest scope for competitive market forces to shape the infrastructure services. Market forces have a strong influence because the infrastructure suppliers and their customers draft the private contracts without interference from government. This means that the parties are free to search for opportunities to redesign or tailor the services in ways that add value for the consumer or reduce costs for the infrastructure company. Under both concession contracts and discretionary regulation, by contrast, the government regulatory agency typically requires that the firm offer customers only a limited menu of service offerings. A limited menu makes it easier for the regulator to monitor the firm's

Table 2.1 Principal options for regulating a private infrastructure monopoly: Advantages and disadvantages

Option	Advantages	Disadvantages
Private contract	• Clear and specific commitment of a commercial contract • Greatest scope for competitive market forces to shape service offerings and control costs	• Risk of an incomplete contract • Depends on integrity of commercial law courts • Small customers may make it costly to negotiate and enforce private contracts
Concession contract	• Clear and specific commitment of a commercial contract • Competitive market can be used to determine supplier costs	• Risk of an incomplete contract • Depends on integrity of commercial law courts • Regulator must second-guess consumer preferences
Discretionary regulation	• Flexibility to respond to changing and unforeseen circumstances	• Regulatory agency must second-guess both consumer preferences and supplier costs • Greater risk of capture of regulatory agency by special interests

compliance with the tariff rules set out in the concession contract or by the regulator. A limited menu is often politically more acceptable as well, because to the general public it often seems fairer to offer broad classes of consumers similar services and prices than to vary price and service offerings from one consumer to the next. But the limited menu may suppress important differences among customers and reduce opportunities for market-driven tailoring and innovation. The more heterogeneous the customers, the more advantageous the private contractual approach.

The main disadvantages of private contracts are the risk of an incomplete contract and the potentially high costs of negotiating contracts with many small customers. In many cases, the needs and circumstances of the two parties are predictable enough to draft a complete contract. If they are not so predictable, however, then the parties must either find extra-contractual forms of protection or reconcile themselves to the risk that the contract will fail and leave them vulnerable to opportunism. Similarly, not all infrastructure systems have many small customers, and sometimes strategies can be developed for reducing the costs of negotiating with small customers (as explained in Chapter 8). Nevertheless, the costs of negotiating and enforcing a contract usually do not decrease proportionately with the value of the transaction, so that it may not be worthwhile to negotiate individual contracts for small customers.

Concession Contracts

Concession contracts maintain the clear commitment of a commercial contract and some significant influence for market forces while eliminating the difficulty of negotiating with many small customers. The concession contract is no different than a private contract in its aspiration to describe obligations completely in advance. Usually a government regulatory agency monitors the company's compliance with a concession contract, but that agency cannot unilaterally change the terms once the contract is awarded. Contractual disputes are normally adjudicated by the same commercial law courts that enforce private contracts rather than by the regulatory agency. The presence of the government as a party to the contract may influence the courts, but the fact that the entire economy depends on the integrity of the contract law system provides a potentially powerful countervailing pressure.

The influence of market forces is not as strong as in the case of the private contract, because consumers are represented by the government instead of participating directly. Often the government specifies the minimum quality of services to be provided and then awards the concession to the bidder proposing the lowest tariff. Another common variant is for the government to

specify both the minimum service and the maximum tariff and then award the concession to the bidder offering the largest concession fee or requesting the lowest subsidy. Either way, the government must estimate what consumers want when it drafts the basic concession framework.

Nevertheless, market forces are involved, because the concession is usually awarded through competitive bidding. A competitive award should ensure that the infrastructure supplier is driven to offer terms that reflect his costs, and no more. And if the concession expires and is rebid periodically, then the terms should reflect reasonably up-to-date market conditions. Equally important, open and competitive bidding can enhance popular perceptions of fairness and help maintain political support for the regulatory process.

Concession contracts share with private contracts the limitation of contractual completeness. Again, some situations are predictable enough to make a complete contract a realistic goal. And there are ways to build some flexibility in contracts through such schemes as arbitration and buyout clauses. But foreseeing the future and drafting appropriate contingencies is often difficult.

Discretionary Regulation

The best-known examples of discretionary regulation are cost-of-service regulation as developed in the United States and price-cap regulation as developed in the United Kingdom.¹¹ Discretionary regulation is often called commission regulation in the scholarly literature, since, until recently, the primary examples were in the United States, where it is customary to appoint a commission of three to seven members who decide regulatory issues by majority vote. The British gave individual regulators the authority to make decisions in the original price-cap schemes of the 1980s, but they began to replace individual regulators with commissions in 1998. The term *discretionary regulation* better reflects the distinctive characteristic of this approach.

The basic advantage of discretionary regulation is its flexibility, enabling it to adapt to changing and unforeseen circumstances. Unlike contract regulation, the discretionary approach does not attempt to anticipate all the developments that may happen during the life of the investments that need protection. Instead, a regulatory commission or an individual regulator is granted substantial discretion to set the prices and service standards for the regulated firm. The authorizing statute usually constrains the regulatory body to some degree by, for example, setting out the factors that it must consider. And there is usually some provision for appeal to the courts or a similar body if the firms or consumers feel the regulator's rulings are inconsistent with the statute. But the statutory guidance is seldom formulaic, so the regulatory body enjoys substantial freedom in interpreting it.

A major disadvantage of the discretionary approach is that the influence of market forces is drastically reduced. In essence, the regulatory agency must predict almost all of the decisions that the infrastructure supplier and its customers would have made in a market with no transaction costs. Markets are still involved, in that the private capital market will not finance new investments if the regulator treats the infrastructure company too harshly. But the constraints of the capital market are obvious only in the long term or when capacity is limited, and even within those constraints the regulator has plenty of latitude to influence the services provided. As in the concession contract, the regulator must represent consumers in deciding what combination of quality and service they prefer. But unlike the concession contract, the regulator can no longer rely on competitive bidding to ensure that the infrastructure supplier is providing those services at the lowest cost and without earning excess profits.

The need to estimate what the efficient market solution would have been in the absence of transaction costs makes discretionary regulation technically challenging. Under cost-of-service regulation as developed in the United States, for example, the regulator is supposed to set tariffs high enough to cover the costs of an efficient firm, including operating expenses, depreciation, and a reasonable return on invested capital. But to make this calculation, the regulator needs to know what the operating expenses for an efficiently run firm are, how rapidly investments depreciate, whether the investments being made are prudent and well managed, and the minimum rate of return that the private capital market requires to finance investments of comparable risk. Many of these questions are hard for the firm itself to answer, let alone the regulator.

The tariff-setting task is complicated by the fact that the regulator inevitably has less information and analytic staff than the firm has. Although the firm is obligated to provide actual cost data and other information that the regulator asks for, the data seldom tell the regulator all he needs to know. A regulator can ask the firm to supply figures on actual and projected operating expenses and output, for example, but the regulator will have to determine independently whether the costs can be reduced further or not. The British price-cap system was developed in part to reduce the regulator's need for information but, as explained in Chapter 9, it has not been totally successful in this regard.

The second major drawback of discretionary regulation is the risk that the agency will be captured by special interests and therefore will not exercise its discretion in ways that are in the long-term interests of the infrastructure supplier and its customers. The agency needs discretion because tariff setting is too complicated to be reduced to a contract or a formula that can be specified

in advance. But the fact that the agency has so much discretion makes it a more inviting target for capture. Moreover, a regulatory agency that is specialized in a particular industry does not have the same built-in defenses against capture found in the commercial law courts. While all industries and consumers have a stake in the integrity of the courts, only the regulated firms and their customers have a stake in the integrity of the specialized regulatory agency.

Usually the regulatory agency is designed to provide some insulation from day-to-day political pressures. The commissioners or individual regulators are typically appointed to fixed terms, for example, and cannot be removed except for specific and limited causes. In the case of commissions, the terms of the members are usually staggered so that it takes years before the president, prime minister, or governor has appointed a majority of its members. Commissions are sometimes required to take evidence or deliberate in public, and to provide written justifications for their decisions. Nevertheless, political pressures inevitably affect regulators. Many regulators hope for reappointment at the end of their terms, for example, while those that don't usually still want the regulatory system to be regarded as successful, or at least to maintain enough popular support to survive. The fact that the risk of capture cannot be eliminated entirely by the design of the regulatory statute or agency often leads regulated firms and consumers to rely on other sources of protection against opportunism.

Variants and Hybrids

In practice, regulatory schemes are often mixtures or hybrids of the private contract, concession contract, and discretionary regulation approaches. The scheme to regulate U.S. freight railroads relies primarily on private contracts, for example, but it includes the opportunity to appeal to a discretionary regulatory agency. The appeals process is designed primarily as a safety valve for small shippers, and appeals are permitted only under limited circumstances. Similarly, British price-cap regulation relies primarily on a discretionary approach but with an important contractual component.

Figure 2.2 provides a sense of the variation by arraying a number of schemes along the continuum from markets to politics. There is substantial variation even among schemes that would be classified as predominantly using a single approach. Concession contracts that are awarded through competitive bidding are subject to more direct market forces, for example, than concession contracts awarded without bidding. The variation among discretionary schemes is perhaps even broader. In some schemes politics enters di-

rectly and the scope of discretion is virtually unlimited. An example is the bus industry in Sri Lanka where, as described in Chapter 4, ministers and the Parliament set tariffs. In other schemes the scope of discretion and the potential influence of politics are substantial but limited by various statutory and institutional constraints. Most modern forms of discretionary regulation fit this description, including the U.S. cost-of-service and the British price-cap approaches.

Some schemes include such a strong mixture of the contractual and discretionary approaches that they can only be classified as hybrids. One example is the French system of municipal concession contracts with a special adminis-

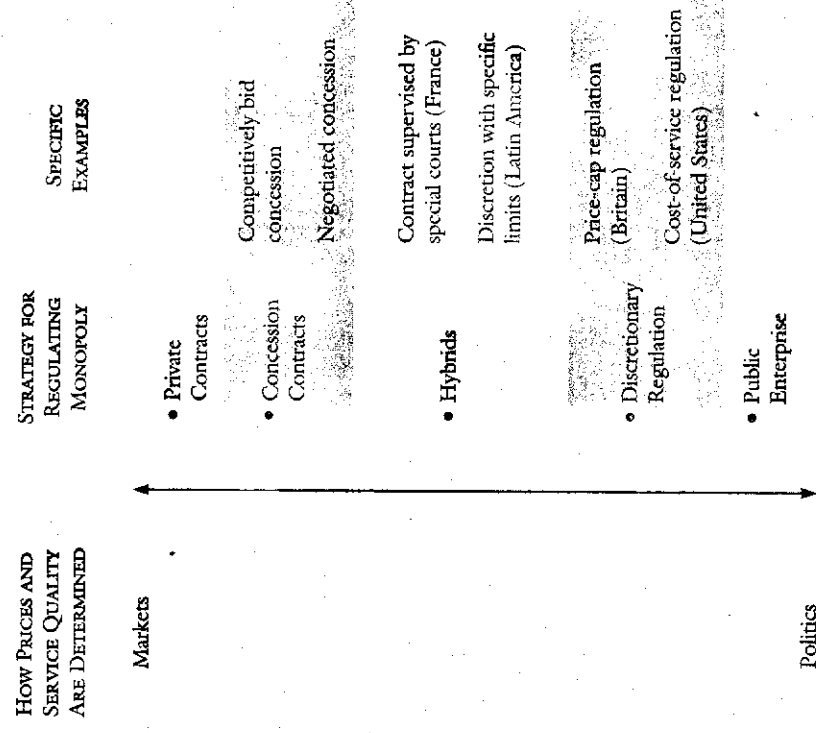


Figure 2.2 The continuum of strategies for regulating monopoly.

trative court to supervise renegotiations. This system, which has been developed over many years, is used by municipal governments in France and in some of the former French colonies in Africa to regulate private water, sewage, solid waste, and bus and streetcar companies.¹² As in conventional contract regulation, the government awards a concession for a fixed term and the concession contract specifies the prices to be charged and services to be provided. Unlike the conventional contract, however, the French contract anticipates that the municipality and the concessionaire may need to renegotiate price or service terms during the life of the contract. The contract sets out the circumstances under which either party can request renegotiations and the general criteria the parties must consider in negotiating changes, but the language is not as specific or formulaic as in the adjustment clauses found in conventional contracts. Equally important, disputes over renegotiations can be appealed to specialized administrative courts that apply broad principles for adjusting municipal contracts that are set out in French administrative law. In essence, the specialized administrative courts play the role of a discretionary regulator in that they are empowered to change a contract, not just interpret its terms. But these courts are used only in the event that the two parties cannot reach agreement on their own.

Similarly, some discretionary systems attempt to establish explicit commitments that are more characteristic of a contract. The most common method of doing so is by legislating a few very specific constraints on the conduct of the discretionary regulatory agency. For example, the law might specify the minimum rate of return that the regulator must provide private investors. Mexico adopted this approach in the 1930s, when it passed a law requiring regulators to allow private electric companies to earn a return at least as large as the rate of interest paid on government debt. Similarly, in the 1930s and again in the 1980s Chilean law required regulators to allow a return of at least 10 percent per year.¹³ These constraints were not sufficient to prevent the gradual impoverishment and nationalization of many private electricity companies in Latin America between the 1930s and the 1970s, as described in Chapter 6. But more modern versions may prove successful in combining the flexibility of discretion with some of the specific commitments found in contracts.

Finally, it is worth remembering that the performance of any regulatory scheme can be deeply influenced by its broader social and political environment, sometimes enough that it has the effect of changing the basic regulatory approach. The most obvious example is a concession contract scheme in a country without an independent judiciary. If the courts do the bidding of the executive or the legislative branches of government, then concession contracts amount to discretionary regulation in disguise.

Conclusion

The more market-oriented solutions to monopoly have two basic advantages. First, the more market-oriented the remedy, the more closely it is likely to approximate the agreements that infrastructure companies and their customers would reach voluntarily in a world with no transaction costs. Any intervention will only approximate the results desired, and the broader the scope of the intervention the more prescient the regulator must be to come close. No intervention is the best course of action if intervention cannot improve on the protections against opportunism that the companies and customers could develop for themselves, despite the transaction costs involved. And the more extensive forms of government regulation are likely to be justified only where the transaction costs of private long-term contracts are high, so that the uncorrected market performs poorly relative to the ideal.

The second advantage of the more market-oriented approaches is that they have the potential to provide stronger protection against opportunism. One reason is that the market-oriented approaches typically rely on some form of explicit contract, either private or concession. These contracts work only if they are reasonably complete and if the legal system that enforces them is reasonably fair and efficient. If these two conditions are met, however, contracts provide a clearer and stronger commitment than discretionary regulation.

Market-oriented solutions are also more stable because they raise fewer concerns about the use of government powers and the fairness of regulatory proceedings, and thus are less likely to generate the kinds of political controversies that lead to intervention and broken commitments. Private contracts will be politically controversial if some customers remain vulnerable to opportunism because the transaction costs of negotiating satisfactory long-term contracts are prohibitively high for them. These political objections may be fatal if the remaining problems of opportunism are too broadly or intensely felt.

With the shift from private contracts to concession contracts or discretionary regulation, however, the focus shifts from whether the market is fair to whether the government is fair. And as one moves from concession contracts to discretionary regulation, the difficulty of demonstrating fairness increases. The concession contract has the important advantage that it can be awarded through an open and competitive bidding process, which the public is likely to accept as fair. Critics may object that the government should have set a different minimum service standard or maximum tariff when it drafted the concession contract, but if the contract is then awarded competitively it is harder to argue that the firm is making unreasonably high profits. With discretionary regulation, by contrast, the government has no analogous procedure that it

can use to win the instant credibility of competitive bidding. As a result, the task of demonstrating fairness and maintaining public acceptance becomes much harder in discretionary regulation than it is with concessions.

Nevertheless, market-oriented solutions have important limitations. Private contracts are potentially vulnerable to the problems posed by large numbers of small customers, and both private and concession contracts are vulnerable to courts that lack integrity and to the difficulties of drafting a complete contract. If the contract proves to be incomplete, moreover, all its advantages of procedural fairness and protection against opportunism disappear quickly. Any political advantages of awarding a concession contract competitively are lost, for example, if the contract has to be renegotiated.

Given these limitations, discretionary regulation may be the best solution where the circumstances are too complex or unstable to draft a complete contract. Discretionary regulation should not be entered into lightly, however, because the difficulty of approximating what the market might have done is great and the risk of capture and difficulties of maintaining public support are high.

I

Regulatory Politics and Dynamics

The Behavior of Regulatory Agencies

The Public Interest

To the extent that regulatory agencies have the discretion to set tariffs and service standards, how will they use that discretion? Three theories are often advanced to explain why government regulation is established and how regulatory agencies behave once they are created. The first views the agency as a tool for discovering and advancing the broad public interest in the performance of the regulated industry. The second presumes that special interests will capture the agency in order to use the agency's powers for their own parochial ends. A final theory suggests that the mix of public and special interests that the agency will pursue depends on the design of the agency and the institutional environment in which the agency operates. These theories have been explored most intensively in the United States, because of its longer continuous experience with regulated private utilities. But they are increasingly the topics of research in other countries as well.

The concept of the public interest assumes that members of society have common or widely shared goals, and that the best means to identify and advance these goals is through rational and expert analysis. In the case of infrastructure regulation, the primary goal is usually the promotion of economic efficiency that is threatened by the presence of natural monopoly. Often there are secondary goals as well, such as a desire to promote universal access to basic infrastructure services for the poor or for those in remote or lagging areas.

Regulatory agencies are almost always created in the name of the public interest. In the United States, for example, the pursuit of the public interest was the avowed reason why many states replaced regulation by municipal concession contracts with regulation by state public utility commissions (PUCs) between 1907 and 1914. As explained in Chapter 7, the PUCs were the product of a debate within the Progressive political movement in the United States and Canada about whether public ownership, municipal regulation, or state regulation would best protect the public from private mo-

nopolies. The PUCs triumphed in part because they embodied the rationalist, public interest perspective. PUC commissioners were appointed to fixed terms and were removable only for specific causes, for example, so that they could make decisions that might be unpopular in the short term without fear of political interference. Similarly, by shifting regulation from the municipal to the state level, advocates of PUCs hoped to create regulatory agencies that had enough responsibilities and resources to justify and support an expert staff.

The public interest was also the rationale for the creation of federal regulatory commissions in the United States. Congress established the Interstate Commerce Commission (ICC) to regulate railroads in 1887, but federal regulatory agencies did not proliferate until the 1930s, when the Depression undermined popular confidence in private enterprise. President Franklin D. Roosevelt, as part of his New Deal, argued that government coordination was needed to ensure that industries performed in the public interest, and new regulatory agencies were created to supervise the securities, telecommunications, energy, airline, and other industries.¹ A similar faith in the pursuit of the public interest was professed by the designers of the regulatory agencies for the utilities privatized in Britain during the 1980s and in many other countries during the 1990s. (See, for example, the description of the origins of the British price-cap system in Chapter 9.)

Critics of this perspective often argue that the public interest is an illusion or that it cannot be discovered through rational analysis. There may be widespread agreement when goals are stated at a very general and abstract level, but the consensus often dissolves when the goals are translated into specific policies. We may all agree that some basic level of electricity or telephone service should be widely available and affordable, for example, but then disagree about what constitutes basic service, how to define affordability, and whether taxpayers or other utility customers should finance any losses on basic service.

Even if there is a public interest, moreover, it may not be discoverable through the rational deliberations of an expert regulatory agency. Some critical information simply may not be available with reasonable time or effort, or so much information may be required that even a large and expert staff cannot comprehend and analyze it all. Given that we don't and can't know everything necessary to make completely informed and rational decisions, some critics argue that relying on experts is actually risky. There is virtue in developing policy incrementally, through political competition among parochial interests.

To proponents of the public interest, these criticisms seem overblown. Just because there is controversy or uncertainty about the nature of the public interest doesn't mean that it doesn't exist. Since all citizens rely on infrastruc-

ture services, they have a common interest in seeing that infrastructure is provided reasonably efficiently and priced not too much above cost. The promotion of universal access is more controversial, but often enjoys widespread support. The general public has an interest in promoting universal piped water and sewerage, for example, inasmuch as they help protect everyone against the spread of disease. Similarly, rural roads, electricity, and telephones may be important to the general public because they promote social integration and political stability, particularly in developing countries.

In addition, many proponents have faith that the public interest will emerge from the give and take of different interest groups in a democracy. This belief was particularly strong among political scientists in the United States in the 1950s, perhaps because the principal alternative at the time seemed to be some form of fascism. The belief became popular once again toward the end of the twentieth century after the fall of many communist states. The presumption is that society is "pluralist," in that power is widely distributed rather than concentrated. As a result, public policies are adopted only after debate, compromise, and the formation of broad and heterogeneous coalitions.

Finally, public interest proponents do not put their faith in democracy alone, but also seek to perfect regulatory institutions that will encourage the development of the public interest. The issue of institutional design is discussed later in this chapter, but the state PUCs established in the first third of the twentieth century are examples of regulatory institutions crafted to encourage the rational discovery of the public interest.

Capture by Special Interests

Whether or not there is a real and knowable public interest, many scholars argue that in practice it has proved difficult to prevent regulatory agencies from being captured by special interests.² The concern about capture is as old as regulation, although our understanding of who might capture the agency, how, and with what degree of influence has evolved over the years.

During the late nineteenth and early twentieth centuries there were many allegations in the United States, Latin America, and elsewhere that municipal and state officials had been bribed by the companies they were supposed to regulate. Just how extensive the problem was and who had corrupted whom were not always obvious. Nevertheless, outrage over corruption was a major factor in the regulatory reforms put in place in the United States during the first decades of the twentieth century.³

During the 1950s, some U.S. scholars began to point out that the reforms had not prevented the regulated firms from exercising substantial influence

over regulatory decisions.⁴ In an influential 1955 book about federal regulatory agencies, Marver Bernstein proposed that agencies go through a life cycle, beginning as aggressive advocates of consumer interests but then declining into protectors of the status quo or, worse, servants of the firms they regulate.⁵

Bernstein and others noted that firms could capture the regulators with much more subtle and legal methods of influence than bribes. It was common at the time, for example, for the commissioners and senior staff of U.S. regulatory agencies to take well-paid jobs with the firms they were regulating when they retired. This practice was eventually restricted, but typically by requiring retirees to wait only a year or two, so that the prospect of a comfortable job presumably still made at least some regulators more sympathetic to the industry's perspective. Moreover, it has always been perfectly legal for a regulated firm to lobby and provide campaign contributions to the legislators who oversee a regulatory agency. Indeed, the industry representatives might capture the regulators intellectually simply because they spent so much time in each other's company.

The heart of the argument by Bernstein and his successors was not just that there were legal ways for the firms to influence regulators, but that the firms also typically had stronger incentives and more resources to press their cases than consumers did. Soon after Bernstein's book appeared, Anthony Downs and Mancur Olson described more formally the difficulties of organizing a group to pursue policies of collective benefit.⁶ Among the obstacles were the time and resources it took for the members of a group to become informed and to influence policy. In addition, if all members of the group stood to benefit once favorable policies were enacted, then members had an incentive not to contribute resources to the common campaign but to "free ride" on the efforts of their peers instead. The larger the membership of the group, the harder it was to detect and discourage free riding. In a regulated industry the producers were more easily mobilized than their customers, because they were fewer and had more at stake. Regulatory decisions were critical to the financial health of the firms, while they affected only a small part of the budget of a typical customer.

During the 1960s, other researchers advanced the proposition that regulatory agencies were actually established at the behest of producers rather than being captured by them later. Producers might lobby to be regulated, for example, because the industry had elements of competition as well as monopoly, and regulation could be used to suppress those competitive sectors or elements. The historian Gabriel Kolko was among the first to make this argument in two books on the origins of railroad and economic regulation in the

United States.⁷ Kolko claimed, for example, that the railroads were the single most important advocates for the ICC in the 1870s and 1880s, because the ICC could protect them from aggressive state regulators and provide the means for suppressing rate wars between railroads with parallel trunk lines. Historians soon found that firms had been active in lobbying for regulation in other industries. These revisionist histories coincided with the publication of several influential studies by political scientists that argued that pluralism was an illusion in the United States. Although not specifically focused on regulation, these studies contended that business and economic groups held disproportionate political power and that the public interest was usually a disguise for policies that served their private interests.⁸

Some economists, led by faculty at the University of Chicago, began to think along the same lines as the historians and political scientists. In an influential 1971 article, George Stigler offered an extreme version of Kolko's argument by proposing that the creation and operation of regulatory agencies could be understood entirely as a device to transfer economic resources to various private interests in return for those interests providing votes or campaign contributions to politicians.⁹ Stigler hypothesized that every industry or occupation that had enough political power would seek to establish regulations that limited competition by controlling entry and fixing prices. Public interest goals, such as monopoly protection or promoting universal access, simply provided a plausible cloak to hide the real purposes of regulation.

The Chicago law professor Richard Posner soon advanced an alternative: that regulatory agencies were more likely to be captured by subsets of their consumers rather than by producers.¹⁰ Because the regulated firm is a monopoly, it can charge some customers tariffs well above costs and use the profits to cross-subsidize low tariffs for other customers. Customer groups therefore have incentives to try to influence the regulator to give them favored treatment at the expense of others. In the United States, for example, it was often alleged that pressure from farmers had led railroad regulators to set very low rates for shipments of agricultural products that were cross-subsidized by very high rates on shipments of manufactured products. Posner labeled this strategy "taxation by regulation" and argued that the high prevalence of cross-subsidies in regulated industries strongly suggested that this was the dominant form of capture.

A third Chicago professor, the economist Sam Peltzman, attempted to combine Stigler's and Posner's theories by viewing the regulators (or the legislators who oversee regulatory agencies) as politicians in search of support from competing interest groups.¹¹ Producers may be more motivated and in-

fluent in some circumstances and consumers in others. The regulator will favor the interest group that can most readily deliver the votes or other support the agency needs. If no one group can supply the support the agency needs, moreover, Peltzman hypothesized that regulatory policy would be a compromise that left all the agency's supporters optimally dissatisfied. At its heart, however, Peltzman's approach still viewed regulatory agencies primarily as devices to transfer resources among interest groups in return for votes or other sources of support.

The work of Bernstein, Kolko, and the University of Chicago faculty stimulated several decades of research into the origins and behavior of regulatory agencies. In the midst of this effort, U.S. regulatory policy underwent two dramatic changes that seemed to undermine Kolko and Stigler's views that regulation is created to serve the interests of the regulated. First, the traditional pricing and entry regulations that Kolko and Stigler thought served corporate interests were either wholly or partially abandoned in the U.S. railroad, trucking, airline, telephone, and financial service industries during the late 1970s and early 1980s. Second, environmental and occupational safety regulations expanded dramatically and over the strenuous objections of business during the 1970s. Had U.S. researchers thought foreign experience relevant, they might have been equally intrigued by a third change: the expropriation of many private utilities in Europe and in developing countries between World War II and 1970. In any event, all three of these developments suggested that regulation might not always serve the regulated industry's interests, and that regulatory regimes were not necessarily stable.

Scholars made several attempts to reconcile these developments with the idea that regulation was usually initiated or captured by special interests. One was to argue that industry costs or demand could change over time in ways that made regulation more or less attractive to special interests.¹² Railroad regulation had been desirable for the railroads and their more politically influential shippers in the nineteenth century, for example, because railroads were the dominant mode of transportation and regulators could protect profits and cross-subsidies by limiting entry and preventing rate wars among railroads. The regulators' ability to protect railroad profits and cross-subsidies was undermined in the twentieth century, however, by the development of highway and air competition. As the demand for railroad services weakened, the benefits to special interests from regulation declined, those interests withdrew their support, and railroad regulation was abandoned.

The argument that changes in industry costs or demand could make regulation less attractive to special interests could not be the whole explanation, however, since it seemed to fit developments better in some industries than in others. In the U.S. trucking, airline, and telecommunications industries, for

example, regulation was abandoned before the returns to the special interests thought to benefit from regulation had been dissipated.¹³ And in the case of the U.S. railroads, the returns to regulation had been dissipated decades before deregulation. If the railroads had been earning little or no profits since the early 1950s, why was deregulation delayed until the end of the 1970s?

A second response to the deregulation movement was to acknowledge that competition among special interests could encourage policies that favored broader public interests. Gary Becker, another Chicago economist, pointed out that the introduction of regulation to favor special interests in markets that were otherwise efficient would be discouraged by the fact that the regulations created inefficiencies.¹⁴ The inefficiencies meant that the benefits received by the favored interests would be less than the costs imposed on the other interests. The excess of costs over benefits gave the opponents of regulation a stronger incentive to mobilize than the proponents, all else being equal, thereby limiting the extent to which regulation was likely to be introduced solely as a means to redistribute wealth from one interest group to another. Theodore Keeler extended Becker's argument to markets that were not efficient in the absence of regulation, such as those characterized by natural monopoly.¹⁵ In such markets the introduction of properly designed regulation could correct the inefficiency, which meant that the regulation's benefits would be greater than its costs. The excess of benefits over costs made it easier to form coalitions that favored the efficient corrective regulation of monopoly, again all else being equal.

Unfortunately, Becker and Keeler's models provided little guidance regarding the circumstances under which competition among special interests would lead to the public interest. Although economically efficient regulatory policies produced benefits in excess of costs, that advantage might not be enough to overcome other factors, such as the tendency for the regulated companies to be smaller in number and to have larger stakes in the outcome of regulatory policy than their consumers. It might be easier to organize coalitions that favored efficient regulatory policy all else being equal, but all else was seldom equal.

The most significant response to the deregulation movement was to recognize that the models of interest group competition advanced by the Chicago economists were probably too simplistic. In the first place, they assumed that the principal actors—the interest groups—could effectively control the behavior of the agents acting on their behalf—the voters, the regulators, and the regulated firms. Social scientists had become intrigued with the difficulties that principals face in controlling their agents if the principals' and agents' interests don't coincide and the principals have trouble monitoring and second-guessing the agents' actions. Regulation seemed to be a classic

"principal-agent" problem. As the political scientist James Q. Wilson explained:

To say that firms prefer higher profits, politicians more votes, and bureaucrats larger incomes is to make an important but incomplete assertion. If we wish to explain public policy by reference to such preferences, we must be able to say more—to show that policies are made so that profit-seeking firms can affect the votes won by vote-hungry politicians who will in turn constrain the behavior of money-hungry (or power-hungry, or status hungry) bureaucrats, whose behavior will in turn affect the profitability of firms. It is a long and complex causal chain.¹⁶

Moreover, the Chicago economists generally assumed that the regulatory agency (or the regulated industry) was ultimately supervised by a single political master: the politician who won the most votes. In fact, regulatory agencies were typically subject to supervision from a variety of different political bodies, including the elected executive, the different chambers of the legislative branch, and often several different committees within each chamber. This diffusion of oversight responsibility provided opportunities for a wider variety of politicians or interest groups to try to affect the behavior of the agency, but probably made it harder for any one group to exercise decisive influence. Politicians might be attentive when the agency was initially established or during crises, but the difficulties of exercising oversight and the diffusion of responsibility meant that the agency's activities were often largely unsupervised.¹⁷

These observations helped to stimulate a group of economists, led by Jean-Jacques Laffont and Jean Tirole of the University of Toulouse, to develop mathematical models of the relationships between the principals and their agents in the regulatory process. The central feature of these models is that the principals have imperfect information about the capabilities or performance of their agents. Some models incorporate several layers of principals and agents (for example, voters, politicians, regulators, and regulated firms), while others recognize that an agent might be supervised by more than one principal. Although the models have generated interesting suggestions for reducing principal-agent problems, the yield in new insights about how regulatory agencies behave has been modest so far.¹⁸

More generally, however, recent scholarship tends to support Peltzman's argument that regulators often draw support from a variety of sources instead of being captured by a single interest, and that these patterns of support change over time.¹⁹ In a series of studies of the railroad industry, for example, later scholars argued that Kolko had understated the roles of agricultural and mercantile interest groups in the establishment of the ICC and state regula-

tory agencies.²⁰ Railroad regulation seems to have originated with a broad base of support, including consumers as well as producers. Once an agency is established, however, the political forces that continue to focus on it may be narrower than those that created it. And the political bargain may have to change over time as, for example, changes in technology alter interests or events shape popular perceptions. Regulatory systems seem to have enormous inertia, so that pressures may have to build for a while before significant changes occur. Ultimately, however, the agency's performance has to be reasonably consistent with what a preponderance of interests will support or its statute will be revised.

Another idea that has survived is that whether the costs and benefits of regulation are narrowly concentrated or widely distributed helps determine which interests will influence regulatory policy and whether broader public interests are pursued. James Q. Wilson distinguishes four different patterns. When both the costs and the benefits of a regulatory issue are widely distributed, then interest groups will have little incentive to form around the issue, and what Wilson calls "majoritarian" politics will result. When both costs and benefits are highly concentrated, however, the result is "interest group politics." In this case, the affected interest groups have strong incentives to mobilize and battle for influence, and the general public is not likely to be heard from. Concentrated benefits and widely dispersed costs generate "client" politics, in which the group that benefits will dominate policy, since it has a strong incentive to mobilize while the general public does not. Finally, concentrated costs and widely distributed benefits generate what Wilson calls "entrepreneurial" politics. In such cases, Wilson hypothesizes that the regulatory policy will be adopted only if a skilled political entrepreneur can mobilize latent public sentiment and put the opponents on the defensive.²¹

As Wilson's "entrepreneurial" category indicates, recent research also suggests that interest group pressures alone do not shape regulatory policy, and that individuals and ideas can play an important role. If a regulatory agency is subject to pressures from a variety of interest groups with roughly similar strength and conflicting positions, for example, then regulators and their staff may have room to advance their own ideas about what might be best. And some of their ideas may be of the public interest variety. The need to explain government policy to the public also can constrain interest group influence and increase the role of ideas and ideology.²² Some researchers argue, for example, that individual political or bureaucratic entrepreneurs, armed with the scholarship critical of regulation that had developed since the 1950s, played an important role in stimulating the deregulation of various transportation and telecommunications industries in the United States in the late 1970s and early 1980s.²³

Institutional Protections

The Institutional Perspective

Yet another perspective is that the risk of capture by special interests depends on the design of both the regulatory agency and the other broader institutions that shape the interactions between private citizens and government. The study of institutions was fashionable among political scientists and economists in the first half of the twentieth century and has come into favor again. Many of the more recent studies are rooted in economics, and particularly in the work on transaction costs by Ronald Coase, Oliver Williamson, and Douglass North and others mentioned in the previous chapter.²⁴ From this perspective, institutions are defined as the formal and informal rules that govern the ways in which individuals and organizations interact with each other. Thus institutions include a nation's constitution, laws, and political system as well as its more informal customs or social conventions. Institutions affect the transaction costs of various types of activities by influencing, among other things, the exposure of parties to opportunism and the defenses they can use against it.

The proponents of this perspective argue that institutions change in the long run but tend to be stable in the short run. Institutions evolve in response to political, economic, and other forces. In the long run they are a product of their larger environment, and cannot be too much at odds with it. But institutions are slow to change in part because their stability is advantageous to society. Stable institutions make the behavior of individuals and organizations more predictable, and thus make it easier for individuals and organizations to engage in cooperative activity. And because institutions evolve only gradually, they shape and constrain the behavior of individuals and organizations.

The Design of Regulatory Agencies

During the first half of the twentieth century, U.S. reformers devoted a great deal of effort to the design of regulatory institutions that would resist capture by special interests. These schemes typically tried to balance independence from elected politicians in the executive and legislative branches of government with some form of direct accountability to the larger public and to the regulated companies and their customers. In effect these schemes replaced the normal process for political decision making in a democracy with alternative forms of accountability.

The primary measure for promoting independence was to appoint regula-

tory commissioners to long terms and make them removable only for specific and limited causes, such as corruption. In many cases commissioners were also appointed to staggered terms so that it would take many years before the elected officials who appointed them could name a majority on the commission. In some cases the regulatory agency's budget was removed from the normal process of legislative and executive review so as to make it harder for elected officials to retaliate against the commission by cutting its funding.

A variety of measures were used to promote public accountability. Most PUCs had five to seven commissioners instead of one, for example, in order to make it harder to keep corrupt agreements secret. Many of the commissions were also required to hold public hearings and to deliberate on the record in an effort to further limit the risk of secret influence.²⁵ Similarly, the U.S. Congress imposed a variety of institutional constraints on the federal regulatory agencies created in the New Deal in response to complaints that the agencies were abusing their powers over the firms and individuals they regulated. Under the Administrative Procedure Act of 1946, Congress specified that a federal regulatory agency could issue a rule only if it (1) gave adequate public notice of its intention to issue the rule, (2) provided interested parties an opportunity to participate through the written or oral submission of arguments and data, and (3) provided a written explanation of the basis for the final rule adopted. If an individual case was being considered rather than an industrywide rule, the procedural protections were even stricter. The idea was that regulatory agencies had a quasi-judicial function, and thus their procedures ought to embody the principles of fairness that had developed in common law and were enshrined in the Constitution.²⁶ Many states subsequently adopted similar constraints.

The effectiveness of these institutional strategies came into doubt in the second half of the twentieth century, as scholarly criticism of the regulatory agencies grew. The very idea that a politically independent regulator could best discover the public interest was challenged by Sam Huntington in a 1952 study of the ICC. Huntington argued that the independence designed to allow the ICC to discover the public interest had instead made its capture by special interests possible.²⁷ The railroads had captured the ICC after World War I and then used the agency to suppress competition among railroads, trucks, and barges. Huntington viewed the results as so contrary to the public interest that he proposed that the ICC be abolished as an independent commission and replaced by a government agency that would be directly accountable to Congress and the president and, through them, to the broader electorate. While most scholars were reluctant to recommend such a radical step, Huntington's study made it clear that a regulator's political independence had costs as well as benefits.

In the 1960s and 1970s, scholars began to argue that some of the specific regulatory procedures adopted to promote accountability also made it easier for the regulated firm to capture the regulator.²⁸ Requirements for public hearings, oral testimony, written testimony, and rebuttal made arguing a case before a regulator a long and elaborate process, which gave an advantage to interest groups that could mobilize the resources needed to participate on a sustained basis. And even if the regulator was not captured, the lengthy procedures contributed to the impression that the regulatory agencies were lethargic and out of touch. These criticisms encouraged efforts to streamline the regulatory process and to provide assistance to help the general public and small firms participate.

A less pessimistic view emerged during the 1980s, in which administrative procedures were seen as a device to help politicians maintain control over the regulatory agencies they created.²⁹ Politicians faced a classic principal-agent problem, because the regulatory bureaucrats might not share their preferences or priorities. The politicians could not rely on the enabling statute to control the regulators' behavior, since it was difficult to draft a statute that anticipated all the problems that might develop. And effective oversight was difficult, because it was costly for the politicians to find out what the agency was doing and the sanctions they could apply were limited. The Administrative Procedure Act aided in control because the requirements for notice, comment, and hearings all generated, at no cost to the politicians, much of the information they needed for oversight. Moreover, the politicians could design procedures that favored certain interest groups over others by specifying who bore the burden of proof, who had standing to intervene, and who could initiate a regulatory change. The intent, these scholars argued, was to mandate a set of procedures that reproduced the balance of power among interests when the regulatory agency was established, so that the agency's subsequent decisions would be consistent with the political intent of its creators.

The Broader Institutional Environment

More recent scholarship has focused on the broader institutional environments in which regulatory agencies operate rather than on regulatory agencies or their procedures. A major theme of this research is that governments are less likely to be able to behave opportunistically toward private individuals in political systems where power is divided and there are checks and balances among government institutions. In a famous article, for example, Douglass North and Barry Weingast argued that England's economic development was facilitated by the Glorious Revolution of 1688 because the rebellion led to the establishment of Parliament as a counterweight to the monarchy.³⁰

Parliament limited the discretion of the king to expropriate private property through taxation and other means, and the security of property in turn led to an expanded private capital market that was better able to finance both the wars of the state and the coming industrial revolution. North and Weingast suggested that the lack of a similar check against royal power in France retarded the development of French capital markets and contributed to the relative decline of French power in the eighteenth century.

A related theme is that legal systems are critical for modern economic development, because they reduce the costs of enforcing contracts and protect private property against theft or expropriation. Legal scholars classify the legal systems of countries into different traditions or families. These classifications do not depend on the specific rules in the country's laws but rather on the methods "through which the rules to be applied are discovered, interpreted and evaluated."³¹ The two dominant legal families in the modern world are English common law and the European civil law.³² Britain and its former colonies, including the United States and Canada, have common law systems; the European countries and their former colonies, including Latin America, generally have civil law systems.³³

Recent research suggests that common law systems provide more protection for individual property rights than civil law systems. In a series of papers, Rafael La Porta, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny compared, across a large sample of countries, the size of the markets for private corporate debt and equity and the extent to which major companies are owned by small stockholders or are closely held by families or large institutions.³⁴ They classified the countries according to whether they belonged to the civil or common law traditions. In addition, they looked for the presence of specific legal rules designed to protect shareholders against management, such as provisions that give investors one vote for each share they own or that allow a small minority of shareholders to force a special shareholder meeting. Their statistical analyses demonstrate that the specific legal rules matter: countries that have rules to protect shareholders tend to have larger debt and equity markets and more widespread stock ownership. But these investigators also found that the legal traditions have an independent effect. Even with the same specific rules, common law countries tend to have deeper capital markets and more small shareholders than civil law countries.

Civil law systems may provide less protection for private property and individual liberties, because civil law judges are supposed to only interpret the letter of the law and not go beyond it. A judge can intervene only if she finds a violation of a specific statute, which gives the executive and legislative branches of government, who write the statutes, more control over the scope of the judiciary's activities. In common law systems, by contrast, judges are

not restricted to the statutes but can base decisions on the "common law" of case precedents and past decisions as well. The ability to appeal to a common law that has evolved over centuries provides an independent check on the powers of the executive and the legislature. It is no accident that the common law system evolved out of popular opposition to the English monarchy while civil law assumes that the state is the ultimate arbiter in society.

One implication of this research is that contractual and discretionary regulation may be more viable in some types of political and legal systems than in others. Political institutions that have checks and balances should provide stronger regulatory commitment, for example, because they make it harder for government to change the rules of the regulatory game. Checks and balances that limit the arbitrary use of government powers should also help discourage special interests from abusing the powers of a discretionary regulatory agency. And legal institutions that provide strong protection for private property should protect private contracts and concession contracts as well.

Very little research has been done to confirm whether or how these broader political and legal institutions affect the performance of regulatory agencies. One exception is a study by Brian Levy and Pablo Spiller of the history of private telephone companies in Argentina, Britain, Chile, Jamaica, and the Philippines.³⁵ Telephone systems are usually too complex to be regulated through a concession contract system, so all five of the countries used some form of discretionary regulation during the periods when the telephone companies were privately owned. In most countries and during most years, however, the regulatory scheme also included contractual elements that constrained the regulator. For example, regulators were obligated by statute to provide a specific minimum rate of return to telephone companies in Chile from 1958 to 1970 and in Jamaica before 1966 and after 1987. Similarly, Britain after 1984 and Chile after 1987 used a price-cap scheme, in which the regulator is allowed to adjust the tariff formula only every five years.

According to Levy and Spiller, the experience of the private telephone companies suggests that private investment is more likely with a discretionary regulatory scheme that has contractual elements and in countries that have an independent judiciary and a presidential rather than a parliamentary system of government. The authors' reasoning is that private investment is more likely if investors believe the government is committed to reasonably clear rules. Statutory provisions that limit the regulator's discretion help make the regulator's decisions more predictable. An independent judiciary is needed to enforce the statutory constraints on the regulator. And a presidential system makes it harder to change the statute, since the executive and the legislature are often controlled by different political parties.

Levy and Spiller's results offer some support for the advantage of an inde-

pendent judiciary but less for the advantages of statutory constraints on discretion or a presidential system. Three of the five countries studied have enjoyed sustained periods of private investment, and they are the three that Levy and Spiller classify as having independent judiciaries (Britain, Chile, and Jamaica). But only one of the three countries with sustained investment has a presidential system; the other two have parliamentary governments (Britain and Jamaica).³⁶ Moreover, the three countries with presidential systems (Argentina, Chile, and the Philippines) seemed to change regulatory regimes roughly as often as the two countries with parliamentary systems. All five of the countries imposed fairly specific statutory constraints on the regulators, the only exception being Jamaica between 1966 and 1987. Although the Jamaican telephone investment suffered during the 1966 to 1987 period, it would be difficult to conclude from that one case that detailed statutory constraints on the regulator are highly beneficial.³⁷

Levy and Spiller's study has inspired efforts by other researchers to examine the relationship between regulatory performance and the broader institutional environment using larger samples of countries and regulatory systems. Assembling data sets has been difficult, because most countries outside the United States have only a decade or so of recent experience with private utilities and because a large number of variables are needed to make cross-country comparisons.³⁸ But this approach is likely to yield some interesting results.

The Chances of Capture

In sum, the chances that a regulatory agency will be captured seem to depend on a variety of factors, not all of which are well understood. One factor is the pattern and visibility of the costs and benefits of the actions the agency might take. The traditional fear is that the agency will be captured by the companies it regulates, since the benefits of tariff increases are concentrated in the companies while the costs are often spread thinly over many customers. But utilities often have large customers as well as small ones, and those large customers may mobilize to protect their investments. Moreover, many utility tariffs are highly visible, which may make it easier to mobilize many small consumers against tariff increases even though each customer bears only a small share of the costs.

Institutions also appear to be important, but our understanding of when and why is fairly limited. Institutional differences probably explain why regulatory agencies seem more likely to pursue the public interest soon after their creation or after major restructuring. The creation or restructuring of an agency usually requires a change in legislation, which brings the issue into a more open forum where a broader coalition may be needed to effect change.

The routine business of a regulatory agency does not require legislation, however, which increases the possibility for the agency to be captured by special interests over time, or possibly to escape effective supervision altogether.

The design of the regulatory agency also matters, although here we have learned that the choices are not always easy. Measures that insulate regulatory agencies from day-to-day political pressures make it easier to make decisions that are unpopular in the short run, including decisions that are not in the public interest. And procedures designed to make regulatory agencies accountable to particular coalitions or to the general public may have unintended and undesired effects as well. Provisions to ensure that affected parties have opportunities to be heard and participate, for example, may impart such a strong bias toward the status quo that the regulatory system cannot cope with change.

Finally, the broader institutional environment of the regulatory agency is probably influential as well. Legal systems and other institutions that protect private property may also help protect utility investors from abuse by regulatory agencies. And political systems with checks and balances may make it harder to change regulatory regimes. How strong these influences are is far from clear, however. Moreover, our understanding of the effects of the broader institutional environment is still too primitive to help us choose among regulatory strategies. The institutional environments that seem helpful to private contracts or concession contracts seem to help discretionary regulation as well.

Capture and Instability: Sri Lanka's Buses and U.S. Telephones

Capture as a Source of Instability

Regulatory regimes can change, as the nationalization of regulated utilities in many countries after World War II and the deregulation of some U.S. industries in the 1970s and 1980s demonstrate. Many of the explanations for change recounted in the previous chapter credit technological and other forces exogenous to the regulated industry. The development of highway and air transportation during the twentieth century weakened the monopoly power of U.S. railroads, for example, and made continued regulation either less necessary (from the public interest perspective) or less lucrative (from the special interest perspective). But there is also a tradition, dating back at least to Marver Bernstein's 1955 book on the life cycle of a regulatory agency, that the forces for change may be partly endogenous; in other words, that regulation contains the seeds of its own decay or destruction.¹

With discretionary regulatory regimes, the accounts of endogenous change and instability all involve the capture of the regulatory agency by special interests. The simplest but most compelling versions are based on the observation that the special interests that capture the regulatory agency benefit at the expense of others. Gary Becker noted that the benefits to the agency's captors are inevitably less than the costs imposed on others, since the policies pursued by the captured agency create market inefficiencies.² Becker reasoned that this pattern of costs and benefits made it risky for a regulatory agency to pander too much to special interests. Becker had in mind that excessive pandering might mobilize the losers to overthrow the agency's policies. But the losers might also have the option of exiting the regulatory system, so as to avoid bearing its costs. Moreover, the losers' ability to exit increases over time as any durable investments that are specific to the regulatory regime wear out and have to be replaced. In the long run, the losing interests may be able to withdraw enough to undermine the benefits received by the regulator's captors, so that the regulatory regime collapses.