GOVERNING the COMMONS



ELINOR OSTROM

The Evolution of Institutions for Collective Action

Political Economy
of Institutions and Decisions

THE POLITICAL ECONOMY OF INSTITUTIONS AND DECISIONS

Editors
James E. Alt, Harvard University
Douglass C. North, Washington University in St. Louis

Other books in the series
James E. Alt and Kenneth Shepsle, eds., *Perspectives on Positive Political Economy*
Yoram Barzel, *Economic Analysis of Property Rights*
Robert Bates, *Beyond the Miracle of the Market: The Political Economy of Agrarian Development in Kenya*
Gary W. Cox, *The Efficient Secret: The Cabinet and the Development of Political Parties in Victorian England*
Leif Lewin, *Ideology and Strategy: A Century of Swedish Politics* (English Edition)
Gary Libecap, *Contracting for Property Rights*
Matthew D. McCubbins and Terry Sullivan, eds., *Congress: Structure and Policy*
Douglass C. North, *Institutions, Institutional Change, and Economic Performance*
Charles Stewart III, *Budget Reform Politics: The Design of the Appropriations Process in the House of Representatives, 1865–1921*

GOVERNING THE COMMONS

The evolution of institutions for collective action

ELINOR OSTROM
Indiana University

CAMBRIDGE
UNIVERSITY PRESS
1990

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Cambridge University Press 1990

This book is in copyright. Subject to statutory exception and
to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1990
Reprinted 1991, 1992, 1993, 1994 (twice), 1995, 1996, 1997,
1998, 1999, 2000, 2001, 2002, 2003

Printed in the United States of America

Typeset in Sabon

A catalog record for this book is available from the British Library

Library of Congress Cataloging in Publication data is available

ISBN 0 521 40599 8 paperback

Contents

Series editors' preface *page* xi
Preface xiii

CHAPTER 1
REFLECTIONS ON THE COMMONS 1

Three influential models 2
The tragedy of the commons 2
The prisoner's dilemma game 3
The logic of collective action 5
The metaphorical use of models 7
Current policy prescriptions 8
Leviathan as the "only" way 8
Privatization as the "only" way 12
The "only" way? 13
An alternative solution 15
An empirical alternative 18
Policy prescriptions as metaphors 21
Policies based on metaphors can be harmful 23
A challenge 23

CHAPTER 2
AN INSTITUTIONAL APPROACH TO THE STUDY OF SELF-ORGANIZATION AND SELF-GOVERNANCE IN CPR SITUATIONS 29

The CPR situation 30
CPRs and resource units 30
Rational appropriators in complex and uncertain situations 33

Interdependence, independent action, and collective action 38
The theory of the firm 40
The theory of the state 41
Three puzzles: supply, commitment, and monitoring 42
The problem of supply 42
The problem of credible commitment 43
The problem of mutual monitoring 45
Framing inquiry 45
Appropriation and provision problems 46
Multiple levels of analysis 50
Studying institutions in field settings 55

CHAPTER 3
ANALYZING LONG-ENDURING, SELF-ORGANIZED, AND SELF-GOVERNED CPRs 58

Communal tenure in high mountain meadows and forests 61
Törbel, Switzerland 61
Hirano, Nagaike, and Yamanoka villages in Japan 65
Huerta irrigation institutions 69
Valencia 71
Murcia and Orihuela 76
Alicante 78
Zanjera irrigation communities in the Philippines 82
Similarities among enduring, self-governing CPR institutions 88
Clearly defined boundaries 91
Congruence between appropriation and provision rules and local conditions 92
Collective-choice arrangements 93
Monitoring 94
Graduated sanctions 94
Conflict-resolution mechanisms 100
Minimal recognition of rights to organize 101
Nested enterprises 101

CHAPTER 4
ANALYZING INSTITUTIONAL CHANGE 103

The competitive pumping race 104
The setting 104
The logic of the water-rights game 106
The litigation game 111

The Raymond Basin negotiations 111
The West Basin negotiations 114
The Central Basin litigation 123
Conformance of parties to negotiated settlements 125
The entrepreneurship game 127
Reasons for forming a district to include both basins 130
Reasons against forming a district to include both basins 131
The polycentric public-enterprise game 133
The analysis of institutional supply 136
Incremental, sequential, and self-transforming institutional change in a facilitative political regime 137
Reformulating the analysis of institutional change 139

CHAPTER 5
ANALYZING INSTITUTIONAL FAILURES AND FRAGILITIES 143

Two Turkish inshore fisheries with continuing CPR problems 144
California groundwater basins with continuing CPR problems 146
A Sri Lankan fishery 149
Irrigation development projects in Sri Lanka 157
The fragility of Nova Scotian inshore fisheries 173
Lessons to be learned from comparing the cases in this study 178

CHAPTER 6
A FRAMEWORK FOR ANALYSIS OF SELF-ORGANIZING AND SELF-GOVERNING CPRs 182

The problems of supply, credible commitment, and mutual monitoring 185
A framework for analyzing institutional choice 192
Evaluating benefits 195
Evaluating costs 198
Evaluating shared norms and other opportunities 205
The process of institutional change 207
Predicting institutional change 210
A challenge to scholarship in the social sciences 214

Notes 217
References 245
Index 271

Series editor' preface

The Cambridge series on the Political Economy of Institutions and Decisions is built around attempts to answer two central questions: How do institutions evolve in response to individual incentives, strategies, and choices, and how do institutions affect the performace of political and economic systems? The scope of the series is comparative and historical rather than international or specifically American, and the focus is positive rather than normative.

In this pioneering book Elinor Ostrom tackles one of the most enduring and contentious questions of positive political economy, whether and how the exploration of common-pool resources can be organized in a way that avoids both excessive consumption and administrative cost. These cases, where a resource is held in common by many individuals – that is, well-defined individual property rights over the resource are absent – are often held by economists to be exploitable only where the problem of over-consumption is solved by privatization or enforcement imposed by outside force. Ostrom, by contrast, argues forcefully that other solutions exist, and that stable institutions of self-government can be created if certain problems of supply, credibility, and monitoring are solved. She provides a close study of a uniquely broad range of cases, including high mountain meadows in Japan and Switzerland, water projects in the Philippines and California, and fisheries in Canada and Turkey. Some of these cases involve stable institutions; in other cases the institutions were fragile and failed. Basing her conclusions on comparisons of sources of success and failure in self-government, Ostrom describes some fundamental characteristics of successful common-pool management schemes, and concludes with a challenge to other social scientists to build on her original theoretical work.

Preface

It is difficult to say when I began work on this study. If one asks when I first began to study problems of collective action faced by individuals using common-pool resources, then identifying the beginning is easier. In the early 1960s, I took a graduate seminar with Vincent Ostrom, who was to become my closest colleague and husband. The seminar focused on the development of institutions related to water resources in southern California. I began my dissertation focusing on the entrepreneurship involved in developing a series of public enterprises to halt the process of saltwater intrusion into a groundwater basin underlying a portion of the Los Angeles metropolitan area. A fellow graduate student, Louis Weschler, conducted a parallel study in an adjacent groundwater basin that adopted different institutional arrangements to cope with similar problems. As Weschler and I completed our studies, it appeared that both institutional arrangements had been successful in enabling the water producers to avoid the catastrophic economic loss that would have occurred if both basins had been inundated by the Pacific Ocean (E. Ostrom 1965; Weschler 1968).

In the late 1960s, Vincent and I participated in the Great Lakes Research Program initiated by the Batelle Memorial Institute (V. Ostrom and E. Ostrom 1977b), but most of my work as a young faculty member focused on problems of urban service delivery and public economies in metropolitan areas. In 1981 I was asked by Paul Sabatier, a colleague for a year at the Center for Interdisciplinary Research at the University of Bielefeld, to make a seminar presentation on "organizational learning." I used as my example of organizational learning the set of rules that groundwater producers had developed in the southern California groundwater basins. Paul then wanted to know why I was so confident that the systems I had studied

15 years earlier were still operating and performing well. At the time, I had no effective answer other than that the institutions had been so well crafted to fit local circumstances that I presumed they had survived and were faring well.

When I returned from Bielefeld, I suggested to one of my doctoral students, William Blomquist, that he answer Sabatier's question as his dissertation. Blomquist (1987b) found that the institutions that the water producers themselves had designed were still in place and operating effectively. The physical condition of the basins had improved substantially. The very substantial "success" involved in these cases led us to undertake a study, funded by the U.S. Geological Survey (grant number 14-08–0001-G1476), of a larger set of groundwater basins in southern California and a limited set in northern California to ascertain what factors were associated with the successful evolution of new institutions and with the efficiency and equity of those institutions. Eventually, we will have completed a comparative study of institutional, economic, and physical changes in 12 groundwater basins over a 30–50-year period.

Although I have been excited about what one can learn from a concentrated effort to study a dozen groundwater basins and the institutions that have evolved for their governance and management over time, such studies alone are not sufficient for the development of a broader theory of institutional arrangements related to the effective governance and management of common-pool resources (CPRs). One needs similar information from many other settings to begin to gain the empirical base necessary to improve our theoretical understanding of how institutions work and how individuals change their own institutions.

My awareness of the possibility of using detailed case studies written by other authors to obtain a sufficiently rich empirical base for understanding CPRs came about as a result of joining the National Academy of Sciences' "Panel on Common Property Resource Management" in 1985. By the time I was asked to join the panel, its members had commissioned a series of papers to be written by field researchers. The authors were all asked to organize their papers using a framework prepared by Ronald Oakerson (1986). That meant that all of the papers would address not only the physical properties of the resource systems but also what types of rules were used to regulate entry and use of these systems, what types of interactions resulted, and what types of outcomes were obtained. The papers were presented at an international conference in Annapolis and were published by the National Academy Press (National Research Council 1986). Some of those papers and some new chapters have been brought together in a new volume (Bromley in press).

Reading those studies, as well as some of the studies cited by those authors, made me aware of two major facts: First, an extraordinarily rich case-study literature already existed, written by field researchers who had invested years of effort in obtaining detailed information about the strategies adopted by the appropriators of CPRs and the rules they used. Second, that literature had been written by authors in diverse fields and frequently had appeared in obscure publications. Almost no syntheses of the findings from that literature had been undertaken.

Several colleagues at Indiana University began to collect citations to relevant cases, and within a short time Fenton Martin, who compiled the resultant bibliography, had identified nearly 1,000 cases. More recently, the number was approaching 5,000 (Martin 1989). The disciplines represented in the bibliography include rural sociology, anthropology, history, economics, political science, forestry, irrigation sociology, and human ecology; included also are area studies, such as African studies, Asian studies, West European studies, and so forth. Scholars had cited primarily studies conducted by others in their own disciplines and perhaps others focusing on the same resource sector or geographic region. Few citations had come from outside each author's disciplinary, sectoral, or regional frame of reference. Consequently, a vast amount of highly specialized knowledge had been accumulated without much synthesis or application of the knowledge to the policy problems involved.

Given the importance of understanding how institutions help users cope with CPR problems, and given the existence of a rich theoretical literature concerning these problems, it seemed to me that it was important to use these case studies as an empirical basis for learning more about the effects of institutions on behaviors and outcomes in diverse field settings. With the help of a grant from the U.S. National Science Foundation (grant number SES 8619498), several colleagues and I have been able to gather many of these cases into an archive. We have systematically screened these cases and selected a much smaller subset for further scrutiny, coding, and analysis. Our selection criteria required that the case be written as a result of extended fieldwork and that information be provided about (1) the structure of the resource system, (2) the attributes and behaviors of the appropriators, (3) the rules that the appropriators were using, and (4) the outcomes resulting from the behaviors of the appropriators. We have now developed a structured coding form that enables us to transform the in-depth qualitative data into a structured data base amenable to quantitative analysis.

The development of the coding forms was itself an exercise in theory development. We used the method of institutional analysis that had grown

out of our earlier work (E. Ostrom 1986a,b) as the organizing framework for the design of these coding forms. In addition, we paid serious attention to the hypotheses stated by field researchers who had conducted multiple studies or were themselves reviewing findings from multiple studies. We tried to include ways of measuring their concepts and proposed relationships in our coding forms. Because we were working with qualitative data, most of our concepts had to be formulated as variables with ordinal or nominal values. Some years of hard work were required simply to read sufficient numbers of cases, study earlier efforts to synthesize findings from specialized fields, and develop the coding forms.

During this process, several papers were written in an attempt to elucidate a theory that would help us understand the patterns we were beginning to see in reading these diverse materials (Gardner and E. Ostrom 1990; Gardner, E. Ostrom, and Walker 1990; E. Ostrom 1985b, 1987, 1989; Schlager and E. Ostrom 1987; Walker, Gardner, and E. Ostrom 1990). It is my conviction that knowledge accrues by the continual process of moving back and forth from empirical observation to serious efforts at theoretical formulation. This book can thus be viewed as an intermediate "progress report" for an ongoing research effort. Given the complexity of the empirical phenomena being studied and the type of theory that is needed to explain these phenomena, the effort may well continue for another decade.

The stimulus to write this volume came from James Alt and Douglass North after I presented a lecture at Washington University in St. Louis during the fall of 1986. Given that the CPR project was still "in process," I would never have dreamed of writing a book without their continued prodding. When Kenneth Shepsle and James Alt asked me to present a series of lectures at Harvard University, during a semester of sabbatical leave, the die was cast.

Actual work on the manuscript began in January 1988, when again I was fortunate to spend a sabbatical semester at the Center for Interdisciplinary Research at the University of Bielefeld. During that time I participated in a Research Group on Game Theory and the Behavioral Sciences organized by Dr. Reinhard Selten, Department of Economics, University of Bonn. I benefited greatly from the opportunity to participate in that research group. Although only a few game-theoretical examples are used in this book, the way that game theorists think about strategic possibilities in social settings strongly influences the way I analyze the central questions addressed here. Working with Roy Gardner and Franz Weissing on two game-theoretical analyses of CPR situations greatly increased my apprecia-

tion for the power and utility of game theory as a general theoretical tool for scholars interested in studying the consequences of diverse institutions.

The writing of this book was undertaken in tandem with participation in the "Decentralization: Finance and Management Project" sponsored by the Office of Rural and Institutional Development of the Bureau for Science and Technology (ST&RD) of the U.S. Agency for International Development. The challenge of making theoretical ideas relevant for application, the support for fieldwork in Nepal, Bangladesh, and Pakistan, and the opportunity to discuss these ideas with Larry Schroeder, Susan Wynne, Jamie Thomson, Louis Siegel, James Wunsch, Ed Connerley, Jerry Miner, Ken Kornher, and Eric Chetwynd, as well as mission personnel and host-government officials, have been of considerable value to me, and I hope that this volume is of value to the project.

I have also benefited greatly from the opportunity to present lectures based on parts of this work while the manuscript was in process. Besides the lectures at Harvard in April 1988, I have made presentations based on one or more chapters at the following places: the Sociology Department at the University of Bielefeld; the first Udall Lecture at the University of Arizona; a conference on "Democracy and Development" organized by the Sequoia Institute; a Liberty Fund summer series held in Victoria, British Columbia, May 15–20, 1989; and several different occasions at the Workshop in Political Theory and Policy Analysis at Indiana University.

Many people have commented on earlier papers or draft copies of this book, and I am deeply appreciative of their frank and helpful critiques. I hope I have responded adequately to their suggestions. Readers of the whole manuscript included Arun Agrawal, James Alt, Oliver Avens, Fikret Berkes, Elizabeth Case, David Feeny, Roy Gardner, Larry Kiser, Hartmut Kliemt, Robert Netting, Douglass C. North, Vincent Ostrom, Christine Picht, Russell Roberts, Edella Schlager, Jane Sell, Michael Taylor, Norman Uphoff, James Walker, Franz Weissing, and Rick Wilson. Readers of the individual chapters, drawing on their own prior research, included Paul Alexander, Fikret Berkes, William Blomquist, Peter Bogason, Thomas F. Glick, Arthur Maass, Robert Netting, and Norman Uphoff. Readers of prior papers that were drawn on in preparing the manuscript included Wulf Albers, Christi Barbour, William Blomquist, James Coleman, James Cooper, David Feeny, Margaret McKean, Fritz Scharpf, Kenneth Shepsle, Rick Wilson, and James Wunsch. I extend a special note of appreciation to those colleagues who have been associated with the CPR project from the beginning – William Blomquist, Roy Gardner, Edella Schlager, S. Y. Tang, and James Walker – and have spent hours refining concepts, developing models, designing instruments and experiments, and discussing how we

can better search out variables from what we are reading and gathering. The help of Elizabeth Case, associate editor for this series, Sophia Prybylski at Cambridge University Press, and Emily Loose, editor at Cambridge University Press, is gratefully acknowledged. Patty Dalecki has, as always, provided professional editorial and production support that has greatly improved the quality of the manuscript. Her cheerful spirit has relieved pressures on many occasions.

1

Reflections on the commons

Hardly a week goes by without a major news story about the threatened destruction of a valuable natural resource. In June of 1989, for example, a *New York Times* article focused on the problem of overfishing in the Georges Bank about 150 miles off the New England coast. Catches of cod, flounder, and haddock are now only a quarter of what they were during the 1960s. Everyone knows that the basic problem is overfishing; however, those concerned cannot agree how to solve the problem. Congressional representatives recommend new national legislation, even though the legislation already on the books has been enforced only erratically. Representatives of the fishers argue that the fishing grounds would not be in such bad shape if the federal government had refrained from its sporadic attempts to regulate the fishery in the past. The issue in this case – and many others – is how best to limit the use of natural resources so as to ensure their long-term economic viability. Advocates of central regulation, of privatization, and of regulation by those involved have pressed their policy prescriptions in a variety of different arenas.

Similar situations occur on diverse scales ranging from small neighborhoods to the entire planet. The issues of how best to govern natural resources used by many individuals in common are no more settled in academia than in the world of politics. Some scholarly articles about the "tragedy of the commons" recommend that "the state" control most natural resources to prevent their destruction; others recommend that privatizing those resources will resolve the problem. What one can observe in the world, however, is that neither the state nor the market is uniformly successful in enabling individuals to sustain long-term, productive use of natural resource systems. Further, communities of individuals have relied on institutions resembling neither the state nor the market to govern some resource systems with reasonable degrees of success over long periods of time.

We do not yet have the necessary intellectual tools or models to understand the array of problems that are associated with governing and managing natural resource systems and the reasons why some institutions seem to work in some settings and not others. This book is an effort to (1) critique the foundations of policy analysis as applied to many natural resources, (2) present empirical examples of successful and unsuccessful efforts to govern and manage such resources, and (3) begin the effort to develop better intellectual tools to understand the capabilities and limitations of self-governing institutions for regulating many types of resources. To do this, I first describe the three models most frequently used to provide a foundation for recommending state or market solutions. I then pose theoretical and empirical alternatives to these models to begin to illustrate the diversity of solutions that go beyond states and markets. Using an institutional mode of analysis, I then attempt to explain how communities of individuals fashion different ways of governing the commons.

THREE INFLUENTIAL MODELS

The tragedy of the commons

Since Garrett Hardin's challenging article in *Science* (1968), the expression "the tragedy of the commons" has come to symbolize the degradation of the environment to be expected whenever many individuals use a scarce resource in common. To illustrate the logical structure of his model, Hardin asks the reader to envision a pasture "open to all." He then examines the structure of this situation from the perspective of a rational herder. Each herder receives a direct benefit from his own animals and suffers delayed costs from the deterioration of the commons when his and others' cattle overgraze. Each herder is motivated to add more and more animals because he receives the direct benefit of his own animals and bears only a share of the costs resulting from overgrazing. Hardin concludes:

> Therein is the tragedy. Each man is locked into a system that compels him to increase his herd without limit – in a world that is limited. Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons. (Hardin 1968, p. 1,244)

Hardin was not the first to notice the tragedy of the commons. Aristotle long ago observed that "what is common to the greatest number has the least care bestowed upon it. Everyone thinks chiefly of his own, hardly at all of the common interest" (*Politics*, Book II, ch. 3). Hobbes's parable of man in a state of nature is a prototype of the tragedy of the commons: Men

seek their own good and end up fighting one another. In 1833, William Forster Lloyd (1977) sketched a theory of the commons that predicted improvident use for property owned in common. More than a decade before Hardin's article, H. Scott Gordon (1954) clearly expounded similar logic in another classic: "The Economic Theory of a Common-Property Research: The Fishery." Gordon described the same dynamic as Hardin:

> There appears then, to be some truth in the conservative dictum that everybody's property is nobody's property. Wealth that is free for all is valued by no one because he who is foolhardy enough to wait for its proper time of use will only find that it has been taken by another. . . . The fish in the sea are valueless to the fisherman, because there is no assurance that they will be there for him tomorrow if they are left behind today. (Gordon 1954, p. 124)

John H. Dales (1968, p. 62) noted at the same time the perplexing problems related to resources "owned in common because there is no alternative!" Standard analyses in modern resource economics conclude that where a number of users have access to a common-pool resource, the total of resource units withdrawn from the resource will be greater than the optimal economic level of withdrawal (Clark 1976, 1980; Dasgupta and Heal 1979).

If the only "commons" of importance were a few grazing areas or fisheries, the tragedy of the commons would be of little general interest. That is not the case. Hardin himself used the grazing commons as a metaphor for the general problem of overpopulation. The "tragedy of the commons" has been used to describe such diverse problems as the Sahelian famine of the 1970s (Picardi and Seifert 1977), firewood crises throughout the Third World (Norman 1984; Thomson 1977), the problem of acid rain (R. Wilson 1985), the organization of the Mormon Church (Bullock and Baden 1977), the inability of the U.S. Congress to limit its capacity to overspend (Shepsle and Weingast 1984), urban crime (Neher 1978), public-sector/private-sector relationships in modern economies (Scharpf 1985, 1987, 1988), the problems of international cooperation (Snidal 1985), and communal conflict in Cyprus (Lumsden 1973). Much of the world is dependent on resources that are subject to the possibility of a tragedy of the commons.

The prisoner's dilemma game

Hardin's model has often been formalized as a prisoner's dilemma (PD) game (Dawes 1973, 1975).[1] Suppose we think of the players in a game as being herders using a common grazing meadow. For this meadow, there is an upper limit to the number of animals that can graze on the meadow for

a season and be well fed at the end of the season. We call that number L. For a two-person game, the "cooperate" strategy can be thought of as grazing $L/2$ animals for each herder. The "defect" strategy is for each herder to graze as many animals as he thinks he can sell at a profit (given his private costs), assuming that this number is greater than $L/2$. If both herders limit their grazing to $L/2$, they will obtain 10 units of profit, whereas if they both choose the defect strategy they will obtain zero profit. If one of them limits his number of animals to $L/2$, while the other grazes as many as he wants, the "defector" obtains 11 units of profit, and the "sucker" obtains -1. If each chooses independently without the capacity to engage in a binding contract, each chooses his dominant strategy, which is to defect. When they both defect, they obtain zero profit. Call this the Hardin herder game, or Game 1. It has the structure of a prisoner's dilemma game.[1]

The prisoner's dilemma game is conceptualized as a noncooperative game in which all players possess complete information. In noncooperative games, communication among the players is forbidden or impossible or simply irrelevant as long as it is not explicitly modeled as part of the game. If communication is possible, verbal agreements among players are presumed to be nonbinding unless the possibility of binding agreements is explicitly incorporated in the game structure (Harsanyi and Selten 1988, p. 3). "Complete information" implies that all players know the full structure of the game tree and the payoffs attached to outcomes. Players either know or do not know the current moves of other players depending on whether or not they are observable.

In a prisoner's dilemma game, each player has a dominant strategy in the sense that the player is always better off choosing this strategy – to defect

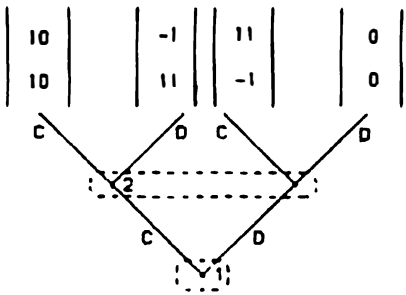


Figure 1.1. Game 1: The Hardin herder game.

– no matter what the other player chooses. When both players choose their dominant strategy, given these assumptions, they produce an equilibrium that is the third-best result for both. Neither has an incentive to change that is independent of the strategy choice of the other. The equilibrium resulting from each player selecting his or her "best" individual strategy is, however, not a Pareto-optimal outcome. A Pareto-optimal outcome occurs when there is no other outcome strictly preferred by at least one player that is at least as good for the others. In the two-person prisoner's dilemma game, both players prefer the (cooperate, cooperate) outcome to the (defect, defect) outcome. Thus, the equilibrium outcome is Pareto-inferior.

The prisoner's dilemma game fascinates scholars. The paradox that individually rational strategies lead to collectively irrational outcomes seems to challenge a fundamental faith that rational human beings can achieve rational results. In the introduction to a recently published book, *Paradoxes of Rationality and Cooperation*, Richmond Campbell explains the "deep attraction" of the dilemma:

> Quite simply, these paradoxes cast in doubt our understanding of rationality and, in the case of the Prisoner's Dilemma suggest that it is impossible for rational creatures to cooperate. Thus, they bear directly on fundamental issues in ethics and political philosophy and threaten the foundations of the social sciences. It is the scope of these consequences that explains why these paradoxes have drawn so much attention and why they command a central place in philosophical discussion.
> (Campbell 1985, p. 3)

The deep attraction of the dilemma is further illustrated by the number of articles written about it. At one count, 15 years ago, more than 2,000 papers had been devoted to the prisoner's dilemma game (Grofman and Pool 1975).

The logic of collective action

A closely related view of the difficulty of getting individuals to pursue their joint welfare, as contrasted to individual welfare, was developed by Mancur Olson (1965) in *The Logic of Collective Action*. Olson specifically set out to challenge the grand optimism expressed in group theory: that individuals with common interests would voluntarily act so as to try to further those interests (Bentley 1949; Truman 1958). On the first page of his book, Olson summarized that accepted view:

> The idea that groups tend to act in support of their group interests is supposed to follow logically from this widely accepted premise of rational, self-interested behavior. In other words, if the members of some group have a common interest or object, and if they would all be better off if that objective were achieved, it has been

thought to follow logically that the individuals in that group would, if they were rational and self-interested, act to achieve that objective. (Olson 1965, p. 1)

Olson challenged the presumption that the possibility of a benefit for a group would be sufficient to generate collective action to achieve that benefit. In the most frequently quoted passage of his book, Olson argued that

unless the number of individuals is quite small, or unless there is coercion or some other special device to make individuals act in their common interest, *rational, self-interested individuals will not act to achieve their common or group interests.* (Olson 1965, p. 2; emphasis in original)

Olson's argument rests largely on the premise that one who cannot be excluded from obtaining the benefits of a collective good once the good is produced has little incentive to contribute voluntarily to the provision of that good. His book is less pessimistic than it is asserted to be by many who cite this famous passage. Olson considers it an open question whether intermediate-size groups will or will not voluntarily provide collective benefits. His definition of an intermediate-size group depends not on the number of actors involved but on how noticeable each person's actions are.

The tragedy of the commons, the prisoner's dilemma, and the logic of collective action are closely related concepts in the models that have defined the accepted way of viewing many problems that individuals face when attempting to achieve collective benefits. At the heart of each of these models is the free-rider problem. Whenever one person cannot be excluded from the benefits that others provide, each person is motivated not to contribute to the joint effort, but to free-ride on the efforts of others. If all participants choose to free-ride, the collective benefit will not be produced. The temptation to free-ride, however, may dominate the decision process, and thus all will end up where no one wanted to be. Alternatively, some may provide while others free-ride, leading to less than the optimal level of provision of the collective benefit. These models are thus extremely useful for explaining how perfectly rational individuals can produce, under some circumstances, outcomes that are not "rational" when viewed from the perspective of all those involved.

What makes these models so interesting and so powerful is that they capture important aspects of many different problems that occur in diverse settings in all parts of the world. What makes these models so dangerous – when they are used metaphorically as the foundation for policy – is that the constraints that are assumed to be fixed for the purpose of analysis are taken on faith as being fixed in empirical settings, unless external author-

ities change them.[1] The prisoners in the famous dilemma cannot change the constraints imposed on them by the district attorney; they are in jail. Not all users of natural resources are similarly incapable of changing their constraints. As long as individuals are viewed as prisoners, policy prescriptions will address this metaphor. I would rather address the question of how to enhance the capabilities of those involved to change the constraining rules of the game to lead to outcomes other than remorseless tragedies.

THE METAPHORICAL USE OF MODELS

These three models and their many variants are diverse representations of a broader and still-evolving theory of collective action. Much more work will be needed to develop the theory of collective action into a reliable and useful foundation for policy analysis. Considerable progress has been made during the past three decades by theorists and empirically oriented social scientists. The sweeping conclusions of the first variants of this theory have given way to a more qualified body of knowledge involving many more variables and explicit base conditions.

As an evolving, rather than completed, theory, it provokes disagreement regarding the importance or insignificance of some variables and how best to specify key relationships.[4] The results from more recent work, particularly work focusing on the dynamic aspects of relevant empirical settings, have begun to generate more optimistic predictions than did earlier models; see, in particular, the work of Axelrod (1981, 1984) and Kreps and Wilson (1982). This is one of the most exciting areas in the social sciences, for although considerable cumulation has already occurred, some deep questions remain unanswered. Some of these puzzles are key to understanding how individuals jointly using a common-pool resource might be able to achieve an effective form of governing and managing their own commons. These puzzles are examined in Chapter 2.

Much that has been written about common-pool resources, however, has uncritically accepted the earlier models and the presumption of a remorseless tragedy (Nebel 1987). Scholars have gone so far as to recommend that "Hardin's 'Tragedy of the Commons' should be required reading for all students . . . and, if I had my way, for all human beings."[5] Policy prescriptions have relied to a large extent on one of the three original models, but those attempting to use these models as the basis for policy prescription frequently have achieved little more than a metaphorical use of the models.

When models are used as metaphors, an author usually points to the similarity between one or two variables in a natural setting and one or two

variables in a model. If calling attention to similarities is all that is intended by the metaphor, it serves the usual purpose of rapidly conveying information in graphic form. These three models have frequently been used metaphorically, however, for another purpose. The similarity between the many individuals jointly using a resource in a natural setting and the many individuals jointly producing a suboptimal result in the model has been used to convey a sense that further similarities are present. By referring to natural settings as "tragedies of the commons," "collective-action problems," "prisoner's dilemmas," "open-access resources," or even "common-property resources," the observer frequently wishes to invoke an image of helpless individuals caught in an inexorable process of destroying their own resources. An article in the December 10, 1988, issue of *The Economist* goes so far as to assert that fisheries can be managed successfully only if it is recognized that "left to their own devices, fisherman will overexploit stocks," and "to avoid disaster, managers must have effective hegemony over them."

Public officials sometimes do no more than evoke grim images by briefly alluding to the popularized versions of the models, presuming, as self-evident, that the same processes occur in all natural settings. The Canadian minister of fisheries and oceans, for example, captured the color of the models in a 1980 speech:

> If you let loose that kind of economic self-interest in fisheries, with everybody fishing as he wants, taking from a resource that belongs to no individual, you end up destroying your neighbour and yourself. In free fisheries, good times create bad times, attracting more and more boats to chase fewer and fewer fish, producing less and less money to divide among more and more people.
>
> (Romeo LeBlanc, speaking at the 50th anniversary meeting of the United Maritime Fishermen, March 19, 1980; quoted by Matthews and Phyne 1988)

The implication, of course, was that Canadian fisheries universally met that description – an empirically incorrect inference.[6] But many observers have come to assume that most resources are like those specified in the three models. As such, it has been assumed that the individuals have been caught in a grim trap. The resulting policy recommendations have had an equally grim character.

CURRENT POLICY PRESCRIPTIONS

Leviathan as the "only" way

Ophuls (1973, p. 228) argued, for example, that "because of the tragedy of the commons, environmental problems cannot be solved through cooperation . . . and the rationale for government with major coercive

powers is overwhelming." Ophuls concluded that "even if we avoid the tragedy of the commons, it will *only* be by recourse to the tragic necessity of Leviathan" (1973, p. 229; emphasis added).[7] Garrett Hardin argued a decade after his earlier article that we are enveloped in a "cloud of ignorance" about "the true nature of the fundamental political systems and the effect of each on the preservation of the environment" (1978, p. 310). The "cloud of ignorance" did not, however, prevent him from presuming that the only alternatives to the commons dilemma were what he called "a private enterprise system," on the one hand, or "socialism," on the other (1978, p. 314). With the assurance of one convinced that "the alternative of the commons is too horrifying to contemplate" (1968, p. 1,247), Hardin indicated that change would have to be instituted with "whatever force may be required to make the change stick" (1978, p. 314). In other words, "if ruin is to be avoided in a crowded world, people must be responsive to a coercive force outside their individual psyches, a 'Leviathan,' to use Hobbes's term" (Hardin 1978, p. 314).

The presumption that an external Leviathan is necessary to avoid tragedies of the commons leads to recommendations that central governments control most natural resource systems. Heilbroner (1974) opined that "iron governments," perhaps military governments, would be necessary to achieve control over ecological problems. In a less draconian view, Ehrenfeld (1972, p. 322) suggested that if "private interests cannot be expected to protect the public domain then external regulation by public agencies, governments, or international authorities is needed." In an analysis of the problems involved in water resource management in developing countries, Carruthers and Stoner (1981, p. 29) argued that without public control, "overgrazing and soil erosion of communal pastures, or less fish at higher average cost," would result. They concluded that "common property resources *require* public control if economic efficiency is to result from their development" (1981, p. 29; emphasis added).[8] The policy advice to centralize the control and regulation of natural resources, such as grazing lands, forests, and fisheries, has been followed extensively, particularly in Third World countries.

One way to illustrate these proponents' image of centralized control is to modify the Hardin herder game using the assumptions that underlie this policy advice. The proponents of centralized control want an external government agency to decide the specific herding strategy that the central authority considers best for the situation: The central authority will decide who can use the meadow, when they can use it, and how many animals can be grazed. Let us assume that the central authority decides to impose a penalty of 2 profit units on anyone who is considered by that authority to

be using a defect strategy. Assuming that the central agency knows the sustainable yield of the meadow (L) and can unfailingly discover and penalize any herder using the defect strategy, the newly restructured game imposed by the central authority is represented in Game 2. Now, the solution to Game 2 is (cooperate, cooperate). Both players receive 10 profit units each, rather than the zero units they would have received in Game 1. If an external authority accurately determines the capacity of a common-pool resource, unambiguously assigns this capacity, monitors actions, and unfailingly sanctions noncompliance, then a centralized agency can transform the Hardin herder game to generate an optimally efficient equilibrium for the herders. Little consideration is given to the cost of creating and maintaining such an agency. This is seen as exogenous to the problem and is not included as a parameter of Game 2.[9]

The optimal equilibrium achieved by following the advice to centralize control, however, is based on assumptions concerning the accuracy of information, monitoring capabilities, sanctioning reliability, and zero costs of administration. Without valid and reliable information, a central agency could make several errors, including setting the carrying capacity or the fine too high or too low, sanctioning herders who cooperate, or not sanctioning defectors. The implications of all forms of incomplete information are interesting. However, as an example, I shall focus entirely on the implications arising from a central agency's incomplete information about the herders' strategies. The implicit assumption of Game 2 is that the central agency monitors all actions of the herders costlessly and imposes sanctions correctly.

In Game 3, we assume that the central agency has complete information about the carrying capacity of the meadow, but incomplete information about the particular actions of the herders. The central agency conse-

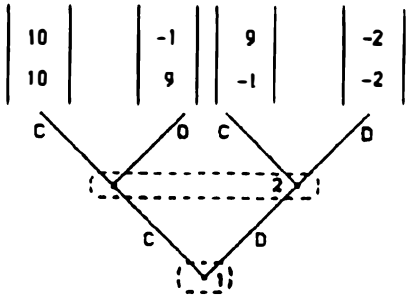


Figure 1.2. Game 2: The central-authority game with complete information.

quently makes errors in imposing punishments. Let us assume that the central agency punishes defections (the correct response) with probability y and fails to punish defections with probability $1 - y$ (the erroneous response). Let us also assume that the central agency punishes cooperative actions (the erroneous response) with probability x and does not punish cooperative actions (the correct response) with probability $1 - x$. The payoff parameters are illustrated in Figure 1.3.

A central agency with complete information would make no errors in its punishment level; in that case, $x = 0$ and $y = 1$. Game 2 would then be a special case of Game 3 in which $x = 0$ and $y = 1$. However, if the central agency does not have complete information about the actions of the herders, it imposes both types of sanctions correctly with a probability of 0.7 ($x = 0.3$, $y = 0.7$). An example of the specific payoffs for this game is shown as Game 4 in Figure 1.4. Given this payoff structure, the herders again face a prisoner's dilemma game. They will defect (overgraze) rather than cooperate (graze within the carrying capacity). In Game 4, as in the original Game 1, the equilibrium outcomes for the herders were (0, 0). In a game in which a central agency sanctions correctly with a probability of 0.7, the equilibrium outcomes are (−1.6, −1.6). The equilibrium of the regulated game has a lower value than that of the unregulated game. Given the carrying capacity and profit possibilities of Game 1, the central agency must have sufficient information so that it can correctly impose sanctions with a probability greater than 0.75 to avoid pushing the herders to the (*D*, *D*) equilibrium.[10]

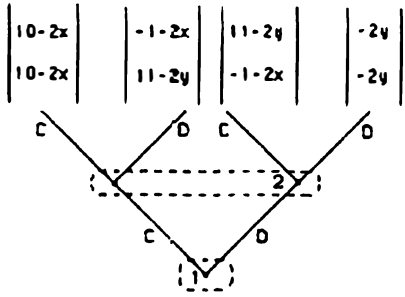


Figure 1.3. Game 3: The central-authority game with incomplete information.

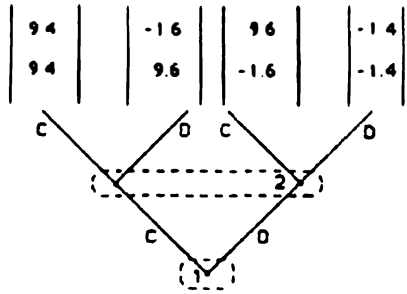


Figure 1.4. Game 4: An example of the central-authority game with incomplete information.

Privatization as the "only" way

Other policy analysts, influenced by the same models, have used equally strong terms in calling for the imposition of private property rights whenever resources are owned in common (Demsetz 1967; O. Johnson 1972). "Both the economic analysis of common property resources and Hardin's treatment of the tragedy of the commons" led Robert J. Smith (1981, p. 467) to suggest that "the *only* way to avoid the tragedy of the commons in natural resources and wildlife is to end the common-property system by creating a system of private property rights" (emphasis added); see also the work of Sinn (1984). Smith stressed that it is "by treating a resource as a common property that we become locked in its inexorable destruction" (1981, p. 465). Welch advocated the creation of full private rights to a commons when he asserted that "the establishment of full property rights is necessary to avoid the inefficiency of overgrazing" (1983, p. 171). He asserted that privatization of the commons was the optimal solution for all common-pool problems. His major concern was how to impose private ownership when those currently using a commons were unwilling to change to a set of private rights to the commons.

Those recommending the imposition of privatization on the herders would divide the meadow in half and assign half of the meadow to one herder and the other half to the second herder. Now each herder will be playing a *game against nature* in a smaller terrain, rather than a game against another player in a larger terrain. The herders now will need to invest in fences and their maintenance, as well as in monitoring and sanctioning activities to enforce their division of the grazing area (B. Field 1984, 1985b). It is presumed that each herder will now choose $X/2$ animals

to graze as a result of his own profit incentive.[11] This assumes that the meadow is perfectly homogeneous over time in its distribution of available fodder. If rainfall occurs erratically, one part of the grazing area may be lush with growth one year, whereas another part of the area may be unable to support X/2 animals. The rain may fall somewhere else the next year. In any given year, one of the herders may make no profit, and the other may enjoy a considerable return. If the location of lush growth changes dramatically from year to year, dividing the commons may impoverish both herders and lead to overgrazing in those parts where forage is temporarily inadequate. Of course, it will be possible for the herder who has extra fodder in one year to sell it to the other herder. Alternatively, it will be possible for the herders to set up an insurance scheme to share the risk of an uncertain environment. However, the setup costs for a new market or a new insurance scheme would be substantial and will not be needed so long as the herders share fodder and risk by jointly sharing a larger grazing area.

It is difficult to know exactly what analysts mean when they refer to the necessity of developing private rights to some common-pool resources (CPRs). It is clear that when they refer to land, they mean to divide the land into separate parcels and assign individual rights to hold, use, and transfer these parcels as individual owners desire (subject to the general regulations of a jurisdiction regarding the use and transfer of land). In regard to nonstationary resources, such as water and fisheries, it is unclear what the establishment of private rights means. As Colin Clark has pointed out, the "'tragedy of the commons' has proved particularly difficult to counteract in the case of marine fishery resources where the establishment of individual property rights is virtually out of the question" (1980, p. 117). In regard to a fugitive resource, a diversity of rights may be established giving individuals rights to use particular types of equipment, to use the resource system at a particular time and place, or to withdraw a particular quantity of resource units (if they can be found). But even when particular rights are unitized, quantified, and salable, the resource *system* is still likely to be owned in common rather than individually.[12] Again, referring to fisheries, Clark has argued that "common ownership is the fundamental fact affecting almost every regime of fishery management" (1980, p. 117).

The "only" way?

Analysts who find an empirical situation with a structure presumed to be a commons dilemma often call for the imposition of a solution by an external actor: The "only way" to solve a commons dilemma is by doing

X. Underlying such a claim is the belief that *X* is necessary and sufficient to solve the commons dilemma. But the content of *X* could hardly be more variable. One set of advocates presumes that a central authority must assume continuing responsibility to make unitary decisions for a particular resource. The other presumes that a central authority should parcel out ownership rights to the resource and then allow individuals to pursue their own self-interests within a set of well-defined property rights. Both centralization advocates and privatization advocates accept as a central tenet that institutional change must come from outside and be imposed on the individuals affected. Despite sharing a faith in the necessity and efficacy of "the state" to change institutions so as to increase efficiency, the institutional changes they recommend could hardly be further apart.

If one recommendation is correct, the other cannot be. Contradictory positions cannot both be right. I do not argue for either of these positions. Rather, I argue that both are too sweeping in their claims. Instead of there being a single solution to a single problem, I argue that many solutions exist to cope with many different problems. Instead of presuming that optimal institutional solutions can be designed easily and imposed at low cost by external authorities, I argue that "getting the institutions right" is a difficult, time-consuming, conflict-invoking process. It is a process that requires reliable information about time and place variables as well as a broad repertoire of culturally acceptable rules. New institutional arrangements do not work in the field as they do in abstract models unless the models are well specified and empirically valid and the participants in a field setting understand how to make the new rules work.

Instead of presuming that the individuals sharing a commons are inevitably caught in a trap from which they cannot escape, I argue that the capacity of individuals to extricate themselves from various types of dilemma situations *varies* from situation to situation. The cases to be discussed in this book illustrate both successful and unsuccessful efforts to escape tragic outcomes. Instead of basing policy on the presumption that the individuals involved are helpless, I wish to learn more from the experience of individuals in field settings. Why have some efforts to solve commons problems failed, while others have succeeded? What can we learn from experience that will help stimulate the development and use of a better theory of collective action – one that will identify the key variables that can enhance or detract from the capabilities of individuals to solve problems?

Institutions are rarely either private or public – "the market" or "the state." Many successful CPR institutions are rich mixtures of "private-like" and "public-like" institutions defying classification in a sterile dichotomy.

Reflections on the commons

By "successful," I mean institutions that enable individuals to achieve productive outcomes in situations where temptations to free-ride and shirk are ever present. A competitive market – the epitome of private institutions – is itself a public good. Once a competitive market is provided, individuals can enter and exit freely whether or not they contribute to the cost of providing and maintaining the market. No market can exist for long without underlying public institutions to support it. In field settings, public and private institutions frequently are intermeshed and depend on one another, rather than existing in isolated worlds.

An alternative solution

To open up the discussion of institutional options for solving commons dilemmas, I want now to present a fifth game in which the herders themselves can make a binding contract to commit themselves to a cooperative strategy that they themselves will work out. To represent this arrangement within a noncooperative framework, additional moves must be overtly included in the game structure. A binding contract is interpreted within noncooperative game theory as one that is unfailingly enforced by an

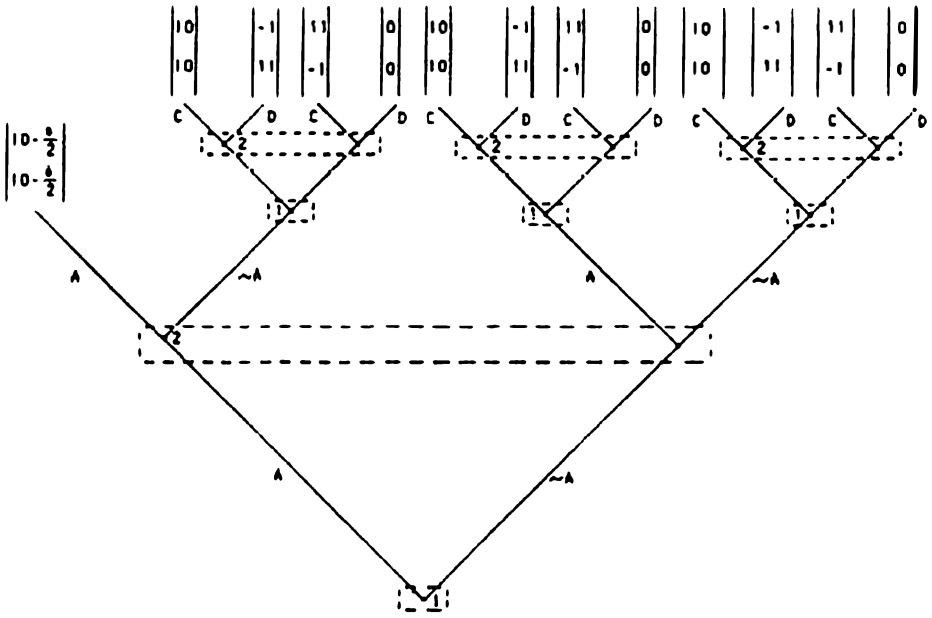


Figure 1.5. Game 5: Self-financed contract-enforcement game.

external actor – just as we interpreted the penalty posited earlier as being unfailingly enforced by the central authority.

A simple way to represent this is to add one parameter to the payoffs and a strategy to both herders' strategy sets.[13] The parameter is the cost of enforcing an agreement and will be denoted by e. The herders in Game 5 must now negotiate prior to placing animals on the meadow. During negotiations, they discuss various strategies for sharing the carrying capacity of the meadow and the costs of enforcing their agreement. Contracts are not enforceable, however, unless agreed to unanimously by the herders. Any proposal made by one herder that did not involve an equal sharing of the carrying capacity and of enforcement costs would be vetoed by the other herder in their negotiations. Consequently, the only feasible agreement – and the equilibrium of the resulting game – is for both herders to share equally the sustainable yield levels of the meadow and the costs of enforcing their agreement so long as each herder's share of the cost of enforcement is less than 10.[14]

Further, in Game 5, players can *always* guarantee that the worst they will do is the (defect, defect) outcome of Game 1. They are not dependent on the accuracy of the information obtained by a distant government official regarding their strategies. If one player suggests a contract based on incomplete or biased information, the other player can indicate an unwillingness to agree. They determine their own contract and ask the enforcer to enforce only that on which they have agreed. If the enforcer should decide to charge too much for its services [any number equal to or greater than $P_i(C, C) - P_i(D, D)$, $i = 1, 2$], neither player would agree to such a contract.

The "solution" of a commons-dilemma game through instrumentalities similar to Game 5 is not presented as the "only way" to solve a commons dilemma. It is merely one way. But this way has been almost totally ignored in both the policy-analysis literature and the formal-theory literature. Contemplating such an option raises numerous questions. First, might it be possible for the herders to hire a private agent to take on the role of enforcer? This is not as farfetched as it might seem at first. Many long-term business exchanges have the structure of a prisoner's dilemma.[15] Businesses are hesitant to accept promises of future performance rather than enforceable contracts, especially when beginning new business relationships. To reduce enforcement costs, however, a frequent practice is to use a private arbitrator rather than a civil court as the mechanism to achieve enforcement.[16] In N-person settings, all professional athletic leagues face problems similar to those illustrated here. During the play of a professional game, the temptation to cheat and break the rules is ever present. Further,

accidents do happen, and rules get broken, even by players who were intending to follow the rules. Athletic leagues typically employ private monitors to enforce their rules.[17]

As soon as we allow the possibility of a private party to take on the role of an external enforcer, the nature of the "solution" offered by Game 5 to the commons dilemma begins to generate a rich set of alternative applications. A self-financed contract-enforcement game allows the participants in the situation to exercise greater control over decisions about who will be allowed to graze and what limits will be placed on the number of animals, as compared with either Game 2 or Game 3. If the parties use a private arbitrator, they do not let the arbitrator impose an agreement on them. The arbitrator simply helps the parties find methods to resolve disputes that arise within the set of working rules to which the parties themselves have agreed. Arbitrators, courts, and other arrangements for enforcement and dispute resolution make it possible for individuals to initiate long-term arrangements that they could not otherwise undertake.[18] Further, as soon as one thinks about a "solution" like Game 5, it is a small step to thinking about the possibility of several arbitrators offering enforcement services at varying charges during the negotiation stage. The payoff-dominant equilibrium is to agree on that arbitrator who will enforce the contract at the lowest e.

The key difference between Game 5 and Games 2 and 3 is that the participants themselves design their own contracts in Game 5 in light of the information they have at hand. The herders, who use the same meadow year after year, have detailed and relatively accurate information about carrying capacity. They observe the behavior of other herders and have an incentive to report contractual infractions. Arbitrators may not need to hire monitors to observe the activities of the contracting parties. The self-interest of those who negotiated the contract will lead them to monitor each other and to report observed infractions so that the contract is enforced. A regulatory agency, on the other hand, always needs to hire its own monitors. The regulatory agency then faces the principal–agent problem of how to ensure that its monitors do their own job.

The proponents of the central-authority "solution" presume that such agencies have accurate information and are able to change incentives to produce something like Game 2. It is difficult for a central authority to have sufficient time-and-place information to estimate accurately both the carrying capacity of a CPR and the appropriate fines to induce cooperative behavior. I believe that situations like that in Game 3, in which incomplete information leads to sanctioning errors, occur more frequently than has been presumed in the policy literature. The need for external monitors and

enforcers is particularly acute when what is being enforced is a decision by an external agent who may impose excess costs on participants.

A further problem for consideration is that games in which enforcers have been arranged for by mutual agreement may be mistaken by analysts and public officials for games in which there have been *no* agreements about how to cooperate and enforce agreements. In other words, some examples of a "Game 5" may be mistaken for a "Game 1."[19] These situations may be construed to be "informal," carrying a presumption that they are not lawful. This goes to fundamental presumptions about the nature of governments as external authorities governing over societies.

As will be seen in the later discussion of empirical cases, users of CPRs have developed a wide diversity in their own agreements, which are enforced by many mechanisms. Some of the enforcement mechanisms are external governmental agencies. Some enforcement mechanisms involve members of the users' community who have been employed as monitors and enforcers. Some enforcement mechanisms involve the users themselves as their own monitors. When the enforcement mechanism is not an external governmental agency, some analysts presume that there is no enforcement. That is why Game 5 is mistaken for Game 1.

A self-financed contract-enforcement game is no panacea. Such institutional arrangements have many weaknesses in many settings. The herders can overestimate or underestimate the carrying capacity of the meadow. Their own monitoring system may break down. The external enforcer may not be able to enforce ex post, after promising to do so ex ante. A myriad of problems can occur in natural settings, as is also the case with the idealized central-regulation or private-property institutions.

The structure of the institutional arrangements that one finds in natural settings is, of course, far more complicated than the structure of any of the extremely simple games presented here for discussion. What I attempt to do with these simple games is to generate different ways of thinking about the mechanisms that individuals may use to extricate themselves from commons dilemmas – ways different from what one finds in much of the policy literature. To challenge this mind-set, one needs only simple mechanisms that illustrate alternatives to those that normally are presented as the dominant solutions.

An empirical alternative

Game 5 illustrated a theoretical alternative to centralization or privatization as ways to solve CPR problems. Let us now briefly consider a solution devised by participants in a field setting – Alanya, Turkey – that cannot be

characterized as either central regulation or privatization. The inshore fishery at Alanya, as described by Fikret Berkes (1986b), is a relatively small operation. Many of the approximately 100 local fishers operate in two- or three-person boats using various types of nets. Half of the fishers belong to a local producers' cooperative. According to Berkes, the early 1970s were the "dark ages" for Alanya. The economic viability of the fishery was threatened by two factors: First, unrestrained use of the fishery had led to hostility and, at times, violent conflict among the users. Second, competition among fishers for the better fishing spots had increased production costs, as well as the level of uncertainty regarding the harvest potential of any particular boat.

Early in the 1970s, members of the local cooperative began experimenting with an ingenious system for allotting fishing sites to local fishers. After more than a decade of trial-and-error efforts, the rules used by the Alanya inshore fishers are as follows:

- Each September, a list of eligible fishers is prepared, consisting of all licensed fishers in Alanya, regardless of co-op membership.
- Within the area normally used by Alanya fishers, all usable fishing locations are named and listed. These sites are spaced so that the nets set in one site will not block the fish that should be available at the adjacent sites.
- These named fishing locations and their assignments are in effect from September to May.
- In September, the eligible fishers draw lots and are assigned to the named fishing locations.
- From September to January, each day each fisher moves east to the next location. After January, the fishers move west. This gives the fishers equal opportunities at the stocks that migrate from east to west between September and January and reverse their migration through the area from January to May (Berkes 1986b, pp. 73–4).

The system has the effect of spacing the fishers far enough apart on the fishing grounds that the production capabilities at each site are optimized. All fishing boats also have equal chances to fish at the best spots. Resources are not wasted searching for or fighting over a site.[20] No signs of overcapitalization are apparent.

The list of fishing locations is endorsed by each fisher and deposited with the mayor and local gendarme once a year at the time of the lottery. The process of monitoring and enforcing the system is, however, accomplished by the fishers themselves as a by-product of the incentive created by the

rotation system. On a day when a given fisher is assigned one of the more productive spots, that fisher will exercise that option with certainty (leaving aside last-minute breakdowns in equipment). All other fishers can expect that the assigned fisher will be at the spot bright and early. Consequently, an effort to cheat on the system by traveling to a good spot on a day when one is assigned to a poor spot has little chance of remaining undetected. Cheating on the system will be observed by the very fishers who have rights to be in the best spots and will be willing to defend their rights using physical means if necessary. Their rights will be supported by everyone else in the system. The others will want to ensure that their own rights will not be usurped on the days when they are assigned good sites. The few infractions that have occurred have been handled easily by the fishers at the local coffeehouse (Berkes 1986b, p. 74).

Although this is not a private-property system, rights to use fishing sites and duties to respect these rights are well defined. And though it is not a centralized system, national legislation that has given such cooperatives jurisdiction over "local arrangements" has been used by cooperative officials to legitimize their role in helping to devise a workable set of rules. That local officials accept the signed agreement each year also enhances legitimacy. The actual monitoring and enforcing of the rules, however, are left to the fishers.

Central-government officials could not have crafted such a set of rules without assigning a full-time staff to work (actually fish) in the area for an extended period. Fishing sites of varying economic value are commonly associated with inshore fisheries (Christy 1982; Forman 1967), but they are almost impossible to map without extensive on-site experience. Mapping this set of fishing sites, such that one boat's fishing activities would not reduce the migration of fish to other locations, would have been a daunting challenge had it not been for the extensive time-and-place information provided by the fishers and their willingness to experiment for a decade with various maps and systems. Alanya provides an example of a self-governed common-property arrangement in which the rules have been devised and modified by the participants themselves and also are monitored and enforced by them.

The case of the Alanya inshore fishery is only one empirical example of the many institutional arrangements that have been devised, modified, monitored, and sustained by the users of renewable CPRs to constrain individual behavior that would, if unconstrained, reduce joint returns to the community of users. In addition to the case studies discussed in Chapters 3, 4, and 5, productive CPR institutional arrangements have been

well documented for many farmer-managed irrigation systems, communal forests, inshore fisheries, and grazing and hunting territories.[21]

Game 5 and empirical cases of successfully governed CPRs provide theoretical and empirical alternatives to the assertion that those involved cannot extricate themselves from the problems faced when multiple individuals use a given resource. The key to my argument is that some individuals have broken out of the trap inherent in the commons dilemma, whereas others continue remorsefully trapped into destroying their own resources.[22] This leads me to ask what differences exist between those who have broken the shackles of a commons dilemma and those who have not. The differences may have to do with factors *internal* to a given group. The participants may simply have no capacity to communicate with one another, no way to develop trust, and no sense that they must share a common future. Alternatively, powerful individuals who stand to gain from the current situation, while others lose, may block efforts by the less powerful to change the rules of the game. Such groups may need some form of external assistance to break out of the perverse logic of their situation.

The differences between those who have and those who have not extricated themselves from commons dilemmas may also have to do with factors *outside* the domain of those affected. Some participants do not have the autonomy to change their own institutional structures and are prevented from making constructive changes by external authorities who are indifferent to the perversities of the commons dilemma, or may even stand to gain from it. Also, there is the possibility that external changes may sweep rapidly over a group, giving them insufficient time to adjust their internal structures to avoid the suboptimal outcomes. Some groups suffer from perverse incentive systems that are themselves the results of policies pursued by central authorities. Many potential answers spring to mind regarding the question why some individuals do not achieve collective benefits for themselves, whereas others do. However, as long as analysts presume that individuals cannot change such situations themselves, they do not ask what internal or external variables can enhance or impede the efforts of communities of individuals to deal creatively and constructively with perverse problems such as the tragedy of the commons.

Policy prescriptions as metaphors

Policy analysts who would recommend a single prescription for commons problems have paid little attention to how diverse institutional arrange-

ments operate in practice. The centrists presume that unified authorities will operate in the field as they have been designed to do in the textbooks – determining the best policies to be adopted for a resource based on valid scientific theories and adequate information. Implementation of these policies without error is assumed. Monitoring and sanctioning activities are viewed as routine and nonproblematic.

Those advocating the private-property approach presume that the most efficient use patterns for CPRs will actually result from dividing the rights to access and control such resources. Systematic empirical studies have shown that private organization of firms dealing in goods such as electricity, transport, and medical services tends to be more efficient than governmental organization of such firms; for a review of this literature, see De Alessi (1980). Whether private or public forms are more efficient in industries in which certain potential beneficiaries cannot be excluded is, however, a different question. We are concerned with the types of institutions that will be most efficient for governing and managing diverse CPRs for which at least some potential beneficiaries cannot be excluded. Privatizing the ownership of CPRs need not have the same positive results as privatizing the ownership of an airline. Further, privatizing may not mean "dividing up" at all. Privatization can also mean assigning the exclusive right to harvest from a resource system to a single individual or firm.

Many policy prescriptions are themselves no more than metaphors. Both the centralizers and the privatizers frequently advocate oversimplified, idealized institutions – paradoxically, almost "institution-free" institutions. An assertion that central regulation is necessary tells us nothing about the way a central agency should be constituted, what authority it should have, how the limits on its authority should be maintained, how it will obtain information, or how its agents should be selected, motivated to do their work, and have their performances monitored and rewarded or sanctioned. An assertion that the imposition of private property rights is necessary tells us nothing about how that bundle of rights is to be defined, how the various attributes of the goods involved will be measured, who will pay for the costs of excluding nonowners from access, how conflicts over rights will be adjudicated, or how the residual interests of the right-holders in the resource system itself will be organized.

An important lesson that one learns by carefully studying the growing number of systematic studies by scholars associated with "the new institutionalism" is that these "institutional details" are important.[23] Whether or not any equilibria are possible and whether or not an equilibrium would be an improvement for the individuals involved (or for others who are in turn affected by these individuals) will depend on the particular structures

of the institutions. In the most general sense, all institutional arrangements can be thought of as games in extensive form. As such, the particular options available, the sequencing of those options, the information provided, and the relative rewards and punishments assigned to different sequences of moves can all change the pattern of outcomes achieved. Further, the particular structure of the physical environment involved also will have a major impact on the structure of the game and its results. Thus, a set of rules used in one physical environment may have vastly different consequences if used in a different physical environment.

Policies based on metaphors can be harmful

Relying on metaphors as the foundation for policy advice can lead to results substantially different from those presumed to be likely. Nationalizing the ownership of forests in Third World countries, for example, has been advocated on the grounds that local villagers cannot manage forests so as to sustain their productivity and their value in reducing soil erosion. In countries where small villages had owned and regulated their local communal forests for generations, nationalization meant expropriation. In such localities, villagers had earlier exercised considerable restraint over the rate and manner of harvesting forest products. In some of these countries, national agencies issued elaborate regulations concerning the use of forests, but were unable to employ sufficient numbers of foresters to enforce those regulations. The foresters who were employed were paid such low salaries that accepting bribes became a common means of supplementing their income. The consequence was that nationalization created *open-access resources* where limited-access *common-property resources* had previously existed. The disastrous effects of nationalizing formerly communal forests have been well documented for Thailand (Feeny 1988a), Niger (Thomson 1977; Thomson, Feeny, and Oakerson 1986), Nepal (Arnold and Campbell 1986; Messerschmidt 1986), and India (Gadgil and Iyer 1989). Similar problems occurred in regard to inshore fisheries when national agencies presumed that they had exclusive jurisdiction over all coastal waters (Cordell and McKean 1986; W. Cruz 1986; Dasgupta 1982; Panayoutou 1982; Pinkerton 1989a).

A CHALLENGE

An important challenge facing policy scientists is to develop theories of human organization based on realistic assessment of human capabilities and limitations in dealing with a variety of situations that initially share

some or all aspects of a tragedy of the commons. Empirically validated theories of human organization will be essential ingredients of a policy science that can inform decisions about the likely consequences of a multitude of ways of organizing human activities. Theoretical inquiry involves a search for regularities. It involves abstraction from the complexity of a field setting, followed by the positing of theoretical variables that underlie observed complexities. Specific models of a theory involve further abstraction and simplification for the purpose of still finer analysis of the logical relationships among variables in a closed system. As a theorist, and at times a modeler, I see these efforts at the core of a policy science.

One can, however, get trapped in one's own intellectual web. When years have been spent in the development of a theory with considerable power and elegance, analysts obviously will want to apply this tool to as many situations as possible. The power of a theory is exactly proportional to the diversity of situations it can explain. All theories, however, have limits. Models of a theory are limited still further because many parameters must be fixed in a model, rather than allowed to vary. Confusing a model – such as that of a perfectly competitive market – with the theory of which it is one representation can limit applicability still further.

Scientific knowledge is as much an understanding of the diversity of situations for which a theory or its models are relevant as an understanding of its limits. The conviction that all physical structures could be described in terms of a set of perfect forms – circles, squares, and triangles – limited the development of astronomy until Johannes Kepler broke the bonds of classical thought and discovered that the orbit of Mars was elliptical – a finding that Kepler himself initially considered to be no more than a pile of dung (Koestler 1959). Godwin and Shepard (1979) pointed out a decade ago that policy scientists were doing the equivalent of "Forcing Squares, Triangles and Ellipses into a Circular Paradigm" by using the commons-dilemma model without serious attention to whether or not the variables in the empirical world conformed to the theoretical model. Many theoretical and empirical findings have been reported since Godwin and Shepard's article that should have made policy scientists even more skeptical about relying on a limited set of models to analyze the diversity of situations broadly referred to as CPR problems. Unfortunately, many analysts – in academia, special-interest groups, governments, and the press – still presume that common-pool problems are all dilemmas in which the participants themselves cannot avoid producing suboptimal results, and in some cases disastrous results.

What is missing from the policy analyst's tool kit – and from the set of accepted, well-developed theories of human organization – is an ade-

quately specified theory of collective action whereby a group of principals can organize themselves voluntarily to retain the residuals of their own efforts. Examples of self-organized enterprises abound. Most law firms are obvious examples: A group of lawyers will pool their assets to purchase a library and pay for joint secretarial and research assistance. They will develop their own internal governance mechanisms and formulas for allocating costs and benefits to the partners. Most cooperatives are also examples. The cases of self-organized and self-governed CPRs that we consider in Chapter 3 are also examples. But until a theoretical explanation – based on human choice – for self-organized and self-governed enterprises is fully developed and accepted, major policy decisions will continue to be undertaken with a presumption that individuals cannot organize themselves and always need to be organized by external authorities.

Further, all organizational arrangements are subject to stress, weakness, and failure. Without an adequate theory of self-organized collective action, one cannot predict or explain when individuals will be unable to solve a common problem through self-organization alone, nor can one begin to ascertain which of many intervention strategies might be effective in helping to solve particular problems. As discussed earlier, there is a considerable difference between the presumption that a regulatory agency should be established and the presumption that a reliable court system is needed to monitor and enforce self-negotiated contracts. If the theories being used in a policy science do not include the possibility of self-organized collective action, then the importance of a court system that can be used by self-organizing groups to monitor and enforce contracts will not be recognized.[14]

I hope this inquiry will contribute to the development of an empirically supported theory of self-organizing and self-governing forms of collective action. What I attempt to do in this volume is to combine the strategy used by many scholars associated with the "new institutionalism" with the strategy used by biologists for conducting empirical work related to the development of a better theoretical understanding of the biological world.

As an institutionalist studying empirical phenomena, I presume that individuals try to solve problems as effectively as they can. That assumption imposes a discipline on me. Instead of presuming that some individuals are incompetent, evil, or irrational, and others are omniscient, I presume that individuals have very similar limited capabilities to reason and figure out the structure of complex environments. It is my responsibility as a scientist to ascertain what problems individuals are trying to solve and what factors help or hinder them in these efforts. When the problems that I observe involve lack of predictability, information, and trust, as well as high levels

of complexity and transactional difficulties, then my efforts to explain must take these problems overtly into account rather than assuming them away. In developing an explanation for observed behavior, I draw on a rich literature written by other scholars interested in institutions and their effects on individual incentives and behaviors in field settings.

Biologists also face the problem of studying complex processes that are poorly understood. Their scientific strategy frequently has involved identifying for empirical observation the simplest possible organism in which a process occurs in a clarified, or even exaggerated, form. The organism is not chosen because it is representative of all organisms. Rather, the organism is chosen because particular processes can be studied more effectively using this organism than using another.

My "organism" is a type of human situation. I call this situation a CPR situation and define exactly what I mean by this and other key terms in Chapter 2. In this volume, I do not include all potential CPR situations within the frame of reference. I focus entirely on small-scale CPRs, where the CPR is itself located within one country and the number of individuals affected varies from 50 to 15,000 persons who are heavily dependent on the CPR for economic returns. These CPRs are primarily inshore fisheries, smaller grazing areas, groundwater basins, irrigation systems, and communal forests. Because these are relatively small-scale situations, serious study is more likely to penetrate the surface complexity to identify underlying similarities and processes. Because the individuals involved gain a major part of their economic return from the CPRs, they are strongly motivated to try to solve common problems to enhance their own productivity over time. The effort to self-organize in these situations may be somewhat exaggerated, but that is exactly why I want to study this process in these settings. Further, when self-organization fails, I know that it is not because the collective benefits that could have been obtained were unimportant to the participants.

There are limits on the types of CPRs studied here: (1) renewable rather than nonrenewable resources, (2) situations where substantial scarcity exists, rather than abundance, and (3) situations in which the users can substantially harm one another, but not situations in which participants can produce major external harm for others. Thus, all asymmetrical pollution problems are excluded, as is any situation in which a group can form a cartel and control a sufficient part of the market to affect market price.

In the empirical studies, I present a synopsis of important CPR cases that have aided my understanding of the processes of self-organization and self-governance. These cases are in no sense a "random" sample of cases. Rather, these are cases that provide clear information about the processes

involved in (1) governing long-enduring CPRs, (2) transforming existing institutional arrangements, and (3) failing to overcome continued CPR problems. These cases can thus be viewed as a collection of the most salient raw materials with which I have worked in my effort to understand how individuals organize and govern themselves to obtain collective benefits in situations where the temptations to free-ride and to break commitments are substantial.

From an examination and analysis of these cases, I attempt to develop a series of reasoned conjectures about how it is possible that some individuals organize themselves to govern and manage CPRs and others do not. I try to identify the underlying design principles of the institutions used by those who have successfully managed their own CPRs over extended periods of time and why these may affect the incentives for participants to continue investing time and effort in the governance and management of their own CPRs. I compare the institutions used in successful and unsuccessful cases, and I try to identify the internal and external factors that can impede or enhance the capabilities of individuals to use and govern CPRs.

I hope these conjectures contribute to the development of an empirically valid theory of self-organization and self-governance for at least one well-defined universe of problematical situations. That universe contains a substantial proportion of renewable resources heavily utilized by human beings in different parts of the world. It is estimated, for example, that 90% of the world's fishermen and over half of the fish consumed each year are captured in the small-scale, inshore fisheries included within the frame of this study (Panayoutou 1982, p. 49). Further, my choice of the CPR environment for intensive study was based on a presumption that I could learn about the processes of self-organization and self-governance of relevance to a somewhat broader set of environments.

Given the similarity between many CPR problems and the problems of providing small-scale collective goods, the findings from this volume should contribute to an understanding of the factors that can enhance or detract from the capabilities of individuals to organize collective action related to providing local public goods. All efforts to organize collective action, whether by an external ruler, an entrepreneur, or a set of principals who wish to gain collective benefits, must address a common set of problems. These have to do with coping with free-riding, solving commitment problems, arranging for the supply of new institutions, and monitoring individual compliance with sets of rules. A study that focuses on how individuals avoid free-riding, achieve high levels of commitment, arrange for new institutions, and monitor conformity to a set of rules in CPR environments should contribute to an understanding of how in-

dividuals address these crucial problems in some other settings as well.

Let me now give a brief sketch of how this book is organized. In Chapter 2, I define what I mean by a CPR situation and individual choice in a CPR situation. Then I examine a series of crucial questions that any theory of collective action must answer. To conclude the chapter, I examine two assumptions that have framed prior work and discuss the alternatives that frame my analysis. The empirical part of this volume is contained in Chapters 3, 4, and 5, where I examine specific cases of long-enduring CPR institutions and resources, the origin and development of CPR institutions, and CPR failures and fragilities. At the end of each empirical chapter, I consider what can be learned from the cases in that chapter that will contribute toward the development of a better theory of self-organization related to CPR environments. In Chapter 6, I pull together the theoretical reflections contained at the ends of Chapters 3, 4, and 5 and address the implications of these conjectures for the design of self-organizing and self-governing institutions.

2

An institutional approach to the study of self-organization and self-governance in CPR situations

In Chapter 1, I described my strategy as that of a "new institutionalist" who has picked small-scale CPR situations to study because the processes of self-organization and self-governance are easier to observe in this type of situation than in many others. The central question in this study is how a group of principals who are in an interdependent situation can organize and govern themselves to obtain continuing joint benefits when all face temptations to free-ride, shirk, or otherwise act opportunistically. Parallel questions have to do with the combinations of variables that will (1) increase the initial likelihood of self-organization, (2) enhance the capabilities of individuals to continue self-organized efforts over time, or (3) exceed the capacity of self-organization to solve CPR problems without external assistance of some form.

This chapter has several objectives. First, I define what I mean by CPRs and how I view individual behaviors in complex and uncertain CPR situations. Then I examine the general problem facing individuals in CPR situations: how to organize to avoid the adverse outcomes of independent action. This general problem is solved by external agents in two well-accepted theories: the theory of the firm and the theory of the state. These explain how new institutions are supplied, how commitments are obtained, and how the actions of agents and subjects are monitored effectively, using in one case the firm, and in the other state, as an organizational device. How a group of principals – a community of citizens – can organize themselves to solve the problems of institutional supply, commitment, and monitoring is still a theoretical puzzle. Given that some individuals solve this puzzle, whereas others do not, a study of successful and unsuccessful efforts to solve CPR problems should address important issues related to the theory of collective action and the development of better policies related to CPRs. Many efforts to analyze collective-action problems have

framed the analysis by presuming that all such problems can be represented as prisoner's dilemma (PD) games, that a single level of analysis is sufficient, and that transactions costs are insignificant and can be ignored. In the last section of this chapter, I propose assumptions that are alternatives to those that normally frame the analysis of collective action.

THE CPR SITUATION

CPRs and resource units

The term "common-pool resource" refers to a natural or man-made resource system that is sufficiently large as to make it costly (but not impossible) to exclude potential beneficiaries from obtaining benefits from its use. To understand the processes of organizing and governing CPRs, it is essential to distinguish between the *resource system* and the flow of *resource units* produced by the system, while still recognizing the dependance of the one on the other.

Resource systems are best thought of as stock variables that are capable, under favorable conditions, of producing a maximum quantity of a flow variable without harming the stock or the resource system itself. Examples of resource systems include fishing grounds, groundwater basins, grazing areas, irrigation canals, bridges, parking garages, mainframe computers, and streams, lakes, oceans, and other bodies of water. Resource units are what individuals appropriate or use from resource systems. Resource units are typified by the tons of fish harvested from a fishing ground, the acre-feet or cubic meters of water withdrawn from a groundwater basin or an irrigation canal, the tons of fodder consumed by animals from a grazing area, the number of bridge crossings used per year by a bridge, the parking spaces filled, the central processing units consumed by those sharing a computer system, and the quantity of biological waste absorbed per year by a stream or other waterway. The distinction between the resource as a *stock* and the harvest of use units as a *flow* is especially useful in connection with *renewable* resources, where it is possible to define a replenishment rate. As long as the average rate of withdrawal does not exceed the average rate of replenishment, a renewable resource is sustained over time.[1]

Access to a CPR can be limited to a single individual or firm or to multiple individuals or teams of individuals who use the resource system at the same time. The CPRs studied in this volume are used by multiple individuals or firms. Following Plott and Meyer (1975), I call the process of withdrawing resource units from a resource system "appropriation." Those who withdraw such units are called "appropriators."[2] One term –

"appropriator" – can thus be used to refer to herders, fishers, irrigators, commuters, and anyone else who appropriates resource units from some type of resource system. In many instances appropriators use or consume the resource units they withdraw (e.g., where fishers harvest primarily for consumption). Appropriators also use resource units as inputs into production processes (e.g., irrigators apply water to their fields to produce rice). In other instances, the appropriators immediately transfer ownership of resource units to others, who are then the users of the resource units (e.g., fishers who sell their catch as soon as possible after arrival at a port).

The analysis of scarce, renewable resources is made from the perspective of the appropriators. This is not the only perspective that can be used in an analysis of complex CPR problems. If the appropriators of a resource unit gain considerable market power, such as by creating a cartel to influence price, their strategies affect themselves as well as others. This analysis relates to situations in which CPR appropriators have no power in a final-goods market, nor do their actions have significant impact on the environment of others living outside the range of their CPR.

The term I use to refer to those who arrange for the provision of a CPR is "providers." I use the term "producer" to refer to anyone who actually constructs, repairs, or takes actions that ensure the long-term sustenance of the resource system itself. Frequently, providers and producers are the same individuals, but they do not have to be (V. Ostrom, Tiebout, and Warren 1961). A national government may provide an irrigation system in the sense of arranging for its financing and design. It may then arrange with local farmers to produce and maintain it. If local farmers are given the authority to arrange for maintenance, then they become both the providers and the producers of maintenance activities related to a CPR.

A resource system can be jointly provided and/or produced by more than one person or firm. The actual process of appropriating resource units from the CPR can be undertaken by multiple appropriators simultaneously or sequentially. The resource units, however, *are not subject to joint use or appropriation.* The fish harvested by one boat are not there for someone else. The water spread on one farmer's fields cannot be spread onto someone else's fields. Thus, the resource units are not jointly used, but the resource system is subject to joint use. Once multiple appropriators rely on a given resource system, improvements to the system are simultaneously available to all appropriators. It is costly (and in some cases infeasible) to exclude one appropriator of a resource system from improvements made to the resource system itself. All appropriators benefit from maintenance performed on an irrigation canal, a bridge, or a computer system whether they contribute or not.

Failure to distinguish between the subtractability of the resource units and the jointness of the resource system has in the past contributed to confusion about the relationship of CPRs to public or collective goods.[3] Michael Taylor recognized the difference between CPRs and collective goods when he wrote the following:

> There is, in particular, a very important class of collective action problems which arise in connection with the use of resources to which there is open access – resources, that is, which nobody is prevented from using. These resources need not be public goods. (M. Taylor 1987, p. 3)

The relatively high costs of physically excluding joint appropriators from the resource or from improvements made to the resource system are similar to the high costs of excluding potential beneficiaries from public goods. This shared attribute is responsible for the ever present temptation to free-ride that exists in regard to both CPRs and public goods. There is as much temptation to avoid contributing to the provision of a resource system as there is to avoid contributing to the provision of public security or weather forecasts. Theoretical propositions that are derived solely from the difficulty of exclusion are applicable to the *provision* of both CPRs and collective goods.

But one's use of a weather forecast does not subtract from the availability of that forecast to others, just as one's consumption of public security does not reduce the general level of security available in a community.[4] "Crowding effects" and "overuse" problems are chronic in CPR situations, but absent in regard to pure public goods. The subtractability of the resource unit leads to the possibility of approaching the limit of the number of resource units produced by a CPR. When the CPR is a man-made structure, such as a bridge, approaching the limit of crossing units will lead to congestion. When the CPR is a biological resource, such as a fishery or a forest, approaching the limit of resource units not only may produce short-run crowding effects but also may destroy the capability of the resource itself to continue producing resource units. Even a physical resource, such as a bridge, can be destroyed by heavier use than was allowed for in its engineering specifications.

Thus, propositions derived from a theory of public goods that are based on the nonsubtractive attributes of those goods are *not* applicable to an analysis of *appropriation* and *use* of subtractable resource units. Appropriation and use of the resource units are more closely related to the theory of private goods than to the theory of public goods. On the other hand, the process of designing, implementing, and enforcing a set of rules to coordinate provision activities is equivalent to the provision of a local collec-

tive good. CPR appropriators who organize themselves to govern and manage a CPR are faced with some problems that are similar to those of appropriating private goods and other problems that are similar to those of providing public goods. Both aspects are intimately bound together physically and analytically. In a particular CPR, if problems associated with the appropriation of subtractable resource units become severe, local appropriators may refuse to undertake provision activities.[5] No appropriation of resource units can occur without a resource system. Without a fair, orderly, and efficient method of allocating resource units, local appropriators have little motivation to contribute to the continued provision of the resource system.

Rational appropriators in complex and uncertain situations

The decisions and actions of CPR appropriators to appropriate from and provide a CPR are those of broadly rational individuals who find themselves in complex and uncertain situations. An individual's choice of behavior in any particular situation will depend on how the individual learns about, views, and weighs the benefits and costs of actions and their perceived linkage to outcomes that also involve a mixture of benefits and costs.[6]

Organizing appropriators for collective action regarding a CPR is usually an uncertain and complex undertaking. Uncertainty has many external sources: the quantity and timing of rainfall, the temperature and amount of sunlight, the presence or absence of disease-bearing vectors, and the market prices of various inputs or final products. Other sources of uncertainty are internal to the CPR and the appropriators using the CPR. A major source of uncertainty is lack of knowledge. The exact structure of the resource system itself – its boundary and internal characteristics – must be established. Ascertaining the structure of the resource system may come about as a by-product of extended use and careful observation, as in the case of appropriating from a fishing ground or grazing range. Moreover, this folk knowledge must be preserved and passed along from one generation to the next. For a groundwater basin, on the other hand, the discovery of the internal structure may require a major investment in research by geologists and engineers.

How appropriators' actions affect the resource system, the yield of resource units, and each other's outcomes must also be ascertained.[7] It is not immediately apparent, for example, how one irrigator's forbearance in taking water from a canal will affect the yield obtained by that farmer or by other farmers. In some cases, a farmer located near the head of a system

may be able to curtail his water use substantially without a major impact on his own yield, while substantially enhancing the yields of downstream farmers. In other cases, the excess water taken by the farmer located near the headworks may subsequently also flow to farmers located lower in the system. Restraint by the farmer located higher in the system may not increase total yield. Uncertainties stemming from lack of knowledge may be reduced over time as a result of skillful pooling and blending of scientific knowledge and local time-and-place knowledge. Uncertainty reduction is costly and never fully accomplished. The uncertainty stemming from strategic behavior by the appropriators remains even after one acquires considerable knowledge about the resource system itself.

Given these levels of uncertainty about the basic structure of the problems appropriators face, the only reasonable assumption to make about the discovery and calculation processes employed is that appropriators engage in a considerable amount of trial-and-error learning. Many actions are selected without full knowledge of their consequences. Some dams wash out after the first heavy rains. Some rules cannot be enforced because no one is able to monitor conformance to them. By definition, trial-and-error methods involve error, perhaps even disasters. Over time, appropriators gain a more accurate understanding of the physical world and what to expect from the behavior of others.

Appropriators in many settings are strongly motivated to find better solutions to their problems if they can. The economic livelihood of the appropriators depends on their ingenuity in solving individual and joint problems. How complete and accurate the information local appropriators obtain about their situation will vary from one situation to another, depending on the number of appropriators involved, the complexity of the situation, and the stability of factors affecting individual behaviors and resource-system responses. The symmetry of information available to appropriators will also vary from situation to situation, depending on how expensive it is to acquire information and the rules used for disseminating information to appropriators.

Collective-action problems related to the provision of CPRs and appropriation from CPRs extend over time. Individuals attribute less value to benefits that they expect to receive in the distant future, and more value to those expected in the immediate future. In other words, individuals discount future benefits – how severely depends on several factors. Time horizons are affected by whether or not individuals expect that they or their children will be present to reap these benefits, as well as by opportunities they may have for more rapid returns in other settings. The dis-

count rates applied to future yields derived from a particular CPR may differ substantially across various types of appropriators. In a fishery, for example, the discount rates of local fishers who live in nearby villages will differ from the discount rates of those who operate the larger trawlers, who may fish anywhere along a coastline. The time horizons of the local fishers, in relation to the yield of the inshore fishery, extend far into the future. They hope that their children and their children's children can make a living in the same location. More mobile fishers, on the other hand, can go on to other fishing grounds when local fish are no longer available.

Discount rates are affected by the levels of physical and economic security faced by appropriators. Appropriators who are uncertain whether or not there will be sufficient food to survive the year will discount future returns heavily when traded off against increasing the probability of survival during the current year. Similarly, if a CPR can be destroyed by the actions of others, no matter what local appropriators do, even those who have constrained their harvesting from a CPR for many years will begin to heavily discount future returns, as contrasted with present returns.[8] Discount rates are also affected by the general norms shared by the individuals living in a particular society, or even a local community, regarding the relative importance of the future as compared with the present.

Discount rates are not the only aspects of human choice that are affected by shared norms of behavior. Although I stress the importance that the expected consequences will have on one's decisions, individuals vary in regard to the importance they place on acting in ways that they and others view as right and proper. Norms of behavior reflect valuations that individuals place on actions or strategies in and of themselves, not as they are connected to immediate consequences.[9] When an individual has strongly internalized a norm related to keeping promises, for example, the individual suffers shame and guilt when a personal promise is broken. If the norm is shared with others, the individual is also subject to considerable social censure for taking an action considered to be wrong by others.

Norms of behavior therefore affect the way alternatives are perceived and weighed. For many routine decisions, actions that are considered wrong among a set of individuals interacting together over time will not even be included in the set of strategies contemplated by the individual. If the individual's attention is drawn to the possibility of taking such an action by the availability of a very large payoff for doing so, the action may be included in the set of alternatives to be considered, but with a high cost attached. Actions that are strongly proscribed among a set of individuals will occur less frequently (even though they promise to yield high net

payoffs to individuals) than will those same actions in a community that does not censure such actions.

The most important impact that the type and extent of shared norms will have on the strategies available to individuals has to do with the level of opportunistic behavior that appropriators can expect from other appropriators. Opportunism is defined as "self-interest with guile" (Williamson 1975). In a setting in which few individuals share norms about the impropriety of breaking promises, refusing to do one's share, shirking, or taking other opportunistic actions, each appropriator must expect all other appropriators to act opportunistically whenever they have the chance. In such a setting it is difficult to develop stable, long-term commitments. Expensive monitoring and sanctioning mechanisms may be needed. Some long-term arrangements that once were productive are no longer feasible, given their costs of enforcement. In a setting in which there are strong norms against opportunistic behavior, each appropriator will be less wary about the dangers of opportunism.

In every group there will be individuals who will ignore norms and act opportunistically when given a chance. There are also situations in which the potential benefits will be so high that even strongly committed individuals will break norms. Consequently, the adoption of norms of behavior will not reduce opportunistic behavior to zero. Opportunistic behavior is a possibility that must be dealt with by all appropriators trying to solve CPR problems.

In some settings, however, rampant opportunistic behavior severely limits what can be done jointly without major investments in monitoring and sanctioning arrangements. Substantial benefits have to be obtained to make costly monitoring and sanctioning activities worthwhile. In other settings, long-term joint commitments can be undertaken with only a modest investment in monitoring and sanctioning arrangements. Shared norms that reduce the cost of monitoring and sanctioning activities can be viewed as social capital to be utilized in solving CPR problems.

Because CPR settings extend over time, and individuals adopt internal norms, it is possible for individuals to utilize contingent strategies, not simply independent strategies, in relating to one another. By "contingent strategies" I mean a whole class of planned actions that are contingent on conditions in the world. The contingent strategy that has been the object of the most scholarly attention is tit for tat in a two-person game in which an individual adopts a cooperative action in the first round and then mimics the action of the opponent in future rounds (Axelrod 1981, 1984). There are many other contingent strategies that can be adopted; they vary

in terms of the level of initial cooperation extended and the actions of others required for switching behavioral patterns. That individuals utilize contingent strategies in many complex and uncertain field settings is an important foundation for later analysis.

Thus, I use a very broad conception of rational action, rather than a narrowly defined conception. The internal world of individual choice that I use is illustrated in Figure 2.1. Four internal variables – expected benefits, expected costs, internal norms, and discount rates – affect an individual's choice of strategies. Individuals selecting strategies jointly produce outcomes in an external world that impinge on future expectations concerning the benefits and costs of actions. What types of internal norms an individual possesses are affected by the shared norms held by others in regard to particular types of situations. Similarly, internal discount rates are affected by the range of opportunities that an individual has outside any particular situation.

This general model of individual choice is thus open to many particular specifications. The particular assumptions made about the completeness, shape, and differentiability of preference functions depend on the situation of relevance for a particular model in this theory. In simple, highly constrained situations where individuals have interacted for long periods of time, assumptions about convex, twice-differentiable preference functions

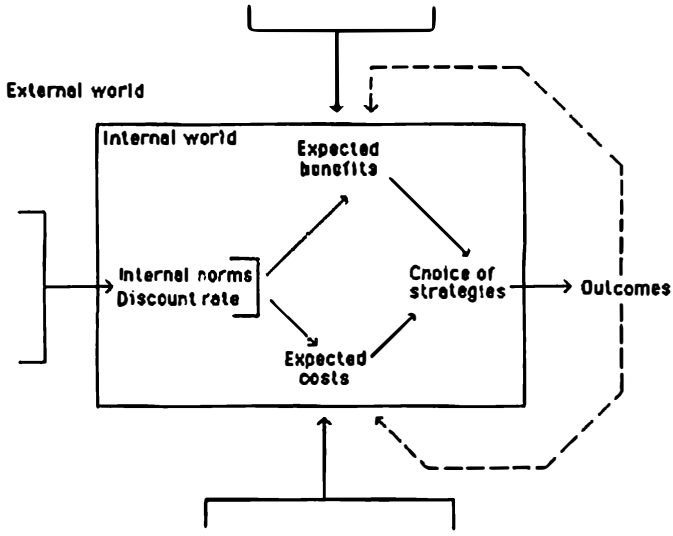


Figure 2.1. The internal world of individual choice.

may be appropriate. In complex situations involving unstructured problems, assuming complete preference functions of any shape is not meaningful. The most one can say is that the individuals in such situations are engaged in a trial-and-error effort to learn more about the results of their actions so that they can evaluate benefits and costs more effectively over time.

This general conception is one way of fulfilling Popper's advice to make the rationality principle "an almost empty principle" (Popper 1967). It places the primary weight of theoretical analysis on specifying rigorously and fully the models of the situations in which individuals find themselves. It accepts Popper's methodological advice to emphasize the way we describe the situations in which individuals find themselves so that we can use observable variables to reject our theories, rather than internal, in-the-mind, subjective variables, which are far more difficult to measure.

Thus, most of the analysis contained in this volume examines the combinations of situational variables that are most likely to affect individuals' choices of strategies and how those situational variables occur.

INTERDEPENDENCE, INDEPENDENT ACTION, AND COLLECTIVE ACTION

When multiple appropriators are dependent on a given CPR as a source of economic activity, they are jointly affected by almost everything they do. Each individual must take into account the choices of others when assessing personal choices. If one fisher occupies a good fishing site, a second fisher arriving at the same location must invest more resources to travel to another site, or else fight for the first site. If one irrigator allocates time and materials to repairing a broken control gate in an irrigation canal, all other irrigators using that canal are affected by that action, whether or not they want the control gate fixed and whether or not they contribute anything to the repair. The key fact of life for coappropriators is that they are tied together in a lattice of interdependence so long as they continue to share a single CPR. The physical interdependence does not disappear when effective institutional rules are utilized in the governance and management of the CPR. The physical interdependence remains; what changes is the result the appropriators obtain.

When appropriators act independently in relationship to a CPR generating scarce resource units, the total net benefits they obtain usually will be less than could have been achieved if they had coordinated their strategies in some way. At a minimum, the returns they receive from their appropriation efforts will be lower when decisions are made independently than they

would have been otherwise. At worst, they can destroy the CPR itself. As long as the appropriators stay "unorganized," they cannot achieve a joint return as high as they could have received if they had organized in some way to undertake collective action. Mancur Olson stated the key problem facing appropriators who rely on a single CPR:

> ... when a number of individuals have a common or collective interest – when they share a single purpose or objective – individual, *unorganized* action [either will] not be able to advance that common interest at all, or will not be able to advance that interest adequately. (Olson 1965, p. 7; emphasis added)

Prisoners who have been placed in separate cells and cannot communicate with one another are also in an interdependent situation in which they must act independently. Acting independently in this situation is the result of coercion, not its absence. The herders in Hardin's model also act independently. Each decides on the number of animals to put on the meadow without concern for how that will affect the actions chosen by others.

At the most general level, the problem facing CPR appropriators is one of organizing: how to change the situation from one in which appropriators act independently to one in which they adopt coordinated strategies to obtain higher joint benefits or reduce their joint harm. That does not necessarily mean creating an organization. Organizing is a process; an organization is the result of that process. An organization of individuals who constitute an ongoing enterprise is only one form of organization that can result from the process of organizing.

The core of organization involves changes that order activities so that sequential, contingent, and frequency-dependent decisions are introduced where simultaneous, noncontingent, and frequency-independent actions had prevailed.[10] Almost all organization is accomplished by specifying a sequence of activities that must be carried out in a particular order.[11] Because of the repeated situations involved in most organized processes, individuals can use contingent strategies in which cooperation will have a greater chance of evolving and surviving. Individuals frequently are willing to forgo immediate returns in order to gain larger joint benefits when they observe many others following the same strategy. By requiring the participation of a minimal set of individuals, organizations can draw on this frequency-dependent behavior to obtain willing contributions on the part of many others. Changing the positive and negative inducements associated with particular actions and outcomes and the levels and types of information available can also encourage coordination of activities.[12]

Unlike prisoners, most CPR appropriators are not coerced into acting independently. Making the switch, however, from independent to co-

ordinated or collective action is a nontrivial problem. The costs involved in transforming a situation from one in which individuals act independently to one in which they coordinate activities can be quite high. And the benefits produced are shared by all appropriators, whether or not they share any of the costs of transforming the situation. Empirically, we know that some appropriators are able to solve this problem, and some are not. Theoretically, we do not have a coherent explanation for why some succeed and others fail.

The theory of the firm and the theory of the state can each provide an explanation for one way in which collective action can be achieved. Each involves the creation of a new institutional arrangement in which the rules in use are fundamentally different from those that structure independent action. Let us briefly and in a stylized fashion consider how each theory can "solve" the problem of independent action in an interdependent situation. By doing this, we can better illustrate the absence of a similar theory that would identify the mechanisms by which a group of individuals could organize themselves.

The theory of the firm

In the theory of the firm, an entrepreneur recognizes an opportunity to increase the return that can be achieved when individuals are potentially involved in an interdependent relationship.[13] The entrepreneur then negotiates a series of contracts with various participants that specify how they are to act in a coordinated, rather than independent, fashion. Each participant voluntarily chooses whether or not to join the firm, but gives up to the entrepreneur discretion over some range of choices. The participants become the agents of the entrepreneur. After paying each of the agents, the entrepreneur retains residual profits (or absorbs losses).

Consequently, the entrepreneur is highly motivated to organize the activity in a manner as efficient as possible. The entrepreneur attempts to craft contracts with agents that will induce them to act so as to increase the returns to the entrepreneur, and the entrepreneur monitors the agents' performances. The entrepreneur can terminate the contract of an agent who does not perform to the satisfaction of the entrepreneur. Because agents freely decide whether or not to accept the terms of the entrepreneur's contract, the organization is considered private, voluntary, and, at least by some individuals, nonexploitative. If there are large residuals to be obtained, however, it is the entrepreneur, not the agents, who receives them.[14] When a firm is located in an open market, one can presume that

external competition will pressure the entrepreneur toward developing efficient internal institutions.

The theory of the state

The theory of the state can also be presented in a brief and stylized version. Instead of an entrepreneur, we posit a ruler who recognizes that substantial benefits can be obtained by organizing some activities. As Hobbes first formulated the theory, individuals who independently engage in protection activities overinvest in weapons and surveillance and consequently live in constant fear. If a ruler gains a monopoly on the use of force, the ruler can use coercion as the fundamental mechanism to organize a diversity of human activities that will produce collective benefits. The ruler obtains taxes, labor, or other resources from subjects by threatening them with severe sanctions if they do not provide the resources.

The "wise" ruler uses the resources thus obtained to increase the general level of economic well-being of the subjects to a degree sufficient that the ruler can increase tax revenues while being able to reduce the more oppressive uses of coercion. Rulers, like entrepreneurs, keep the residuals. Subjects, like agents, may be substantially better off as a result of subjecting themselves to the coercion exercised by rulers. If the effort is highly successful, the ruler captures a substantial portion of the surplus.[15] There is no mechanism, such as a competitive market, that would exert pressure on the ruler to design efficient institutions. The ruler may face rebellion if the measures selected are too repressive, or military defeat if the realm is not adequately organized to do well in warfare.

In both the theory of the firm and the theory of the state, the burden of organizing collective action is undertaken by one individual, whose returns are directly related to the surplus generated. Both involve an outsider taking primary responsibility for supplying the needed changes in institutional rules to coordinate activities. The entrepreneur or the ruler makes credible commitments to punish anyone who does not follow the rules of the firm or the state. Because they gain the residuals, it is in their interest to punish nonconformance to their rules if they are confronted with nonconformance. Consequently, their threats to punish are credible (Schelling 1960; Williamson 1983). It is also in their interest to monitor the actions of agents and subjects to be sure they conform to prior agreements. Both theories thus address how a new institutional arrangement can come about, how credible commitments can be made, and why monitoring must be supplied.[16]

THREE PUZZLES: SUPPLY, COMMITMENT, AND MONITORING

Although the theory of the firm and the theory of the state can resolve these problems, no equivalently well developed and generally accepted theory provides a coherent account for how a set of principals, faced with a collective-action problem, can solve (1) the problem of supplying a new set of institutions, (2) the problem of making credible commitments, and (3) the problem of mutual monitoring.

The problem of supply

In a recent commentary on contractarianism and the new institutionalism, Robert Bates (1988) raises the issue that modern institutional theories do not adequately address the problem of supply. As he points out, "the new institutionalism is contractarian in spirit. Institutions are demanded because they enhance the welfare of rational actors. The problem is: Why are they supplied?" Bates first examines assurance games, where suppling new rules is considered easier to accomplish than it is in PD games, because there are mutually beneficial outcomes that are potential equilibria in the sense that once reached, no one has an incentive independently to switch strategies. Equilibria in assurance games do not, however, necessarily reward participants equally. Participants prefer a set of rules that will give them the most advantageous outcome. Although all will prefer a new institution that will enable them to coordinate their activities to achieve one of these equilibria, in contrast to continuing their independent actions, a fundamental disagreement is likely to arise among participants regarding which institution to choose. "The proposed solution to coordination – or assurance – games thus itself constitutes a collective dilemma" (Bates 1988, p. 394).[17]

Bates then turns to problems faced by a set of symmetric principals facing a collective dilemma in which all would benefit from a change in rules. Because supplying a new set of rules is the equivalent of providing another public good, the problem faced by a set of principals is that obtaining these new rules is a second-order collective dilemma.

> Even if the payoffs were symmetric and all persons were made [equally] better off from the introduction of the institutions, there would still be a failure of supply, since the institution would provide a collective good and rational individuals would seek to secure its benefits for free. The incentives to free-ride would undermine the incentives to organize a solution to the collective dilemma. It is subject to the very incentive problems it is supposed to resolve.
>
> (Bates 1988, pp. 394–5)

Because Bates presumes that the second-order dilemma is no easier to solve than the initial dilemma, he concludes that a new set of rules to solve the collective dilemma will not be provided by a set of principals (M. Taylor 1987).

Bates finds this deeply puzzling as it is obvious to him that some individuals in field settings *do* solve the problem of supply. Bates wishes to remain an institutionalist and a rational-choice theorist. His approach to addressing the inadequacy of current theories to explain how individuals supply their own rules is to turn for inspiration to some of the recent work in the theory of repeated games under uncertainty. Kreps and associates (1982) have demonstrated that in a finitely repeated PD game, some uncertainty about the exact payoff to a player can produce cooperative equilibria, as well as many other equilibria. Given this, it will pay one player to signal to other players an intention to cooperate, in the hope that they will reciprocate for a series of mutually productive plays. Thus, establishing trust and establishing a sense of community are, in Bates's view, mechanisms for solving the problem of supplying new institutions.

> Driven by a concern with institutions, we re-enter the world of the behavioralists. But we do so not in protest against the notion of rational choice, but rather in an effort to understand how rationality on the part of individuals leads to coherence at the level of society. (Bates 1988, p. 399)

Bates's approach is similar to the approach taken in this volume.

The problem of credible commitment

A second puzzle to be solved in explaining how a set of principals can organize themselves to obtain long-term collective benefits is the problem of commitment.[18] To understand the heart of the "commitment" problem, let us consider a highly simplified picture of the choices available to appropriators in CPR situations.[19] In all cases in which individuals have organized themselves to solve CPR problems, rules have been established by the appropriators that have severely constrained the authorized actions available to them. Such rules specify, for example, how many resource units an individual can appropriate, when, where, and how they can be appropriated, and the amounts of labor, materials, or money that must be contributed to various provisioning activities. If everyone, or almost everyone, follows these rules, resource units will be allocated more predictably and efficiently, conflict levels will be reduced, and the resource system itself will be sustained over time.

During an initial time period, an appropriator, calculating his or her

estimated future flow of benefits if most appropriators agree to follow a proposed set of rules, may agree to abide by the set of rules in order to get others to agree. During later time periods, the immediate return to the appropriator for breaking one or another of the rules frequently can be high. When an irrigator's crops are severely stressed, the financial benefit of taking water "out of turn" can be substantial. Breaking the rules may save an entire crop from drought. On many occasions after an initial agreement to a set of rules, each appropriator must make further choices. Minimally, the choice at each decision time subsequent to the agreement can be thought of as the choice between complying to a set of rules, C_t, or breaking the set of rules in some fashion, B_t. On many occasions, B_t will generate a higher immediate return for the appropriator than will C_t, unless B_t is detected and a sanction, S, is imposed that makes $C_t > B_t - S$.[20]

At the beginning of the process, all appropriators know the general configuration of the commitment problem. If they wish to change their appropriation rules, for example, to rotate the authority to withdraw water from an irrigation system among authorized appropriators, how does one appropriator credibly commit himself or herself to follow a rotation system when everyone knows that the temptation to break that commitment will be extremely strong in future time periods? Each appropriator can pledge: "I will keep my commitment if you keep yours." But when the temptation arises, how do past commitments bind the appropriator to future sacrifices? And given that it may be possible to steal water without being observed, how do the other appropriators know that commitments are actually being kept? No one wants to be a "sucker," keeping a promise that everyone else is breaking.

External coercion is a frequently cited theoretical solution to the problem of commitment (Schelling 1984). The presumption is made that if individuals commit themselves to a contract whereby a stiff sanction ($S > B_{max}$) will be imposed by an external enforcer to ensure compliance during all future time periods, then each can make a credible commitment and obtain benefits that would not otherwise be attainable. External coercion is at times a sleight-of-hand solution, because the theorist does not address what motivates the external enforcer to monitor behavior and impose sanctions. That is not, however, the issue at hand; it will be discussed later. The immediate issue is that a self-organized group must solve the commitment problem without an external enforcer. They have to motivate themselves (or their agents) to monitor activities and be willing to impose sanctions to keep conformance high.

These puzzles cumulate. Even if one appropriator took the time and effort to analyze the problems they faced and to devise a set of rules that

could improve their joint returns, the effort at supply would be pointless unless the appropriators could commit themselves to follow the rules. Unless the monitoring problem can be solved, credible commitments cannot be made. So let us now address the problem of mutual monitoring.

The problem of mutual monitoring

The question of how a set of principals can engage in mutual monitoring of conformance to a set of their own rules is not easily addressed within the confines of collective-action theory. In fact, the usual theoretical prediction is that they will not do so. The usual presumption that individuals will not themselves monitor a set of rules, even if they have devised those rules themselves, was summarized by Jon Elster in a recent discussion of the motivations for workers to monitor each other's participation in a union:

> Before a union can force or induce workers to join it must overcome a free-rider problem in the first place. To assume that the incentives are offered in a decentralized way, by mutual monitoring, gives rise to a second-order free-rider problem. Why, for instance, should a rational, selfish worker ostracize or otherwise punish those who don't join the union? What's in it for him? True, it may be better for all members if all punish non-members than if none do, but for each member it may be even better to remain passive. Punishment almost invariably is costly to the punisher, while the benefits from punishment are diffusely distributed over the members. It is, in fact, a public good: To provide it, one would need second-order selective incentives which would, however, run into a third-order free-rider problem. (Elster 1989, pp. 40–1)[21]

Dilemmas nested inside dilemmas appear to be able to defeat a set of principals attempting to solve collective-action problems through the design of new institutions to alter the structure of the incentives they face. Without monitoring, there can be no credible commitment; without credible commitment, there is no reason to propose new rules. The process unravels from both ends, because the problem of supply is presumed unsolvable in the first place. But some individuals have created institutions, committed themselves to follow rules, and monitored their own conformance to their agreements, as well as their conformance to the rules in a CPR situation. Trying to understand how they have done this is the challenge of this study.

FRAMING INQUIRY

Understanding how individuals solve particular problems in field settings requires a strategy of moving back and forth from the world of theory to the world of action. Without theory, one can never understand the general

underlying mechanisms that operate in many guises in different situations. If not harnessed to solving empirical puzzles, theoretical work can spin off under its own momentum, reflecting little of the empirical world.

When theoretical predictions and empirical observations are inconsistent, adjustments in theory are needed.[22] Predictions that individuals will not devise, precommit to, and monitor their own rules to change the structure of interdependent situations so as to obtain joint benefits are inconsistent with evidence that some individuals have overcome these problems, though others have not.

Theories affect the way that a problem is framed, not simply the particular assumptions used in an explanation. The way a problem is framed affects which questions are asked and what one looks for in conducting empirical inquiries. Several of the presumptions that have framed the way that scholars have approached the analysis of collective action have led them to an overly pessimistic view of the capacity of individuals to restructure their own interdependent situations.

Scholars addressing the problem of collective action frequently presume (1) that the underlying structure is always that of a PD game and (2) that one level of analysis is sufficient. When CPR problems are conceptualized as collective-action problems – a useful way to think of them – these same presumptions continue to frame the analyses, leading to the policy prescriptions described in Chapter 1. Consequently, part of the strategy pursued in this inquiry is to start from an alternative set of initial presumptions:

1 Appropriators in CPR situations face a variety of appropriation and provision problems whose structures vary from one setting to another, depending on the values of underlying parameters.

2 Appropriators must switch back and forth across arenas and levels of analysis.

These presumptions lead me to examine questions in a manner somewhat different from that of an analyst using the "normal" presuppositions of collective-action theory, although I still rely heavily on the work of other scholars.

Appropriation and provision problems

Although some interdependent CPR situations have the structure of a PD game, many do not. Several scholars have shown how some simple situations facing appropriators may be better characterized as "assurance" games and as the game known as "chicken" (Runge 1981, 1984a; M.

Taylor 1987; M. Taylor and Ward 1982). The underlying problem facing the appropriators in the Alanya fishing grounds discussed in Chapter 1 cannot be represented as a PD game. A formal analysis shows that it has the structure of an "assignment" game (Gardner and E. Ostrom 1990). In many irrigation systems similar to those discussed in Chapter 3, the fundamental choices facing appropriators are whether or not to steal water and whether or not to monitor the behaviors of others who might be stealing. The resulting game structure is complex and does not reduce down to any simple game. It does not have a single equilibrium. The amounts of stealing and monitoring that occur will depend on the values of parameters such as the number of appropriators, the cost of monitoring, the benefit from stealing, the punishment imposed when stealing is discovered, and the reward that a monitor receives for detecting a rule-breaker (Weissing and E. Ostrom 1990).

Consequently, instead of presuming that all CPR situations involve one underlying structure, I presume that the appropriators relying on any CPR face a variety of problems to be solved. The structure of these problems will depend on the values of underlying parameters, such as the value and predictability of the flow of resource units, the ease of observing and measuring appropriator activities, and so forth. In an effort to develop a unified framework within which to organize the analysis of CPR situations using the tools of game theory and institutional analysis and the findings from empirical studies in laboratory and field settings, Roy Gardner, James Walker, and I have found it most useful to cluster the problems facing CPR appropriators into two broad classes: appropriation problems and provision problems (Gardner et al. 1990).

When appropriators face appropriation problems, they are concerned with the effects that various methods of allocating a fixed, or time-independent, quantity of resource units will have on the net return obtained by the appropriators. Provision problems concern the effects of various ways of assigning responsibility for building, restoring, or maintaining the resource system over time, as well as the well-being of the appropriators. Appropriation problems are concerned with the allocation of the flow; provision problems are concerned with the stock. Appropriation problems are time-independent; provision problems are time-dependent. Both types of problems are involved in every CPR to a greater or lesser extent, and thus the solutions to one problem must be congruent with solutions to the other. The structure of an appropriation problem or a provision problem will depend on the particular configuration of variables related to the physical world, the rules in use, and the attributes of the individuals involved in a specific setting.

Appropriation problems. In regard to appropriation, the key problem in a CPR environment is how to allocate a fixed, time-independent quantity of resource units so as to avoid rent dissipation and reduce uncertainty and conflict over the assignment of rights. Rents are dissipated whenever the marginal returns from an appropriation process are smaller than the marginal costs of appropriation. Rent dissipation can occur because too many individuals are allowed to appropriate from the resource, because appropriators are allowed to withdraw more than the economically optimal quantity of resource units, or because appropriators overinvest in appropriation equipment (e.g., fishing gear).

In an open-access[23] CPR, in which no limit is placed on who can appropriate, the time-independent appropriation process frequently can be characterized as a PD game.[24] Rent dissipation is likely to be endemic. No appropriator has any incentive to leave any resource units for other appropriators to harvest (Gordon 1954; Scott 1955). In a limited-access CPR, in which a well-defined group of appropriators must jointly rely on a CPR for access to resource units, the incentives facing the appropriators will depend on the rules governing the quantity, timing, location, and technology of appropriation and how these are monitored and enforced. The structure of a limited-access CPR is not a PD game (Dasgupta and Heal 1979, p. 59) and lacks a dominant strategy for each participant. The incentives of appropriators who act independently, however, will lead them to overinvest in any input factor that is not constrained under the current rules (Townsend and Wilson 1987).

A second type of appropriation problem relates to assignment of spatial or temporal access to the resource. This occurs because spatial and temporal distributions of resource units frequently are heterogeneous and uncertain. Many fishing grounds, such as Alanya, are characterized by "fishing sites" that vary in their productivity. In grazing areas, one region may be drowned out in one year, but lush with growth in another year. Farmers who extract water from the head of an irrigation system can obtain more water than farmers who are located at the tail end. The risks associated with geographic or temporal uncertainty can be very high. Physical works, particularly those with storage, involve somewhat reduced risks, but well-enforced rules to allocate time or location of use or the quantity of resource units to specific users can reduce risks still further if the rules are well crafted to fit the physical attributes of the resource system. If risks are sufficiently reduced, appropriators can invest in productive enterprises that would not otherwise be economically viable. Physical violence occurring among the users of fisheries and irrigation systems is symptomatic of inadequate assignments of spatial or temporal slots to appropriators. When

appropriators consider the assignment of access rights and duties to be unfair, uneconomic, uncertain, or inappropriately enforced, that can adversely affect their willingness to invest in provision activities. The particular rules used to regulate appropriation will affect monitoring and policing costs and the type of strategic behavior that will occur between appropriators and monitors (the detection/deterrence game).[25]

Provision problems. Analyses of provision problems focus on the time-dependent, productive nature of investment in the resource itself. Provision problems may occur on the supply side, on the demand side, or on both sides. The supply-side problem faced in a CPR environment is related to the construction of the resource itself and its maintenance. Construction problems are like any long-term investment in capital infrastructure. Maintenance problems involve determining the type and level of regular maintenance (and reserves for emergency repair) that will sustain the resource system over time. Given that an investment in maintenance will affect the future rate at which a capital infrastructure will deteriorate, decisions about these activities are difficult to make even when a single entrepreneur makes them. When this difficult long-term problem is combined with the free-riding incentives of multiple appropriators, we see that organizing to maintain a system is a challenging task.

Supply-side provision problems are similar to the supply-side problems in providing a continuing, rather than a one-shot, public good. If appropriators act independently, they can expect that less than an optimal effort will be devoted to the construction, and particularly to the maintenance, of the system because of free-riding. What makes the problem more difficult in a CPR situation than in a public-goods situation is that unless appropriation problems are resolved, the provision problems may prove intractable. In a public-goods situation, appropriation problems do not exist, because resource units are not subtractable.

Demand-side provision problems involve regulating withdrawal rates so that they do not adversely affect the resource itself. Many of the dynamic models of "rent dissipation" in the fisheries literature (Clark 1980; Clark, Munro, and Charles 1985) have focused on the time-dependent relationship between current withdrawals and future yields. The same rules that affect the allocation of this year's resource units will have an impact on the availability of resource units next year and the years thereafter.

The underlying uniformities of all CPR situations relate to the nonseparability of one's choice of strategy and the choices made by others, as well as the fact that solving provision problems depends on achieving

adequate solutions to appropriation problems, not the particular game-theoretical representations for these commonalities.[26] Many factors affect the strategic structure of a particular appropriation or provision problem, including the physical structure of a particular CPR, the technology available to the appropriators, the economic environment, and the sets of rules that affect the incentives that appropriators face. As Oliver (1980, p. 1,359) stressed after reviewing many of the efforts to present "the" model of collective action, "there is no one 'right' way to model collective action: different models imply different assumptions about the situation and lead to substantively different conclusions."

Multiple levels of analysis

Most current analyses of CPR problems and related collective-action problems focus on a single level of analysis – what can be called the operational level of analysis (Kiser and E. Ostrom 1982). At the operational level of analysis, one assumes that both the rules of the game and the physical, technological constraints are given and will not change during the time frame of analysis: The actions of individuals in an operational situation directly affect the physical world. Resource units are withdrawn from a CPR. Inputs are transformed into outputs. Goods are exchanged. Appropriation and provision problems occur at an operational level. When doing an analysis of an operational situation, it is necessary for the analyst to assume that the technology and the institutional rules are known and unchanging. Both technology and rules are, however, subject to change over time. Analysis of technological changes has proved to be far more difficult than analysis of production and consumption decisions within a fixed technology (Dosi 1988; Nelson and Winter 1982). Analysis of institutional change is also far more difficult than analysis of operational decisions within a fixed set of rules.[27] The rules affecting operational choice are made within a set of collective-choice rules that are themselves made within a set of constitutional-choice rules. The constitutional-choice rules for a micro-setting are affected by collective-choice and constitutional-choice rules for larger jurisdictions. Individuals who have self-organizing capabilities switch back and forth between operational-, collective-, and constitutional-choice arenas, just as managers of production firms switch back and forth between producing products within a set technology, introducing a new technology, and investing resources in technology development. Given that CPR appropriators in some of the cases to be discussed in this volume do switch back and forth between arenas, we must drop the framing assumption that analysis at a single level will be sufficient. It is also essential to clarify what is meant by "institutions" in the first place.

"Institutions" can be defined as the sets of working rules that are used to determine who is eligible to make decisions in some arena, what actions are allowed or constrained, what aggregation rules will be used, what procedures must be followed, what information must or must not be provided, and what payoffs will be assigned to individuals dependent on their actions (E. Ostrom 1986a). All rules contain prescriptions that forbid, permit, or require some action or outcome. Working rules are those actually used, monitored, and enforced when individuals make choices about the actions they will take (Commons 1957). Enforcement may be undertaken by others directly involved, agents they hire, external enforcers, or any combination of these enforcers. One should not talk about a "rule" unless most people whose strategies are affected by it know of its existence and expect others to monitor behavior and to sanction nonconformance. In other words, working rules are common knowledge and are monitored and enforced. Common knowledge implies that every participant knows the rules, and knows that others know the rules, and knows that they also know that the participant knows the rules.[28] Working rules are always monitored and enforced, to some extent at least, by those directly involved. In any repetitive situation, one can assume that individuals come to know, through experience, good approximations of the levels of monitoring and enforcing involved.

Working rules may or may not closely resemble the formal laws that are expressed in legislation, administrative regulations, and court decisions. Formal law obviously is a major source of working rules in many settings, particularly when conformance to them is actively monitored and sanctions for noncompliance are enforced. When one speaks about a system that is governed by a "rule of law," this expresses the idea that formal laws and working rules are closely aligned and that enforcers are held accountable to the rules as well as others. In many CPR settings, the working rules used by appropriators may differ considerably from legislative, administrative, or court regulations (Wade 1988). The difference between working rules and formal laws may involve no more than filling in the lacunae left in a general system of law. More radically, operational rules may assign de facto rights and duties that are contrary to the de jure rights and duties of a formal legal system. My primary focus in this study will be on the de facto rules actually used in CPR field settings, in an effort to understand the incentives and consequences they produce.

All rules are nested in another set of rules that define how the first set of rules can be changed.[29] This nesting of rules within rules at several levels is similar to the nesting of computer languages at several levels. What can be done at a higher level will depend on the capabilities and limits of the

software (rules) at that level, on the software (rules) at a deeper level, and on the hardware (the CPR). Whenever one addresses questions about *institutional change*, as contrasted to action within institutional constraints, it is essential to recognize the following:

1 Changes in the rules used to order action at one level occur within a currently "fixed" set of rules at a deeper level.

2 Changes in deeper-level rules usually are more difficult and more costly to accomplish, thus increasing the stability of mutual expectations among individuals interacting according to a set of rules.

It is useful to distinguish three levels of rules that cumulatively affect the actions taken and outcomes obtained in using CPRs (Kiser and E. Ostrom 1982). *Operational rules* directly affect the day-to-day decisions made by appropriators concerning when, where, and how to withdraw resource units, who should monitor the actions of others and how, what information must be exchanged or withheld, and what rewards or sanctions will be assigned to different combinations of actions and outcomes. *Collective-choice rules* indirectly affect operational choices. These are the rules that are used by appropriators, their officials, or external authorities in making policies – the operational rules – about how a CPR should be managed. *Constitutional-choice rules* affect operational activities and results through their effects in determining who is eligible and determining the specific rules to be used in crafting the set of collective-choice rules that in turn affect the set of operational rules. One can think of the linkages among these rules and the related level of analysis at which humans make choices and take actions, as shown in Figure 2.2. The processes of appropriation, provision, monitoring, and enforcement occur at the operational level. The processes of policy-making, management, and adjudication of policy decisions occur at the collective-choice level. Formulation, governance, adjudication, and modification of constitutional decisions occur at the constitutional level.[30]

This nesting of rules within rules is the source of considerable confusion and debate. Institutional theorists, who have attempted to make the choice of rules endogenous to an analysis, have been criticized because it is necessary to assume the presence of some rules that govern the choice of other rules.[31] Making the choice of operational-level rules endogenous does not imply making the choice of collective-choice or constitutional-choice rules endogenous at the same time. For purposes of analysis, the theorist has to assume that some rules already exist and are exogenous for purposes of a particular analysis. The fact that they are held constant and unchanging

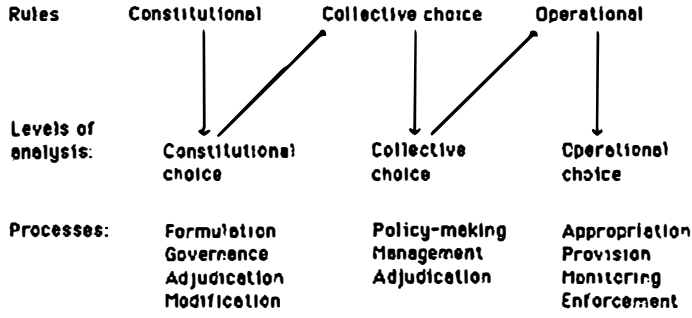


Figure 2.2. Linkages among rules and levels of analysis.

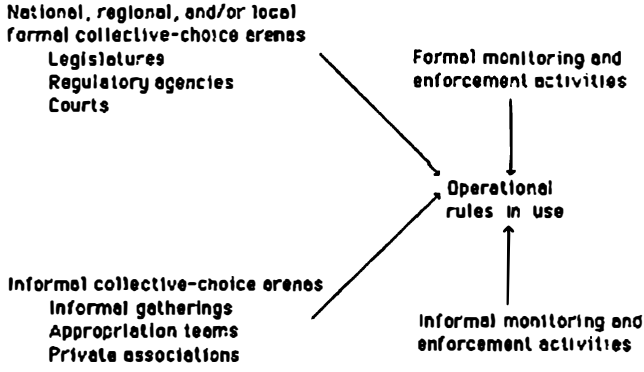


Figure 2.3. Relationships of formal and informal collective-choice arenas and CPR operational rules.

during analysis, however, does not mean that they cannot be changed. Those very same rules may themselves be the objects of choice in a separate analysis or in the context of a different area of choice. At the end of every season, for example, intercollegiate sports leagues consider whether or not to alter the rules of the game for the next season.

On the other hand, rules are changed less frequently than are the strategies that individuals adopt within the rules. Changing the rules at any level of analysis will increase the uncertainty that individuals will face. Rules provide stability of expectations, and efforts to change rules can rapidly

reduce that stability. Further, it is usually the case that operational rules are easier to change than collective-choice rules, and collective-choice rules are easier to change than constitutional-choice rules. Analyses of deeper layers of rules are more difficult for scholars and participants to make. Deciding whether an irrigation association should use a legislative body of five or nine members will depend on the physical and historical environment and the analyst's speculation about different outcomes at several levels.[32]

When doing analysis at any one level, the analyst keeps the variables of a deeper level fixed for the purpose of analysis. Otherwise, the structure of the problem would unravel. But self-organizing and self-governing individuals trying to cope with problems in field settings go back and forth across levels as a key strategy for solving problems. Individuals who have no self-organizing and self-governing authority are stuck in a single-tier world. The structure of their problems is given to them. The best they can do is to adopt strategies within the bounds that are given.

At each level of analysis there may be one or more arenas in which the types of decisions made at that level will occur. The concept of an "arena" does not imply a formal setting, but can include such formal settings as legislatures and courts. An arena is simply the situation in which a particular type of action occurs. Policy-making regarding the rules that will be used to regulate operational-level choices is carried out in one or more collective-choice arenas. If the appropriators using a CPR change at least some of the working rules used to organize appropriation and provision, the arena in which collective-choice decisions will be made may be a local coffeehouse, the meetings of a producers' co-op, or the meetings of an organization that has been set up specifically for the purpose of managing and governing this CPR and possibly others related to it. If the appropriators using a CPR cannot change the rules used to organize operational choices, then the only arenas for collective choice are external to the CPR appropriators. In such cases, choices about the rules to be used will be made by government officials in bureaucratic structures, by elected representatives in local or national legislatures, and by judges in judicial arenas.

The relationships among arenas and rules rarely involve a single arena related to a single set of rules. Most frequently, several collective-choice arenas affect the set of operational rules actually used by appropriators for making choices about harvesting and investment strategies in a CPR. Decisions made in national legislatures and courts concerning access to all resources of particular types, when given legitimacy in a local setting and enforced, are likely to affect the operational rules actually used in particular locations. The relationships among formal and informal collective-choice arenas and the resulting operational rules are illustrated in Figure

2.3. Similarly, formal and informal constitutional-choice processes may occur in local, regional, and/or national arenas.

That the working rules used by appropriators may have multiple sources, and may include de facto as well as de jure rules, greatly complicates the problem of understanding behaviors and outcomes in particular locations and the problem of improving outcomes. The absence of national, formal laws regulating the appropriation from and provision of a CPR is not equivalent to the absence of effective rules. Over a long period of time, local appropriators may have developed working rules that constrain the entry to and use of a CPR. Such rules may or may not lead appropriators to manage their resource efficiently and fairly, but they will affect the strategies that appropriators perceive to be available to them and the resulting outcomes.

STUDYING INSTITUTIONS IN FIELD SETTINGS

In the cases described in Chapters 3, 4, and 5, I present only a fraction of the detailed information to be found in the in-depth case studies from which I draw. A reader is justified in wanting to know how I approach the task of reading in-depth case materials and abstracting from them for the purpose of studying how individuals supply their own institutions, how they commit themselves to conform to their own rules, and how they monitor each other's conformance to these rules. In general, I am relying on the method of institutional analysis that has been described elsewhere (Kiser and E. Ostrom 1982; Oakerson 1986; E. Ostrom 1986a,b) and applied in many papers, doctoral dissertations, and books (Blomquist 1988a–d; Gardner and E. Ostrom 1990; Kaminski in press; V. Ostrom 1989; V. Ostrom, Feeny, and Picht 1988; Sawyer 1989; Schaaf 1989; Tang 1989; Wynne 1988; Yang 1987).

The basic strategy is to identify those aspects of the physical, cultural, and institutional setting that are likely to affect the determination of who is to be involved in a situation, the actions they can take and the costs of those actions, the outcomes that can be achieved, how actions are linked to outcomes, what information is to be available, how much control individuals can exercise, and what payoffs are to be assigned to particular combinations of actions and outcomes. Once one has all the needed information, one can then abstract from the richness of the empirical situation to devise a playable game that will capture the essence of the problems individuals are facing.

To solve appropriation and provision problems, for example, individuals must learn about the structure of the physical system on which they jointly rely, about their own appropriation and use patterns, about the norms of

behavior that are followed in a community, about the incentives they will encourage or discourage as they change rules, and about how all of these factors will cumulatively affect their net benefits and costs over time. Individuals must assess what types of transactions costs will be involved in adopting various strategies within a set of rules or in changing those rules. If the analyst is to understand the structure of the situation, the analyst must learn about the same set of variables.

For the cases that I discuss in Chapter 3, I do not know what the structures of the situations were like before some appropriators in the mists of time began to experiment with various rules to allocate resource units and provisioning responsibilities. What I do know is that the appropriators in the "success" cases described in Chapter 3 were able to allocate resource units and at the same time avoid the conflict, uncertainty, and perceived unfairness of a poorly solved assignment problem, the overinvestment in appropriation efforts involved in an inadequately solved rent-dissipation problem, or the deterioration or destruction of the resources involved when provision problems remain unsolved.

Obviously, I do not know if these appropriators reached optimal solutions to their problems. I strongly doubt it. They solved their problems the way that most individuals solve difficult and complex problems: as well as they were able, given the problems involved, the information they had, the tools they had to work with, the costs of various known options, and the resources at hand. I see my task as one of learning about the structures of the problems they faced and why the rules they adopted seem to work.

This means that I first try to understand something about the structure of the resource itself – its size, clarity of boundary, and internal structure. Then I try to discover the flow patterns involved in the resource units: How much predictability is involved over time, across space, and in quantity? Given the economic circumstances of the appropriators, how reliant are they on the resource, and what are the risks involved in various potential types of allocation schemes? Lastly, I try to ascertain key attributes of the individuals: How many are involved? What are their time horizons likely to be? Are they involved in multiple activities together? Are their interests roughly similar or heterogeneous? Have they established prior norms of behavior that can be drawn on (or pose a disadvantage) in trying to solve these problems? Then I examine the rules that they have devised and try to understand how they work by searching for the design principles that are involved and how these affect the incentives of participants. Given that the appropriators in these cases have engaged in mutual monitoring and generally have kept their commitments to follow their rules to a substantial degree, I try to understand how they have been able to do this.

In Chapter 4, I use this framework again to identify the structure of the situation that existed before a group of appropriators attempted to change their rules to solve several interrelated provision and appropriation problems. Then I examine the process of devising new institutions, in order to address the question of the supply of institutions. The "failure" cases in Chapter 5 are characterized by extreme rent dissipation, unresolved disagreements leading to physical violence, or resource deterioration. The same framework is used to identify the variables that account for that lack of success in solving appropriation and provision problems. I again assume that the individuals involved tried to do as well as they could, given the constraints of the situation. Thus, the problem is to identify what those constraints were, using the same framework for analysis.

In the concluding portions of this study, I discuss how the findings derived from an analysis of these cases can be used to advance theoretical understanding of a theory of self-organized collective action to complement the existing theories of externally organized collective action: the theory of the firm and the theory of the state.

3

Analyzing long-enduring, self-organized, and self-governed CPRs

A direct attack on several of the key questions posed in this book can be launched by an examination of field settings in which (1) appropriators have devised, applied, and monitored their own rules to control the use of their CPRs and (2) the resource systems, as well as the institutions, have survived for long periods of time. The youngest set of institutions to be analyzed in this chapter is already more than 100 years old. The history of the oldest system to be examined exceeds 1,000 years. The institutions discussed in this chapter have survived droughts, floods, wars, pestilence, and major economic and political changes. We shall examine the organization of mountain grazing and forest CPRs in Switzerland and Japan and irrigation systems in Spain and the Philippine Islands.

By indicating that these CPR institutions have survived for long periods of time, I do not mean that their operational rules have remained fixed since they were first introduced. All of the environmental settings included in this chapter are complex and have varied over time. In such settings, it would be almost impossible to "get the operational rules right" on the first try, or even after several tries. These institutions are "robust" or in "institutional equilibrium" in the sense defined by Shepsle. Shepsle (1989b, p. 143) regards "an institution as 'essentially' in equilibrium if changes transpired according to an *ex ante* plan (and hence part of the original institution) for institutional change." In these cases, the appropriators designed basic operational rules, created organizations to undertake the operational management of their CPRs, and modified their rules over time in light of past experience according to their own collective-choice and constitutional-choice rules.

The cases in this chapter are particularly useful for gaining insight re-

garding how groups of self-organized principals solve two of the major puzzles discussed in Chapter 2: the problem of commitment and the problem of mutual monitoring. (The problem of supply of institutions is addressed in Chapter 4.) The continuing commitments of the appropriators to their institutions have been substantial in these cases. Restrictive rules have been established by the appropriators to constrain appropriation activities and mandate provisioning activities. Thousands of opportunities have arisen in which large benefits could have been reaped by breaking the rules, while the expected sanctions were comparatively low. Stealing water during a dry season in the Spanish *huertas* might on occasion save an entire season's crop from certain destruction. Avoiding spending day after day maintaining the Philippine irrigation systems might enable a farmer to earn needed income in other pursuits. Harvesting illegal timber in the Swiss or Japanese mountain commons would yield a valuable product. Given the temptations involved, the high levels of conformance to the rules in all these cases have been remarkable.

Sizable resources are invested in monitoring activities in these cases, but the "guards" are rarely "external" agents. Widely diverse monitoring arrangements are used. In all of them, the appropriators themselves play a major role in monitoring each other's activities. Even though mutual monitoring has aspects of being a second-order dilemma, the appropriators in these settings somehow solve this problem. Further, the fines assessed in these settings are surprisingly low. Rarely are they more than a small fraction of the monetary value that could be obtained by breaking the rules. In the conclusion to this chapter, I argue that commitment and monitoring are strategically linked and that monitoring produces private benefits for the monitor as well as joint benefits for others.

In explaining the robustness of these institutions and the resource systems themselves over time in environments characterized by high levels of uncertainty, one needs to search for the appropriate specificity of underlying commonalities that may explain this level of sustainability. Given the differences in environments and historical developments, one would hardly expect the particular rules used in these settings to be the same. And they are not. Given the length of time that they have had for trial-and-error learning about operational rules, the harshness of these environments as a stimulus toward improvement, and the low transformation costs in changing their own operational rules, one can, however, expect that these appropriators have "discovered" some underlying principles of good institutional design in a CPR environment. I do not claim that the institutions devised in these settings are in any sense "optimal." In fact, given the high

levels of uncertainty involved and the difficulty of measuring benefits and costs, it would be extremely difficult to obtain a meaningful measure of optimality.[1]

On the other hand, I do not hesitate to call these CPR institutions successful. In all instances the individuals involved have had considerable autonomy to craft their own institutions. Given the salience of these CPRs to the appropriators using them, and their capacity to alter rules in light of past performance, these appropriators have had the incentives and the means to improve these institutions over time. The Swiss and Japanese mountain commons have been sustained, if not enhanced, over the centuries while being used intensively. Ecological sustainability in a fragile world of avalanches, unpredictable precipitation, and economic growth is quite an accomplishment for any group of appropriators working over many centuries. Keeping order and maintaining large-scale irrigation works in the difficult terrain of Spain or the Philippine Islands have been similarly remarkable achievements. That record has not been matched by most of the irrigation systems constructed around the world during the past 25 years. Consequently, I have attempted to identify a set of underlying design principles shared by successful CPR institutions and to determine how those design principles affect the incentives of appropriators so that the CPRs themselves and the CPR institutions can be sustained over time. When in Chapter 5 we discuss cases in which appropriators were not able to devise or sustain institutional arrangements to solve CPR problems, we shall consider to what extent the design principles used by appropriators in the "success" cases also characterize the "failure" cases.

The cases discussed in this chapter also help us to examine two other questions. First, the CPR institutions related to the use of precarious and delicately balanced mountain commons to provide fodder and forest products in Switzerland and Japan, in particular, help us to confront the question of the presumed superiority of private-property institutions for most allocational purposes, and specifically those related to the uses of land. Although many resource economists admit that technical difficulties prevent the creation of private property rights to fugitive resources, such as groundwater, oil, and fish, almost all share the presumption that the creation of private property rights to arable or grazing land is an obvious solution to the problem of degradation. Dasgupta and Heal (1979, p. 77), for example, assert that when private property rights are introduced in areas of arable or grazing land, "the resource ceases to be common property and the problem is solved at one stroke."

Many property-rights theorists presume that one of two undesirable outcomes is likely under communal ownership: (1) the commons will be

destroyed because no one can be excluded, or (2) the costs of negotiating a set of allocation rules will be excessive, even if exclusion is achieved.[2] On the contrary, what one observes in these cases is the ongoing, side-by-side existence of private property and communal property in settings in which the individuals involved have exercised considerable control over institutional arrangements and property rights. Generations of Swiss and Japanese villagers have learned the relative benefits and costs of private-property and communal-property institutions related to various types of land and uses of land. The villagers in both settings have *chosen* to retain the institution of communal property as the foundation for land use and similar important aspects of village economies. The economic survival of these villagers has been dependent on the skill with which they have used their limited resources. One cannot view communal property in these settings as the primordial remains of earlier institutions evolved in a land of plenty. If the transactions costs involved in managing communal property had been excessive, compared with private-property institutions, the villagers would have had many opportunities to devise different land-tenure arrangements for the mountain commons.

Second, I have frequently been asked, when giving seminar presentations about the Swiss, Japanese, and Spanish institutions, if the same design principles are relevant for solving CPR problems in Third World settings. The last case discussed in this chapter – the *zanjera* institutions of the Philippines – provides a strong affirmative answer to this question. All of the design principles present in the Swiss, Japanese, and Spanish cases are also present in the Philippine case. An analysis of the underlying similarities of enduring CPR institutions, though based on a limited number of cases, may have broader applications.

COMMUNAL TENURE IN HIGH MOUNTAIN MEADOWS AND FORESTS[3]

Törbel, Switzerland

Our first case concerns Törbel, Switzerland, a village of about 600 people located in the Vispertal trench of the upper Valais canton, as described by Robert McC. Netting in a series of articles (1972, 1976) that were later incorporated into his book *Balancing on an Alp* (1981). Netting (1972, p. 133) identifies the most significant features of the general environment as "(1) the steepness of its slope and the wide range of microclimates demarcated by altitude, (2) the prevailing paucity of precipitation, and (3) the exposure to sunlight." For centuries, Törbel peasants have planted their privately owned plots with bread grains, garden vegetables, fruit trees, and

hay for winter fodder. Cheese produced by a small group of herdsmen, who tend village cattle pastured on the communally owned alpine meadows during the summer months, has been an important part of the local economy.

Written legal documents dating back to 1224 provide information regarding the types of land tenure and transfers that have occurred in the village and the rules used by the villagers to regulate the five types of communally owned property: the alpine grazing meadows, the forests, the "waste" lands, the irrigation systems, and the paths and roads connecting privately and communally owned properties. On February 1, 1483, Törbel residents signed articles formally establishing an association to achieve a better level of regulation over the use of the alp, the forests, and the waste lands.

> The law specifically forbade a foreigner (*Fremde*) who bought or otherwise occupied land in Törbel from acquiring any right in the communal alp, common lands, or grazing places, or permission to fell timber. Ownership of a piece of land did *not* automatically confer any communal right (*genossenschaftliches Recht*). The inhabitants currently possessing land and water rights reserved the power to decide whether an outsider should be admitted to community membership.
>
> (Netting 1976, p. 139)

The boundaries of the communally owned lands were firmly established long ago, as indicated in a 1507 inventory document.

Access to well-defined common property was strictly limited to citizens, who were specifically extended communal rights.[4] As far as the summer grazing grounds were concerned, regulations written in 1517 stated that "no citizen could send more cows to the alp than he could feed during the winter" (Netting 1976, p. 139). That regulation, which Netting reports to be still enforced, imposed substantial fines for any attempt by villagers to appropriate a larger share of grazing rights. Adherence to this "wintering" rule was administered by a local official (*Gewalthaber*) who was authorized to levy fines on those who exceeded their quotas and to keep one-half of the fines for himself. The wintering rule is used by many other Swiss villages as a means for allocating appropriation rights (frequently referred to as "cow rights") to the commons. This and other forms of cow rights are relatively easy to monitor and enforce. The cows are all sent to the mountain to be cared for by the herdsmen. They must be counted immediately, as the number of cows each family sends is the basis for determining the amount of cheese the family will receive at the annual distribution.

The village statutes are voted on by all citizens and provide the general legal authority for an alp association to manage the alp. This association includes all local citizens owning cattle. The association has annual meet-

ings to discuss general rules and policies and elect officials. The officials hire the alp staff, impose fines for misuse of the common property, arrange for distribution of manure on the summer pastures, and organize the annual maintenance work, such as building and maintaining roads and paths to and on the alp and rebuilding avalanche-damaged corrals or huts. Labor contributions or fees related to the use of the meadows usually are set in proportion to the number of cattle sent by each owner. Trees that will provide timber for construction and wood for heating are marked by village officials and assigned by lot to groups of households, whose members then are authorized to enter the forests and harvest the marked trees.

Private rights to land are well developed in Törbel and other Swiss villages. Most of the meadows, gardens, grainfields, and vineyards are owned by various individuals, and complex condominium-type agreements are devised for the fractional shares that siblings and other relatives may own in barns, granaries, and multistory housing units. The inheritance system in Törbel ensures that all legitimate offspring share equally in the division of the private holdings of their parents and consequently in access to the commons, but family property is not divided until surviving siblings are relatively mature (Netting 1972). Prior to a period of population growth in the nineteenth century, and hence severe population pressure on the limited land, the level of resource use was held in check by various population-control measures such as late marriages, high rates of celibacy, long birth spacing, and considerable emigration (Netting 1981).

Netting (1976, p. 140) dismisses the notion that communal ownership is simply an anachronistic holdover from the past by showing that for at least five centuries these Swiss villagers have been intimately familiar with the advantages and disadvantages of both private and communal tenure systems and have carefully matched particular types of land tenure to particular types of land use. He associates five attributes to land-use patterns with the differences between communal and individual land tenure. He argues that communal forms of land tenure are better suited to the problems that appropriators face when (1) the value of production per unit of land is low, (2) the frequency or dependability of use or yield is low, (3) the possibility of improvement or intensification is low, (4) a large territory is needed for effective use, and (5) relatively large groups are required for capital-investment activities. See Runge (1984a, 1986) and Gilles and Jamtgaard (1981) for similar arguments.

Communal tenure "promotes both general access to and optimum production from certain types of resources while enjoining on the entire community the conservation measures necessary to protect these resources from destruction" (Netting 1976, p. 145). Although yields are relatively

low, the land in Törbel has maintained its productivity for many centuries. Overgrazing has been prevented by tight controls. The CPR not only has been protected but also has been enhanced by investments in weeding and manuring the summer grazing areas and by the construction and maintenance of roads.

Netting is clear that Törbel should not be considered the prototype for all Swiss alpine villages. A recent review of the extensive German literature on common-property regimes in Swiss alpine meadows reveals considerable diversity of legal forms for governing alpine meadows (Picht 1987). However, Netting's major findings are consistent with experience in many Swiss locations. Throughout the alpine region of Switzerland, farmers use *private* property for agricultural pursuits and a form of *common* property for the summer meadows, forests, and stony waste lands near their private holdings. Four-fifths of the alpine territory is owned by some form of common property: by local villages (*Gemeinden*), by corporations, or by cooperatives. The remaining alpine territory belongs either to the cantons or to private owners or groups of co-owners (Picht 1987, p. 4). Some villages own several alpine meadows and reallocate grazing rights to the use of a specific meadow every decade or so (Stevenson 1990).

In addition to defining who has access to the CPR, all local regulations specify authority rules to limit appropriation levels (Picht 1987). In most villages, some form of proportional-allocation rule is used. The proportion is based on (1) the number of animals that can be fed over the winter,[5] (2) the amount of meadowland owned by a farmer, (3) the actual amount of hay produced by a farmer, (4) the value of the land owned in the valley, or (5) the number of shares owned in a cooperative. A few villages allow all citizens to send equal numbers of animals to the summer alp (Picht 1987, p. 13). Overuse of alpine meadows is rarely reported.[6] Where overuse has occurred, the combined effects of entry rules and authority rules have not sufficiently limited grazing practices, or else several villages have owned and used a single alp without an overarching set of rules (Picht 1987, pp. 17–18; Rhodes and Thompson 1975; Stevenson 1990).[7]

All of the Swiss institutions used to govern commonly owned alpine meadows have one obvious similarity – the appropriators themselves make all major decisions about the use of the CPR.

> The users/owners are the main decision making unit. They have to decide on all matters of importance and seem to have a considerable degree of autonomy. They can set up statutes and revise them, they can set limits for the use of the pastures and change them, they can adapt their organizational structure. . . . It can also be said that the user organizations are nested in a set of larger organizations (village, Kantone, Bund) in which they are perceived as legitimate. (Picht 1987, p. 28)

Thus, residents of Törbel and other Swiss villages who own communal land spend time governing themselves. Many of the rules they use, however, keep their monitoring and other transactions costs relatively low and reduce the potential for conflict. The procedures used in regard to cutting trees for timber – a valuable resource unit that can be obtained from communal forests – illustrate this quite well. The first step is that the village forester marks the trees ready to be harvested. The second step is that the households eligible to receive timber form work teams and equally divide the work of cutting the trees, hauling the logs, and piling the logs into approximately equal stacks. A lottery is then used to assign particular stacks to the eligible households. No harvesting of trees is authorized at any other time of the year. This procedure nicely combines a careful assessment of the condition of the forest with methods for allocating work and the resulting products that are easy to monitor and are considered fair by all participants. Combining work days or days of reckoning (where the summer's cheese is distributed and assessments are made to cover the costs of the summer's work) with festivities is another method for reducing some of the costs associated with communal management.

In recent times, the value of labor has risen significantly, thus representing an exogenous change for many Swiss villages. Common-property institutions are also changing to reflect differences in relative factor inputs. Villages that rely on unanimity rules for changing their common-property institutions are not adjusting as rapidly as are those villages that rely on less inclusive rules for changing their procedures.

Hirano, Nagaike, and Yamanoka villages in Japan

In Japan, extensive common lands have existed and have been regulated by local village institutions for centuries. In an important study of traditional common lands in Japan, Margaret A. McKean (1986) estimates that about 12 million hectares of forests and uncultivated mountain meadows were held and managed in common by thousands of rural villages during the Tokugawa period (1600–1867) and that about 3 million hectares are so managed today. Although many villages have sold, leased, or divided their common lands in recent times, McKean (1986, p. 534) indicates that she has "not yet turned up an example of a commons that suffered ecological destruction while it was still a commons" (McKean 1982).[8]

McKean provides both a general overview of the development of property law in Japan and a specific view of the rules, monitoring arrangements, and sanctions used in three Japanese villages – Hirano, Nagaike, and Yamanoka – for regulating the commons. The environmental conditions of

the villages studied by McKean are remarkably similar to those of Törbel. The villages are established on steep mountains where many microclimates can be distinguished. Peasants cultivate their own private lands, raising rice, garden vegetables, and horses. The common lands in Japan produce a wide variety of valuable forest products, including timber, thatch for roofing and weaving, animal fodder of various kinds, and decayed plants for fertilizer, firewood, and charcoal. The land held in communal tenure meets the previously cited five conditions that Netting posits as conducive to communal property rather than private property.

Each village in earlier times was governed by an assembly, usually composed of the heads of all the households that had been assigned decision-making authority in the village. The basis for political rights differed from one village to another. Rights were variously based on cultivation rights in land, taxpaying obligations, or ownership rights in land. In some villages, almost all households had political rights and rights to the use of the commons.[9] In others, such rights were more narrowly held (McKean 1986, p. 551; Troost 1985).

Ownership of the uncultivated lands near a village devolved from the imperial court to the villages through several intermediate stages involving land stewards and locally based warriors. National cadastral surveys were conducted late in the sixteenth century at a time of land reform that assigned "most of the rights that we today consider to be 'ownership' of arable land to peasants who lived on and cultivated that land" (McKean 1986, p. 537). Owners of large estates in the earlier systems had employed agents in the various villages and authorized those agents to regulate access to uncultivated lands. As villages asserted their own rights to these lands, they shared a clear image of which lands were private and which were held in common. They took the view that those lands held in common needed management in order to serve the long-term interests of the peasants dependent on them.

In traditional Japanese villages, the household was the smallest unit of account, but the *kumi*, composed of several households, was frequently used as an accounting and distributional unit related to the commons. Each village contained a carefully recorded, defined number of households. A household could not subdivide itself into multiple households without permission from the village. Rights of access to the communally held lands were accorded only to a household unit, not to individuals as such. Consequently, households with many members had no advantage, and considerable disadvantages, in their access to the commons. Population growth was extremely low (0.025% for the period 1721–1846), and ownership patterns within villages were stable (McKean 1986, p. 552).

In addition to delimiting the ownership status of all lands, village assemblies created detailed authority rules specifying in various ways how much of each valued product a household could harvest from the commons and under what conditions. The rules used in these villages, like those in the Swiss villages, were tailored to the specific environment, to the particular economic roles that various forest products played in the local economy, and to the need to minimize the costs of monitoring labor inputs, resource-unit outputs, and compliance with the rules. A village headman usually was responsible for determining the date when the harvesting of a given product could begin. For abundant plants, the date would be selected simply to ensure that plants had matured and had propagated themselves. No limit was placed on the amount to be gathered. For scarce products, various harvesting rules were used. The rules for allocating winter fodder for draft animals by one village from a closed reserve are illustrative:

> . . . each *kumi* was assigned a zone according to an annual rotation scheme, and each household had to send one, but only one adult. On the appointed day, each representative reported to the appropriate *kumi* zone in the winter fodder commons and waited for the temple bell as the signal to begin cutting. However, this grass was cut with large sickles, and since it would be dangerous to have people distributed unevenly around their *kumi* zone swinging sickles in all directions, the individuals in each *kumi* lined up together at one end of their zone and advanced to the other end, whacking in step with each other like a great agricultural drill team. The grass was left to dry . . . and then two representatives from each household entered the fodder commons to tie the grass up into equal bundles. The haul for each *kumi* was grouped together and then divided evenly into one cluster per household. Each household was then assigned its cluster by lottery.
>
> (McKean 1986, pp. 556 7)

Villagers were required to perform collective work to enhance and maintain the yield of the commons, such as annual burning or specific cutting of timber or thatch. Each household had an obligation to contribute a share to such efforts:

> There were written rules about the obligation of each household to contribute a share to the collective work to maintain the commons – to conduct the annual burning (which involved cutting nine-foot firebreaks ahead of time, carefully monitoring the blaze, and occasional fire-fighting when the flames jumped the firebreak), to report to harvest on mountain-opening days, or to do a specific cutting of timber or thatch. Accounts were kept about who contributed what to make sure that no household evaded its responsibilities unnoticed. Only illness, family tragedy, or the absence of able-bodied adults whose labor could be spared from routine chores were recognized as excuses for getting out of collective labor. . . . But, if there was no acceptable excuse, punishment was in order.
>
> (McKean 1986, p. 559)

Each of the villages also devised its own monitoring and sanctioning systems. Given that the mountain usually was closed, except for specified periods, anyone caught in the communally owned territories at other times obviously was not following the rules. Most of the villages hired "detectives" who daily patrolled the commons on horseback in groups of two looking for unauthorized users. In some villages, this position was considered "one of the most prestigious and responsible available to a young man" (McKean 1986, p. 561). In other villages, all eligible males rotated into these positions on a regular basis. One village that did not use formal detectives relied on a form of "citizen's arrest," and anyone was authorized to report violations.

The written codes for each village specified a series of escalating penalties for various violations of the rules to protect the commons, depending on the past behavior of the offender. An occasional infraction would be handled by the detective in a quiet and simple manner. "It was considered perfectly appropriate for the detective to demand cash and saké from violators and to use that as their own entertainment cache" (McKean 1986, p. 561). In addition to the fines paid to the detectives, violators were deprived of their contraband harvest, their equipment, and their horses. The village retained the illegal harvest. The rule-breaker had to pay a fine to the village to retrieve equipment and horses. Fines were graduated from very low levels to extremely high levels to reflect the seriousness of the offense and the willingness of the culprit to make adequate and rapid amends. The most serious sanctions that could be and occasionally were imposed involved complete ostracism or ultimately banishment from the village.

Although the level of rule compliance was very high, violations certainly occurred. McKean reports several types of infractions. Impatience with waiting for mountain-opening day was one reason. In the period just before the official opening of the commons for harvesting a particular plant, the detectives expected – and found – a higher level of infractions and were able to keep themselves well supplied with saké.

A second reason for rule violation sometimes was genuine disagreement about the management decisions of a village headman. McKean illustrates this type of infraction in the following way:

> One former detective in Hirano, now a respected village elder, described how he had been patrolling a closed commons one day and came upon not one or two intruders but thirty, including some of the heads of leading households. It was not yet mountain-opening day, but they had entered the commons en masse to cut a particular type of pole used to build trellises to support garden vegetables raised on private plots. If they could not cut the poles soon enough, their entire vegetable

crop might be lost, and they believed that the village headman had erred in setting opening day later than these crops required. (McKean 1986, p. 565)

In that instance, fines were imposed, but they involved making a donation to the village school, rather than the usual payment of saké. In her conclusion, McKean stresses that the long-term success of these locally designed rule systems indicates "that it is not necessary for regulation of the commons to be imposed coercively from the outside" (McKean 1986, p. 571).

HUERTA IRRIGATION INSTITUTIONS

On May 29, 1435, about 50 years before the residents of Törbel signed their formal articles of association, 84 irrigators served by the Benacher and Faitanar canals in Valencia gathered at the monastery of St. Francis to draw up and approve formal regulations. Those regulations specified who had rights to water from these canals, how the water would be shared in good years as well as bad, how responsibilities for maintenance would be shared, what officials they would elect and how, and what fines would be levied against anyone who broke one of their rules. The canals themselves, like many others in the region, had been constructed in even earlier times. Many rules concerning the distribution of irrigation water were already well established in customary practices. Valencia had been recaptured from the Muslims in 1238 – two centuries before that meeting of the Benacher and Faitanar irrigators. Some of the rules carried into medieval and modern practice were developed *prior* to that reconquest.[10] Thus, for at least 550 years, and probably for close to 1,000 years, farmers have continued to meet with others sharing the same canals for the purpose of specifying and revising the rules that they use, selecting officials, and determining fines and assessments.

Given the limited quantity of rainfall throughout this semiarid region and the extreme variation in rainfall from year to year, its highly developed agriculture would not have been possible without irrigation works bringing water to the farmers' fields. Water was never abundant in this region, not even after major canals were constructed. Given the high stakes, conflict over water has always been just beneath the surface of everyday life, erupting from time to time in fights between the irrigators themselves, between irrigators and their own officials, and between groups of irrigators living in the lower reaches of the water systems and their upstream neighbors. Despite this high potential for conflict – and its actual realization from time to time – the institutions devised many centuries ago for govern-

ing the use of water from these rivers have proved adequate for resolving conflicts, allocating water predictably, and ensuring stability in a region not normally associated with high levels of stability. Maass and Anderson (1986) have devoted much effort to studying the institutions used in the

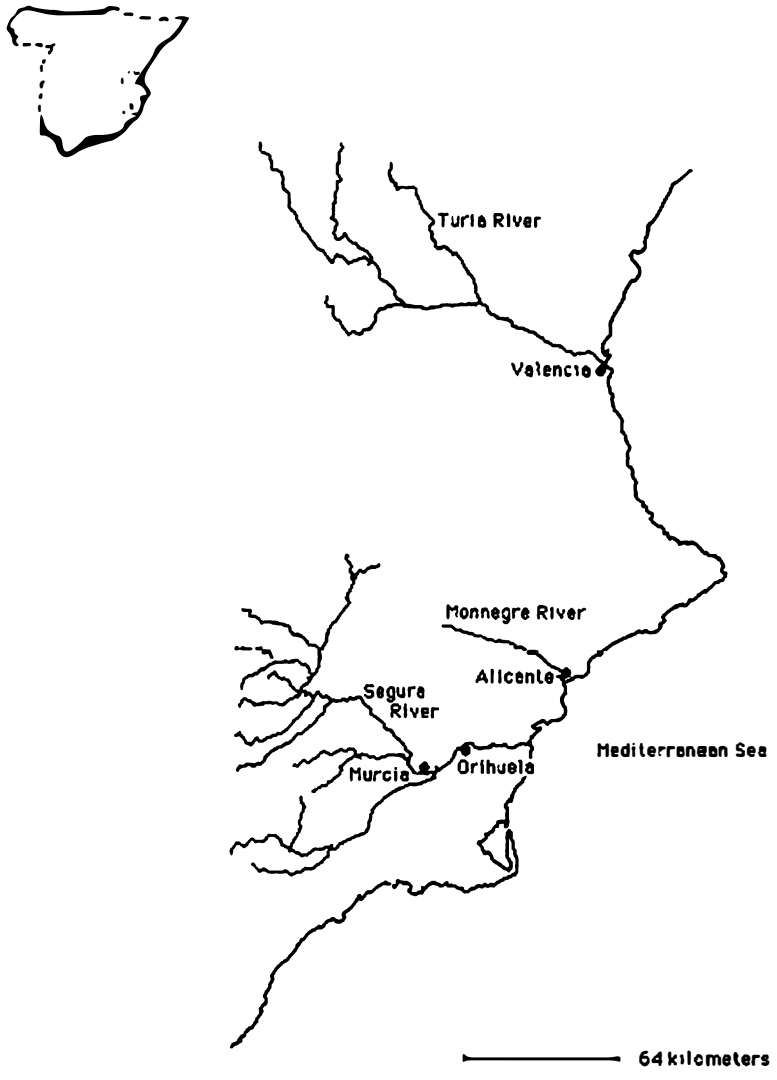


Figure 3.1. Location of Spanish *huertas.*

huertas (well-demarked irrigation areas surrounding or near towns) of Valencia, Murcia, Orihuela, and Alicante, and Glick (1970) has provided us with an authoritative study of the *huerta* of Valencia during the Middle Ages.

Valencia

Near the city of Valencia, the waters of the Turia River are divided into eight major canals serving the 16,000-hectare *huerta*. The farms in Valencia have always been small, but they have become extremely fragmented during the past century. Over 80% of the farms are less than 1 hectare, and few exceed 5 hectares (Maass and Anderson 1986, p. 11). Most winters are frost-free, and the summers are hot and sunny. Farmers are able to harvest two or three crops each year and concentrate largely on potatoes, onions, and a wide diversity of vegetable crops. Each farmer is free to select the cropping patterns he prefers.

Given the low rainfall in Valencia itself, the extensive agriculture of this region would not have been possible without effective use of the Turia River. The variation in the flow of the Turia River has historically been quite high. Years of low water flow have been followed by years of extensive flooding. Until the turn of this century, no dams had been constructed on the Turia River serving the Valencian *huerta*. It was not until 1951, when the Generalisimo Dam was completed, with 228 million cubic meters of storage, that substantial upstream storage was provided to regulate the extreme fluctuations in the river's flow. Some groundwater has been developed in the region to supplement the river's supply, but this has never been a major factor in the supply of irrigation water.

In Valencia, the right to water inheres in the land itself. Land that was watered before the time of the reconquest is specified as irrigated land (*regadiu*), and the remaining lands in these *huertas* are dry lands (*seca*).[11] Some land is entitled to water only in times of abundance (*extremales*). The basic allocation principle in Valencia is that each piece of *regadiu* land is entitled to a quantity of canal water proportionate to its size.

In Valencia, the irrigators from seven of the major canals are organized into autonomous irrigation communities whose syndic,[12] or chief executive, participates in two weekly tribunals. The Tribunal de las Aguas is a water court that has for centuries met on Thursday mornings outside the Apostles' Door of the Cathedral of Valencia. The many Islamic features of its traditions have led scholars to argue that the court evolved during the period of Islamic rule.[13] Its proceedings are carried on without lawyers, but with many onlookers. A presiding officer questions those who are involved in a dispute and others who may be able to provide additional information,

and the members of the court, excluding the syndic whose canal is involved, make an immediate decision regarding the facts of the case in light of the specific rules of the particular canal. Fines and damages are assessed consistent with the rules of the particular canal. The final decisions of the court are recorded, but not the proceedings. After the court session, the syndics may also convene a second tribunal, which serves as a coordinating committee encompassing all seven of the canals to determine when to institute operating procedures related to seasonal low waters or to discuss other intercanal problems.

The farmers (*hereters*) who own lands eligible to receive water from each of these seven canals meet every second or third year to elect the syndic and several other officials for their canal. Besides his role in the two tribunals, the syndic is the executive officer of the individual irrigation unit. His responsibilities include the basic enforcement of the regulations of his own unit. He has the power to make authoritative physical allocations of water when disputes arise in the day-to-day administration of the waterworks, to levy fines, and to determine the order and timing of water deliveries during times of severe shortages (subject to weekly review by the Tribunal de las Aguas). The syndic must own and farm land served by the canal. The syndic usually has a small staff of ditch-riders and guards whom he appoints to help him carry out these assignments.[14]

In medieval times, the *hereters* also elected two or more inspectors (*veedors*) who were representatives of the community of irrigators and were to consult with the syndic about the daily operation of the canal and assist in rendering physical judgments when conflicts between farmers or between a syndic and a farmer erupted. In modern times, the *hereters* elect an executive committee (*junta de gobierno*) to consult with the syndic until the next biannual meeting. The executive committee is composed of delegates from all of the canal's major service areas. Decisions about when to shut down the canals for annual maintenance and how the maintenance work will be organized are made by the members of this committee of irrigators.

The basic rules for allocating water are dependent on the decisions made by the officials of the irrigation community concerning three environmental conditions: abundance, seasonal low water, and extraordinary drought. In years of declared abundance – a relatively infrequent event – farmers are allowed to take as much water as they need whenever water is present in the canal serving their land.

The most frequent condition under which the canals operate is that of seasonal low water. When the low-water condition is in effect, water is distributed to specific farmers through a complex, rule-driven hydraulic

system. Each distributory canal is positioned in a rotation scheme in relation to the other distributory canals.[15] Each farm on a distributory canal receives water in a set rotation order, starting from the head of the canal and culminating in the tail end of the canal:

> On days when water is running in a lateral . . . those farmers who want to irrigate will take it in turn (*por turno*), generally in order from the head to the tail of the channel. Once a farmer opens his headgate, he takes all the water he needs, without any restriction of time; and he defines his own needs, principally in terms of the water requirements of the crops he has chosen to plant. The only limitation is that he may not waste water. If a farmer fails to open his headgate when the water arrives there, he misses his turn and must wait for the water to return to the farm on the next rotation. When a lateral operates in rotation and all users who want water at a given time cannot be served before the rotation passes to another lateral, distribution will begin, when water returns, at the point where it previously terminated. (Maass and Anderson 1986, p. 28)

The basic elements of the *turno* system are that (1) the order in which irrigators receive water is fixed, and (2) each farmer can decide how much water to take as long as water is not wasted. Consequently, no irrigator can tell exactly when his turn will come, because that depends on the volume of water in the canal and the quantity needed by those ahead of him. On the other hand, each irrigator knows that he can take as much water as he needs when his turn eventually comes.

In periods of extraordinary drought, these procedures are modified so that farms whose crops are in the most need of water are given priority over farms whose crops require less water. At the beginning of a drought period, the farmers themselves are expected to apply water only to those crops in most need to shorten their turns in order to allow other farmers in need to obtain the scarce water. As a drought period continues, the syndic and his representatives take more and more responsibility for determining how long each farmer may have water, in light of the condition of the farmer's crops and the needs of others. In recent years, procedures to be used in extraordinary drought have been needed less frequently than in earlier times, because of the increased regulatory capacity of the Generalisimo Dam. Even so, an established procedure is in place for switching rule regimes when environmental conditions change.

The level of monitoring that is used in the *huertas* is very high. In this environment of water scarcity and risk, many temptations occur to take water out of turn, or in some way obtain illegal water. As the time approaches for a farmer to take his turn at the water, he will tend his fields near to the canal so that he can be prepared to open his own gate when the water arrives; if not prepared, he misses his turn entirely and must wait for

the next round. While waiting, it is relatively easy to watch what those ahead of him are doing and watch the ditch-riders, whom he is paying. The ditch-riders patrol the canals regularly and are watched over by the syndic, who can lose respect, and his job, if the allocation of water is not handled fairly and according to the farmers' rules. Challenges to the actions of a syndic, a ditch-rider, or another irrigator can be aired weekly before the Tribunal de las Aguas, with many of the other farmers watching the confrontation with interest. The reciprocal monitoring relationships in Valencia are shown in Figure 3.2. Given that everyone is watching everyone else, there is considerable potential for violence among irrigators and between irrigators and their agents. In medieval times, the norms related to honor probably exacerbated the potential for conflict, and *hereters* "were willing to fight in an instant if they felt that their water supply was jeopardized in any way" (Glick 1970, p. 70). The actual violence never approached the potential.

The survival of the books in which fines were recorded for the years 1443 and 1486 for the nearly similarly structured *huerta* of Castellón, some of the detail of which has been reproduced by Glick, provides a picture of the types of infractions discovered, the high level of monitoring undertaken, and the low level of actual fines during an earlier era.[16] In 1443 there were 441 fines assessed; in 1486 there were 499 fines (Glick 1970, p. 54). The similarity in the distributions of offenses and the numbers of fines for these two periods more than 40 years apart testifies to the stability of the system. Guards assessed fines at a rate of more than one per day.[17] About two-thirds of the actions were initiated by the guards, and the remaining third by farmers. Forty-two percent concerned infractions that clearly were motivated by the temptation to obtain water not legally available to the farmer (taking forbidden water, stealing water, installing or undoing canal checks illegally, taking water by force, irrigating without

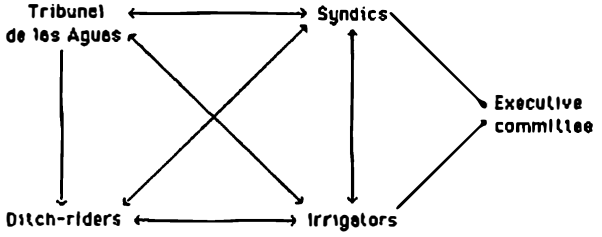


Figure 3.2. Patterns of monitoring and accountability among key actors in the Valencia *huerta*.

right). The remainder of the infractions related to actions that caused harm to others (flooding a road or a fallow field, wasting water) and were also forbidden by the community. Farmers were held publicly accountable for the errors they committed that caused harm to others. Two-thirds of those fined in a year were "one-time offenders" and were not mentioned again in the fine book. Of those who were repeaters, 41% were involved in two actions, 25% in three, 15% in four, 8% in five, and 12% in more than five (Glick 1970, p. 59).

Sufficient data exist to estimate the rate of conformance to the rules for Castellón. There were approximately 1,000 hearths in Castellón in the fifteenth century (T. F. Glick, personal communication). If the rotation system took about two weeks, each of the roughly 1,000 irrigators would have had about 25 opportunities during the year to take water illegally. Thus, approximately 25,000 opportunities for theft occurred, as contrasted to 200 recorded instances of illegal taking of water. That would give a recorded infraction rate of 0.008. One must assume that the guards did not detect all infractions. One could double, triple, or even quadruple the recorded rate, however, and still have a remarkable conformance rate.[18]

Although the conformance rate was high, about one-third of the *hereters* would have had one encounter with a guard at some time during a full year.[19] Consequently, information about the extensive monitoring was regularly conveyed to irrigators. We do not have as detailed a picture of the enforcement patterns in modern times, but both the number of ditch-riders employed and the necessity of holding a weekly court session lead one to suspect that high enforcement levels have been required to dampen the ever present temptation to steal water, as well as the potential for inter-farmer conflict and violence. The stability of this system has been achieved in spite of personal temptations to cheat and engage in violent behavior.

The books of fines also reveal that even though the syndic received two-thirds of the fine (the other third going to the accuser) and the authorized levels for fines were set high, the actual fines assessed "were very low (a few pennies at the most) and also variable, depending on the gravity of the offense, on general economic conditions, and probably on the individual's ability to pay" (Glick 1970, p. 56). Glick comments that this introduced some flexibility into the relatively rigid rotation systems. From time to time, the cost to a farmer of waiting for his next legal turn to receive water, as contrasted to stealing water available in the canal, would be extraordinarily high. Because the fines actually assessed were kept relatively low, the guards did not deeply antagonize the farmers, who generally adhered to the rules. A farmer would suffer some humiliation if detected cheating, but the monetary fine for cooperative farmers would be quite

low. Assessing harsh punishment to someone who usually follows the rules, but in one instance errs in the face of a desperate situation, can engender considerable antagonism and resentment (Oliver 1980).

Only rarely did farmers engage in ongoing harassment of one another. Glick notes one "particularly fractious individual" who made 5 accusations of theft and was himself similarly accused 13 times during 1486; 10 of the 18 incidents were conflicts between members of two families. But such cases are extremely rare in the archival data, and the absence of chronic conflict between farmers is considered by Glick to be "a tribute to both the efficiency of the distribution system and the vigilance of the guards" (Glick 1970, p. 64).[20]

Murcia and Orihuela

The Segura River runs from west to east as it approaches the Mediterranean, flowing first through the *huerta* of Murcia and then through the *huerta* of Orihuela. Of the 13,300 farms included within the service area of the *huerta* of Murcia, 83% are less than a single hectare. Of the 4,888 farms in the *huerta* of Orihuela, 64% are less than a single hectare, and 86% are less than 5 hectares. As in Valencia, water rights in Murcia and Orihuela are tied to the land. *Regadiu* and *seca* lands were designated long ago and have remained stable for centuries. The quantity of rainfall in the *huertas* of Murcia and Orihuela is, on the average, considerably less than in Valencia, and it occurs with greater variation. The terrain in Murcia and Orihuela is more varied than the terrain in Valencia, and local procedures involve much more emphasis on the problem of watering highlands and lowlands from the same canal.

Each farmer is assigned a *tanda*, a fixed time period during which he may withdraw water. Thus, each farmer knows exactly when and for how long he may obtain water, but he does not know exactly how much water may be available at that time. The *tanda* procedure has some advantages over the *turno* procedure used in Valencia. Each farmer can plan his activities with a greater degree of certainty as to when he will be able to irrigate. Each farmer is more motivated to economize on the use of water within his own fields because he must make the decision how to allocate a limited time-slice of water to his own fields. On the other hand, the *tanda* procedure is itself quite rigid, particularly as farms are bought and sold, divided or combined.

The officials of the irrigation community, in consultation with city officials, are responsible for declaring when there is insufficient water to continue the regular *tanda* procedure. When extraordinary low-water con-

ditions are in effect, the officials of each community post a new schedule for each rotation of the season – approximately every two weeks – indicating which crops will be given precedence and the schedule and special rules to be followed for the next rotation period.

There are about 30 irrigation communities in Murcia, 10 in Orihuela, and several more that take from canals just below Orihuela. In both *huertas*, the communities employ guards, who most frequently come from the canal sections where they are employed and are nominated by the farmers of that section.

> The guards patrol the canal and report any violations of the ordinances they observe; act as witnesses where one farmer charges another with a violation or themselves bring charges against farmers; and assist in the distribution of water, frequently opening and closing the principal canal checks and the turnout gate of the principal laterals. (Maass and Anderson 1986, p. 80)

The irrigation communities within both *huertas* have formed *huerta*-wide organizations. The syndics of the canal communities of Murcia meet yearly in a general assembly and elect members to an executive commission, in addition to approving an annual budget and taxes. The syndics of the canals in Orihuela meet in a general assembly of its *huerta*-wide organization every three years to elect a water magistrate, his lieutenant, and a solicitor. The water magistrate presides at all assemblies within Orihuela. The *huerta*-wide agency performs activities similar to those undertaken in Murcia. The city of Orihuela is hardly involved in irrigation activities within its limits.

Both *huertas* have established water courts in which farmers can bring charges against each other or in which officials can charge a farmer with an offense. Murcia's water court – which has the felicitous name of the Council of Good Men (Consejo de Hombres Buenos) – is composed of five canal syndics and two inspectors. Because Murcia has 30 organized communities, the names of all syndics and inspectors for all systems are placed in two bowls at the beginning of each year, and each month a new court is selected by lottery so that each canal will be represented in an equitable fashion. The Murcian court meets every Thursday morning in the City Hall and is presided over by the mayor of Murcia (or his deputy), who votes only in case of a tie. Not only is the day of meeting similar to that in Valencia, but the general procedures are the same: "oral, public, summary, and cheap" (Maass and Anderson 1986, p. 82).

The water court in Orihuela has only a single judge, and its procedures differ substantially from those in Murcia and Valencia. Those who wish to bring charges against others do so to an officer of the court. The person accused is then summoned to appear before the magistrate within a few

days. A sentence is imposed immediately if the person accused confesses to the charge. Otherwise, the magistrate tries to get those involved to come to an agreement that he can accept.

Alicante

Whereas the Segura and Turia rivers drain large watersheds, including mountain ranges where winter precipitation is stored in the form of snow and released later, the Monnegre River serving Alicante rises near the sea and drains only a small area. The even greater shortage of water in this *huerta*, as contrasted to Valencia and Murcia-Orihuela, which themselves do not have an abundant supply of water, has affected the strategies that the irrigators in Alicante have adopted. The basic water right in Alicante is closer to that of Murcia than to that of Valencia. All water rights are to a fixed time period. Originally, these time allocations were tied to land ownership. Shortly after Alicante was recovered from the Muslims, rights to withdraw water for fixed time periods were separated from ownership of land, and a market in these rights existed apart from the market for land. Alicante farmers took the initiative to construct the Tibi Dam in 1594, which at times has led to greater involvement of national and regional authorities in the management of irrigation in Alicante than in Valencia or Murcia-Orihuela. Local irrigators have sought out still other sources of water, and that has involved them in extensive contractual arrangements with large-scale private water companies.

The 3,700 hectares of *huerta* land are divided among 2,400 farms, 63% of which are less than 1 hectare, and 93% of which are less than 5 hectares (Maass and Anderson 1986, p. 101). Alicante farmers have adopted a mixed strategy of growing cereals and vegetables between rows of fruit and nut trees. Prior to the construction of the Tibi Dam, many owners of land sold their water rights to others or regularly rented their rights. Consequently, a fixed quantity of water rights existed prior to construction of the dam, and those rights were traded independently of land transactions. Tibi Dam made available twice as much usable irrigation water as the unregulated river had provided. The rights to the "new water" created by the Tibi Dam were assigned to owners of *huerta* land whose assessments paid for the dam.[21] The rights to the other half of the water supply – the "old water" – were held by those who had already acquired rights prior to construction of the dam. A new proviso was added to these rights that they could be sold or rented only to those who owned land eligible to receive new water. Consequently, the water rights could not be sold to individuals whose land lay outside the *huerta*. Although the rights to new water were

originally attached to the land, those rights were soon "rented" from time to time by farmers who did not need all of their water for a particular rotation.

Prior to a full rotation of water through the irrigation community's canals, a notice is posted by the syndicate providing information about the dates of the next rotation and the times during which "scrip" will be issued. Holders of both new and old water rights obtain scrip equivalent to their recorded water rights in denominations from one hour down to one-third of a minute. All scrip for Tibi water is fully exchangeable. Farmers who hold new-water rights, and thus land within the *huerta*, rarely have sufficient scrip to obtain enough water to irrigate their crops. They can purchase scrip in three ways: at an informal market among holders of rights conducted on Sunday morning before a formal auction is held; at the formal auction; and on market days, when farmers are congregating for trade.

In the formal auction, the irrigation community sells the approximately 90 hours of water that it owns – water rights assigned to it by the irrigators in 1926 to provide a regular income for the syndicate's operations.[12] The syndicate also sells any surplus scrip that was not claimed by right-holders during the previous allotted time period. The minimum quantity of water offered in the formal auction is a full hour, but the purchased scrip is fully divisible and negotiable. Considerable information is made available by the irrigation community to enable farmers to make intelligent choices.

> The ditch riders are present . . . and can tell a farmer when the water is likely to reach his property. The organization posts on a bulletin board outside the tavern a current report of water storage in the reservoir; a full account of all water delivered in the previous rotation, including the names of irrigators and the amounts of water delivered to each; and a full accounting of all water sold at auction in the previous rotation, including the names of all successful bidders, the number of hours each purchased, and the prices paid.
>
> (Maass and Anderson 1986, p. 116)

A farmer who wants to irrigate his land during a particular rotation tells his ditch-rider – who opens and closes all of the relevant control structures – how much time he wishes to use. The ditch-rider, in turn, informs the farmer approximately when the water will be available. The farmer is supposed to pay the ditch-rider when water is delivered, but the practice is to allow a farmer up to three days after a rotation has been completed. At that time, the ditch-rider's report of all water delivered and equivalent scrip must be turned in. Thus, farmers purchase scrip not only for future deliveries but also to cover fully what they have used during the current rotation. The price of water is consequently higher toward the end of a

rotation than at the beginning of one.[23] The price of water also varies in relation to the amount of water available. In years of abundant rainfall, farmers need less irrigation water, and the price of Tibi water falls. In times of extreme drought, there may not be any water to distribute, and the auction will not occur until water is available. In periods of seasonal low water, the price of water may become very high and can be a source of considerable conflict between holders of old rights versus holders of new rights. Alicante farmers may also purchase water from several other sources.[24]

The organization of the irrigation community in Alicante differs significantly from that for the *huertas* discussed earlier. First, there is only one irrigation community for the entire *huerta*. Second, to vote in the general assembly of the community, a farmer must own 1.8 hectares of land; to vote for the executive commission, 1.2 hectares of land; and to be eligible to serve on the commission, 3.6 hectares of land (Maass and Anderson 1986, p. 117). Whereas a farmer must own a minimum of land to participate, the votes of farmers owning more land are not weighted to reflect differences in the amounts of land owned. The executive commission is composed of 12 representatives (*sindicos*) who serve four years each (half rotating every second year). One member from this body is selected as the director.

The general assembly meets annually to approve the budget and taxes and to decide matters brought before it by the executive commission. Special meetings can be called when problems arise requiring action between the annual meetings. Both the executive commission and the assembly have been extremely active in Alicante in the repeated efforts to find new water and to attempt to develop better contractual arrangements with the private firms using the community's canal to sell water in the *huerta*. The regular expenses of the community are assessed against the holders of all water rights. Three rotations each year are designated as those during which regular taxes will be collected, and a right-holder must pay the assessment at the time of applying for scrip. Extraordinary expenses, which at times have been quite high, are also assessed in the same manner, except that the payment is due at a different set of rotations during the year.

The commission employs an executive secretary, as well as all those who operate the control structures and deliver the water directly to the farmer. The ditch-riders open and close all farm headgates in Alicante. The farmers do this themselves in Valencia and Murcia-Orihuela. The ditch-riders of Alicante, however, exercise less discretion in determining who shall receive water and when. One employee is given the responsibility of accounting for

the water that has left the regulating basins until it reaches a ditch-rider, who is again accountable for all of the water that is assigned to him.

National authorities have exerted more control over irrigation matters in Alicante than in the other *huertas*. A large structure, such as Tibi Dam, can be seized and used as a source of revenue and power by a rent-seeking ruler. Although Philip II did not attempt to exercise control over the Tibi Dam when it was built, the dam was transferred to royal ownership for a century in 1739. When control of the dam and responsibility for distributing its water were returned to Alicante in 1840, farmers did not win the right to select syndicate officers for another 25 years. The Spanish Civil War also interrupted the control that farmers exercised over the irrigation syndicate. It was not until 1950 that farmers again selected their own officials.

It should be noted that the degree of freedom to devise and change their own institutions, successfully asserted by the irrigators of eastern Spain, was not typical of the Castilian part of Spain, whose far more centralized institutions were the major influences on the evolution of Spanish national institutions.[25] By the end of the thirteenth century, the *cortes* of the kingdom of Aragon (roughly comprising Valencia, Aragon, and Catalonia) had "already secured the power to legislate and even to limit the king's power to issue legislation under certain conditions" (Veliz 1980, p. 34). The *cortes* in Castile, at the same time in history, was seldom summoned. By the time the centralized monarchy based on the Castilian model came to dominate Spain and Latin America, the autonomy of the *huertas* was well established. The continuing willingness of the irrigators in these regions to stand up for their rights attests that they had greater autonomy than did those in other parts of Spain. One can only wonder if the course of history in Latin America might have differed substantially if the Spanish monarchy established by Ferdinand and Isabella had been modeled on Aragon and not on Castile.[26]

Maass and Anderson have conducted an interesting evaluation of the comparative efficiencies of the Spanish *huertas* and several systems operating in the western part of the United States. Without including the costs of water or the administrative costs associated with governing and managing the canals, they find that the system that has evolved in Alicante enables farmers to be most efficient in using other input factors. The system devised in Valencia is the least efficient of the Spanish systems, with the Murcia-Orihuela systems coming in between. All of the systems generate positive benefits for the farmers they serve, and all have shown an amazing capacity to survive. In 1887, the Murcian historian Diaz Cassou concluded that "the democratic and representative character of the agricultural commune of Murcia had shown a remarkable stability, for a succession of very different

national political epochs had offered no serious obstacles to its continued function" (Maass and Anderson 1986, p. 83). A century later, Cassou's reflection remains valid.

ZANJERA IRRIGATION COMMUNITIES IN THE PHILIPPINES

The earliest recorded reference to the existing irrigation societies in the Ilocanos area of Ilocos Norte in the Philippines derives from Spanish priests writing in 1630 (H. Lewis 1980, p. 153). No serious effort has been made to determine if similar organizations were in existence before the Spanish colonial period, but it would not be unreasonable to assume that the modern *zanjeras* are derived from a mixture of traditions, including that of the Spanish. The most striking similarity between the *huerta* and *zanjera* systems is in the central role given to small-scale communities of irrigators who determine their own rules, choose their own officials, guard their own systems, and maintain their own canals. The internal organization of each *zanjera* has been tailored to its own history, and thus the specific rules in use vary substantially (Keesing 1962). In 1979 there were 686 communal irrigation systems in Ilocos Norte (Siy 1982, p. 25).[27]

Zanjeras have been established both by landowning farmers wanting to construct common irrigation works and by individuals organizing themselves so as to acquire land. The technologies used in *zanjera* systems are relatively crude and labor-intensive. The large number of operating systems and the amount of labor put into these by farmers – tenants as well as landowners – have meant that technological knowledge of how to construct dams and other works has been widely shared. With this knowledge, it has been possible for enterprising tenant farmers to band together to construct an irrigation system on previously nonirrigated land in return for the right to the produce from a defined portion of the newly irrigated land.

This type of contract – called a *biang ti daga* or a "sharing of the land" – allows the landowner to retain ownership. Use rights are extended to the *zanjera* dependent on continued maintenance of the irrigation system. At the time of forming an association, each original participant in the *zanjera* is issued one membership share or *atar*. The total number of *atars* is set at that point.[28] The share gives each member one vote and the right to farm a proportionate share of the land acquired by the *zanjera*, and it defines the obligation of the member for labor and material inputs. Each *atar*-holder is obligated to contribute one day's work during each work season declared by the *zanjera*, plus a share of the material required at construction time. The system was thus developed as a mode of acquiring long-term use rights

to land and the water to irrigate it without prior accumulation of monetary assets.

Each *zanjera* is laid out differently, but all that were set up by a *biang ti daga* contract share an underlying pattern. The area is divided into three or more large sections. Each farmer is assigned a plot in each section. All members are thus in fundamentally symmetrical positions in relation to one another. Not only do they own rights to farm equal amounts of land, but they all farm some land in the most advantageous location near the head of the system, and some near the tail. In years when rainfall is not sufficient to irrigate all of the fields, a decision about sharing the burden of scarcity can be made rapidly and equitably by simply deciding not to irrigate the bottom section of land.

Several parcels are set aside for communal purposes. A few parcels, located at the tail end of the system, are assigned to officials of the association as payment for their services. This system not only provides a positive reward for services rendered but also enhances the incentives for those in leadership positions to try to get water to the tail end of the system. Other lands are retained to secure income for the *zanjera* itself. See the work of Coward (1979, 1985) for a detailed description of this system.

The members of each *zanjera* elect a *maestro* as their executive officer, a secretary, a treasurer, and a cook.[29] In the larger associations, they also select foremen and team leaders to supervise the construction activities. The *maestro* has the challenging job of motivating individuals to contribute many hours of physically exhausting labor in times of emergency, when control structures have been washed out, and for routine maintenance. Given the backbreaking efforts required during the monsoon season or during extremely hot weather, this motivational task is of substantial proportions. The *maestro* is, of course, not dependent simply on his persuasive powers. Many real inducements and sanctions are built into these systems by the rules that *zanjera* members have constructed for themselves.

To illustrate the task involved in governing these systems, we shall consider one of these systems – actually, a federation of nine *zanjeras* – in more detail, based on the work of Robert Siy (1982). The Bacarra-Vintar federation of *zanjeras* constructs and maintains a 100-meter-long brush dam that spans the Bacarra-Vintar River, located on the northwestern tip of Luzon Island approximately 500 kilometers north of Manila. The unpredictable and destructive Bacarra-Vintar River drains the northeastern parts of the provinces. During the rainy season each year, the river destroys the federation's dam, which is constructed of bamboo poles, banana leaves, sand, and rock. During some years the dam will be destroyed three or four times.

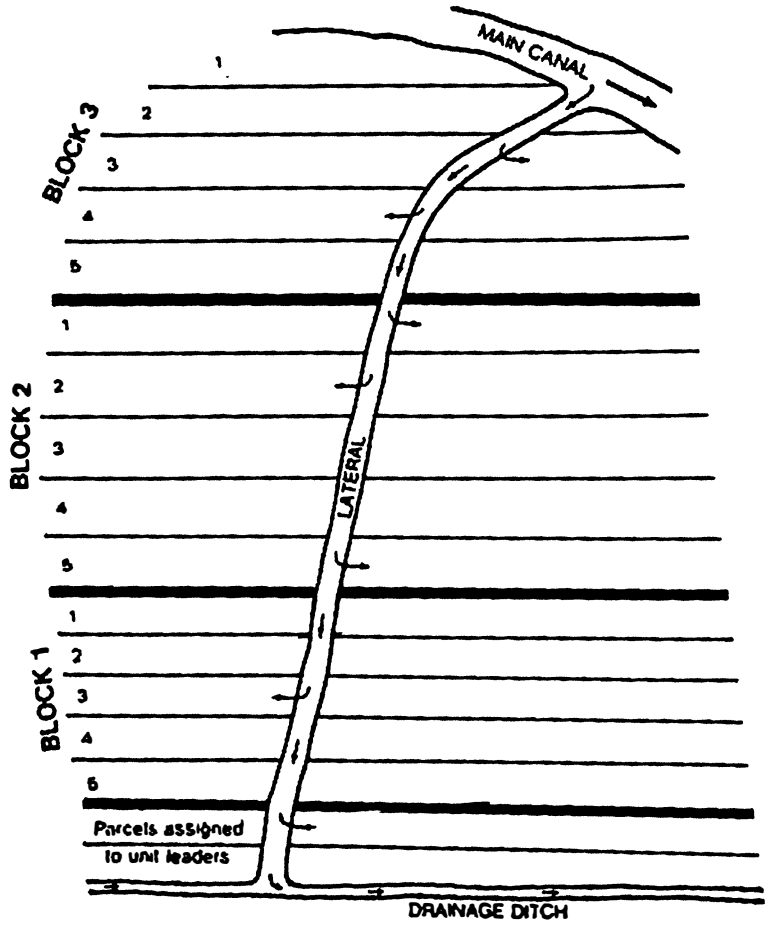


Figure 3.3. Typical layout of *zanjera* fields. (Adapted from Coward 1979.)

The histories of the nine component *zanjeras*, like that of the federation itself, have not been well preserved. What is known is that most of them were established independently and tried to construct and maintain their own diversion works from the river. The river has changed course several times in its history, and at various times some of the *zanjeras* have been cut off from their source of water by such changes. Two of the *zanjeras* were already associated during the nineteenth century and jointly constructed

one dam and canal. A formal agreement dated 1906 was written when a third *zanjera* joined their federation. Other existing systems joined slowly through the 1950s. The last two *zanjeras* entered at the time of their formation (Siy 1982, pp. 67-8).

In 1978 the federation formally incorporated as a private corporation in response to the 1976 Philippine Water Code, which defined only individuals or "juridical persons" as eligible to obtain water rights. Given the history of litigation in the area (M. Cruz, Cornista, and Dayan 1987), members of the federation wanted secure water rights in the name of the federation itself, rather than in the name of individual *zanjeras*. The heads of all the component *zanjeras* form the board of directors, with the *maestro* of the Surgui *zanjera* – one of the founding *zanjeras* – named as the president and chairman of the board. In 1980 there were 431 individuals who owned shares, or parts of shares, in at least one *zanjera*. Many members were involved in more than one of the *zanjeras*. The smallest component *zanjera* had 20 members, and the largest had 73 members (Siy 1982, p. 85). Each *zanjera* is responsible for its own financial and internal affairs and owes no financial obligations to the federation.

The board of directors determines when the dam should be rebuilt or repaired. Rebuilding takes about a week – somewhat more when the weather is unfavorable – and involves several hundred persons. Each *zanjera* is responsible for bringing construction materials and providing work teams (and the cooks and food to feed them). After spending a day preparing banana and bamboo mats, work teams in heavy boats confront the swirling waters to begin pounding in the poles that form the foundation for the dam. Then the mats are woven around the poles and reinforced with sand and rock.

Each of the five *zanjeras* with the largest numbers of *atars* provides one work team. The four smaller associations form two work teams. As the dam is laid out, it is divided, by the use of a "flexible" rod, into seven sections that are roughly proportional to the sizes of the work teams and the difficulty of the terrain. This work assignment pattern allows each group to monitor the progress of other groups and engenders some spirited competition among them. The work of maintaining the main canal is also assigned in a similar manner. Work on distributory canals is organized by each *zanjera*, which has divided itself into smaller work teams called *gunglos*, composed of 5 to 10 members.

Siy computed the total obligations (including work as well as attendance at meetings and celebrations) of *zanjera* members to their own associations and to the federation for 1980. The owner of a full *atar* share of the Santo Rosario *zanjera* was obligated to contribute 86 days during 1980 (the

largest obligation), whereas an owner of a full share in the Nibinib *zanjera* was obligated to contribute 32 days (the lowest). The average across the federation was 53 days (Siy 1982, p. 92). Given that some *atars* are held jointly by several farmers, the average number of days per working member is somewhat less – around 39 days for the year.

In terms of the contemporary schedule of 5 days per week, this amounts to two months of work supplied without direct monetary payment.[30] About 16,000 man-days were supplied by members to their own *zanjera* or federation during the year.[31] As Siy reflects, "there are definitely few rural organizations in the developing world which have been able to regularly mobilize voluntary [*sic*] labor to such extent" (Siy 1982, p. 95).[32] Given the rigorous and at times dangerous nature of the work, the level of attendance at these obligatory sessions is rather amazing.[33] On average, members were absent somewhat over 2 days out of their required 39, making the attendance rate about 94%. Fines assessed for nonattendance were fully paid in five of the *zanjeras*, and only one of the *zanjeras* had a substantial problem with the payment of fines (Siy 1982, p. 98).[34]

Over time, *zanjeras* face the problem of increased fragmentation of the original shares. A founding member with three sons, for example, may bequeath his plots to be distributed evenly among his sons, each of whom then assumes one-third of the obligations that their father had to fulfill (and having access to only one-third of the land). The individual *zanjeras* have responded to fragmentation in several ways. Some *zanjeras* appoint one person to be responsible for the fulfillment of *atar* responsibilities so that the associations do not have to monitor intra-*atar* work contributions or shirking. Some of the *zanjeras* now require prior approval before a share is sold or tenants are allowed to work *zanjera* land.

> Prospective members are "screened," and made to understand the full extent of their obligations to the zanjera before the transaction or tenancy agreement is approved. In a few cases, new members have been required to sign an agreement affirming their recognition of the zanjera's by-laws. These by-laws usually stipulate that erring members may be suspended or expelled from the zanjera, and their lands, confiscated. (Siy 1982, p. 101)

Given the great numbers of the landless population in the area, there is still fierce competition to gain access to land.

Water-allocation rules are not quite as restrictive in these systems as are work-contribution rules. In general, the supply of water to the irrigation system is more than adequate to meet the needs of the farmers, given the current cropping patterns and soil types involved. When water is abundant, water flows throughout the entire system, and anyone can irrigate at will. When water is scarce, rotation systems are established among the *zanjeras*,

and within *zanjeras* among the various distributory canals. During extremely dry periods, downstream *zanjeras* are allowed the full flow of the system for several nights in a row. After notification and agreement, the downstream *zanjera* sends its *gunglos* upstream to set up checks and close turnouts. "Other members 'stand guard' to ensure that such temporary control devices remain in place. Other groups attend to the actual delivery of water to individual parcels" (Siy 1982, p. 122). Precedence is given to parcels with the greatest need, and then a regular rotation system is established.

Several of the downstream *zanjeras* harvest only one crop per year, but two crops are possible in the higher *zanjeras*. Siy presents clear evidence that it would be possible to reallocate water among the nine *zanjeras* so as to increase the productivity of the lower *zanjera* lands without a loss in productivity by the head-end *zanjeras* (Siy 1982, pp. 122–45). On the other hand, the distribution of water is roughly proportional to the contributions of labor and materials and to *atar* shares. Thus, the three *zanjeras* that contribute most of the labor and materials (48%) receive 55% of the water, the three *zanjeras* that contribute 30% of the labor and materials receive 25% of the water, and the three *zanjeras* that contribute 22% of the labor and materials receive 20% of the water.[35]

From the perspective of technical efficiency, the system is not as efficient in its water-allocation scheme as it could be. Siy is, however, extremely careful to point out that many costs besides those of output forgone are involved in designing and running such systems:

> The costs may be in the form of the time and energy expended in deciding on an acceptable arrangement or in adjusting to an externally-imposed procedure. . . . For example, a shift in the distribution of water may necessitate a shift in the distribution of obligations among zanjeras. A zanjera that ends up receiving more water may then be required to contribute a larger proportion of labor and materials for system maintenance in order to satisfy the demands for sharing obligations in proportion to the increased benefits received. However, there is always the danger that the individual zanjera involved may not possess the immediate capability to meet such requirements, and, as such, these new demands on their resources may actually undermine the stability or solidarity of the whole organization. (Siy 1982, p. 146)

The major criterion used by irrigation engineers to evaluate the performance of an irrigation system is whether or not a system is technically efficient in the sense that water is allocated optimally to enhance crop production. The federation falls short in regard to this criterion, but it performs well in regard to mobilization of personnel for construction and maintenance activities. The members of the federation perceive the alloca-

tion of water to conform to legitimate formulas that they have themselves devised, rather than to formulas devised by external experts. As we shall see in Chapter 5, when external experts, working without the participation of the irrigators, have designed systems with the primary aim of achieving technical efficiency, they frequently have failed to achieve either the hoped-for technical efficiency or the level of organized action required to allocate water in a regular fashion or to maintain the physical system itself.

Because many members of the lower *zanjeras* also participate in other *zanjeras*, many own lands that receive adequate or more than adequate quantities of water, thus offsetting those lands that are left dry part of the year. In a survey of *zanjera* members, respondents from the lower *zanjeras* were more likely than members of upstream *zanjeras* to report a lack of water during part of the year. But when asked what major irrigation problems they faced, none "had anything to say about the way water was allocated or about the fairness of water distribution" (Siy 1982, p. 141). The problem cited by 65% of the irrigators surveyed was the hardship associated with the annual damage to their dam.

SIMILARITIES AMONG ENDURING, SELF-GOVERNING CPR INSTITUTIONS

Despite all of the differences among the CPR settings described in this chapter – and substantial differences exist – all share fundamental similarities. One similarity is that all face uncertain and complex environments. In the mountain commons, the location and timing of rainfall cannot be predicted. In the irrigation systems, erratic rainfall is again a major source of uncertainty. Whereas the construction of physical works tends to reduce the level of uncertainty, it tends to increase the level of complexity in these systems. Irrigators must have practical engineering skills as well as farming skills.

In contrast to the uncertainty caused by these environments, the populations in these locations have remained stable over long periods of time. Individuals have shared a past and expect to share a future. It is important for individuals to maintain their reputations as reliable members of the community. These individuals live side by side and farm the same plots year after year. They expect their children and their grandchildren to inherit their land. In other words, their discount rates are low. If costly investments in provision are made at one point in time, the proprietors – or their families – are likely to reap the benefits.

Extensive norms have evolved in all of these settings that narrowly define "proper" behavior. Many of these norms make it feasible for in-

dividuals to live in close interdependence on many fronts without excessive conflict. Further, a reputation for keeping promises, honest dealings, and reliability in one arena is a valuable asset. Prudent, long-term self-interest reinforces the acceptance of the norms of proper behavior. None of these situations involves participants who vary greatly in regard to ownership of assets, skills, knowledge, ethnicity, race, or other variables that could strongly divide a group of individuals (R. Johnson and Libecap 1982).

The most notable similarity of all, of course, is the sheer perseverance manifested in these resource systems and institutions. The resource systems clearly meet the criterion of sustainability. The institutions meet Shepsle's (1989b) criterion of institutional robustness, in that the rules have been devised and modified over time according to a set of collective-choice and constitutional-choice rules. These cases were specifically selected because they have endured while others have failed. Now the task is to begin to explain their sustainability and robustness, given how difficult it must have been to achieve this record in such complex, uncertain, and interdependent environments in which individuals have continuously faced substantial incentives to behave opportunistically.

The specific operational rules in these cases differ markedly from one another. Thus, they cannot be the basis for an explanation across settings. In the Japanese mountain commons, for example, appropriation rights and provision duties are assigned to established family units in a village instead of to individuals. In the Swiss mountains, appropriation rights and provision duties are inherited by individual males who own private property in the village and remain citizens of the village. In eastern Spain, a farmer's right to irrigation water is based on the parcel of land inherited, purchased, or leased, not on a relationship to a village. In the Philippines, a complex contract among long-term usufructuary right-holders determines rights and provision duties. The rules defining when, where, and how an individual's allotted resource units can be harvested or how many labor days are required also vary considerably across cases.

Although the particular rules that are used within these various settings cannot provide the basis for an explanation of the institutional robustness and sustainability across these CPRs, part of the explanation that I offer is based on the fact that the particular rules differ. The differences in the particular rules take into account specific attributes of the related physical systems, cultural views of the world, and economic and political relationships that exist in the setting. Without different rules, appropriators could not take advantage of the positive features of a local CPR or avoid potential pitfalls that might be encountered in one setting but not others.

Instead of turning to the specific rules, I turn to a set of seven design

principles that characterize all of these robust CPR institutions, plus an eighth principle used in the larger, more complex cases. These are listed in Table 3.1. By "design principle" I mean an essential element or condition that helps to account for the success of these institutions in sustaining the CPRs and gaining the compliance of generation after generation of appropriators to the rules in use. This list of design principles is still quite speculative. I am not yet willing to argue that these design principles are necessary conditions for achieving institutional robustness in CPR settings. Further theoretical and empirical work is needed before a strong assertion of necessity can be made. I am willing to speculate, however, that after

Table 3.1. *Design principles illustrated by long-enduring CPR institutions*

1. Clearly defined boundaries
   Individuals or households who have rights to withdraw resource units from the CPR must be clearly defined, as must the boundaries of the CPR itself.
2. Congruence between appropriation and provision rules and local conditions
   Appropriation rules restricting time, place, technology, and/or quantity of resource units are related to local conditions and to provision rules requiring labor, material, and/or money.
3. Collective-choice arrangements
   Most individuals affected by the operational rules can participate in modifying the operational rules.
4. Monitoring
   Monitors, who actively audit CPR conditions and appropriator behavior, are accountable to the appropriators or are the appropriators.
5. Graduated sanctions
   Appropriators who violate operational rules are likely to be assessed graduated sanctions (depending on the seriousness and context of the offense) by other appropriators, by officials accountable to these appropriators, or by both.
6. Conflict-resolution mechanisms
   Appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials.
7. Minimal recognition of rights to organize
   The rights of appropriators to devise their own institutions are not challenged by external governmental authorities.

For CPRs that are parts of larger systems:

8. Nested enterprises
   Appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises.

further scholarly work is completed, it will be possible to identify a set of necessary design principles and that such a set will contain the core of what has been identified here.[16]

For these design principles to constitute a credible explanation for the persistence of these CPRs and their related institutions, I need to show that they can affect incentives in such a way that appropriators will be willing to commit themselves to conform to operational rules devised in such systems, to monitor each other's conformance, and to replicate the CPR institutions across generational boundaries. I shall discuss each of the design principles in turn.

Clearly defined boundaries

1 Individuals or households who have rights to withdraw resource units from the CPR must be clearly defined, as must the boundaries of the CPR itself.

Defining the boundaries of the CPR and specifying those authorized to use it can be thought of as a first step in organizing for collective action. So long as the boundaries of the resource and/or the specification of individuals who can use the resource remain uncertain, no one knows what is being managed or for whom. Without defining the boundaries of the CPR and closing it to "outsiders," local appropriators face the risk that any benefits they produce by their efforts will be reaped by others who have not contributed to those efforts. At the least, those who invest in the CPR may not receive as high a return as they expected. At the worst, the actions of others could destroy the resource itself. Thus, for any appropriators to have a minimal interest in coordinating patterns of appropriation and provision, some set of appropriators must be able to exclude others from access and appropriation rights. If there are substantial numbers of potential appropriators and the demand for the resource units is high, the destructive potential should all be allowed to freely withdraw units from the CPR could push the discount rate used by appropriators toward 100%. The higher the discount rate, the closer the situation is to that of a one-shot dilemma in which the dominant strategy of all participants is to overuse the CPR.

Since the work of Ciriacy-Wantrup and Bishop (1975), the presence of boundaries concerning who is allowed to appropriate from the CPR has been used as the single defining characteristic of "common-property" institutions as contrasted to "open-access" institutions. The impression is sometimes given that this is all that is necessary to achieve successful

regulation. Making this attribute one of seven, rather than a unique attribute, puts its importance in a more realistic perspective. Simply closing the boundaries is not enough. It is still possible for a limited number of appropriators to increase the quantity of resource units they harvest so that they either dissipate all potential rents or totally destroy the resource (Clark 1980). Consequently, in addition to closing the boundaries, some rules limiting appropriation and/or mandating provision are needed.

Congruence between appropriation and provision rules and local conditions

2 Appropriation rules restricting time, place, technology, and/or quantity of resource units are related to local conditions and to provision rules requiring labor, materials, and/or money.

Adding well-tailored appropriation and provision rules helps to account for the perseverance of these CPRs. In all these cases, the rules reflect the specific attributes of the particular resource. Among the four Spanish *huertas* that are located in fairly close proximity to one another, the specific rules for the various *huertas* differ rather substantially. It is only in the one system (Alicante) where there has been substantial storage available since the construction of Tibi Dam in 1594 that a water auction is held. At the time of the Sunday morning auction, substantial information about the level of water in the dam is made available to the Alicante irrigators. Consequently, they can know about how much water they will receive if they purchase an hour of water. In the systems without storage, water is strictly tied to the land, and some form of rotation is used. In Valencia, each farmer takes as much water as he can put to beneficial use in a defined order. Thus, each farmer has a high degree of certainty about the quantity of water to be received, and less certainty about the exact timing. In Murcia and Orihuela, where water is even more scarce, a tighter rotation system is used that rations the amount of time that irrigators can keep their gates open. Further, the rules attempt to solve the problem of getting water to a more diversified terrain than in Valencia. Subtly different rules are used in each system for assessing water fees used to pay for water guards and for maintenance activities, but in all instances those who receive the highest proportion of the water also pay the highest proportion of the fees. No single set of rules defined for all irrigation systems in the region could deal with the particular problems in managing each of these broadly similar, but distinctly different, systems.[37]

3 Most individuals affected by the operational rules can participate in modifying the operational rules.

CPR institutions that use this principle are better able to tailor their rules to local circumstances, because the individuals who directly interact with one another and with the physical world can modify the rules over time so as to better fit them to the specific characteristics of their setting. Appropriators who design CPR institutions that are characterized by these first three principles – clearly defined boundaries, good-fitting rules, and appropriator participation in collective choice – should be able to devise a good set of rules if they keep the costs of changing the rules relatively low.

The presence of good rules, however, does not ensure that appropriators will follow them. Nor is the fact that appropriators themselves designed and initially agreed to the operational rules in our case studies an adequate explanation for centuries of compliance by individuals who were not involved in the initial agreement. It is not even an adequate explanation for the continued commitment of those who were part of the initial agreement. Agreeing to follow rules ex ante is an easy commitment to make. Actually following rules ex post, when strong temptations arise, is the significant accomplishment.

The problem of gaining compliance to the rules – no matter what their origin – often is assumed away by analysts positing all-knowing and all-powerful *external* authorities who enforce agreements. In the cases described here, no external authority has had sufficient presence to play any role in the day-to-day enforcement of the rules in use.[38] Thus, external enforcement cannot be used to explain these high levels of compliance.

Some recent theoretical models of repeated situations do predict that individuals will adopt contingent strategies to generate optimal equilibria without external enforcement, but with very specific information requirements rarely found in field settings (Axelrod 1981, 1984; Kreps et al. 1982; T. Lewis and Cowens 1983). In these models, participants adopt resolute strategies to cooperate so long as everyone else cooperates. If anyone deviates, the models posit that all others will deviate immediately and forever. Information about everyone's strategies in a previous round is assumed to be freely available. No monitoring activities are included in these models, because information is presumed to be already available.

It is obvious from our case studies, however, that even in repeated settings where reputation is important and where individuals share the norm of keeping agreements, reputation and shared norms are insufficient

by themselves to produce stable cooperative behavior over the long run. If they had been sufficient, appropriators could have avoided investing resources in monitoring and sanctioning activities. In all of the long-enduring cases, however, active investments in monitoring and sanctioning activities are quite apparent. That leads us to consider the fourth and fifth design principles:

Monitoring

4 Monitors, who actively audit CPR conditions and appropriator behavior, are accountable to the appropriators or are the appropriators.

Graduated sanctions

5 Appropriators who violate operational rules are likely to be assessed graduated sanctions (depending on the seriousness and context of the offense) by other appropriators, by officials accountable to these appropriators, or by both.

Now we are at the crux of the problem – and with surprising results. In these robust institutions, monitoring and sanctioning are undertaken not by external authorities but by the participants themselves. The initial sanctions used in these systems are also surprisingly low. Even though it is frequently presumed that participants will not spend the time and effort to monitor and sanction each other's performances, substantial evidence has been presented that they do both in these settings. The appropriators in these CPRs somehow have overcome the presumed problem of the second-order dilemma.

To explain the investment in monitoring and sanctioning activities that occurs in these robust, self-governing CPR institutions, the term "quasi-voluntary compliance" can be useful, as applied by Margaret Levi (1988a, ch. 3) to describe the behavior of taxpayers in systems in which most taxpayers comply. Paying taxes is voluntary in the sense that individuals choose to comply in many situations in which they are not being directly coerced. On the other hand, it is "*quasi*-voluntary because the noncompliant are subject to coercion – if they are caught" (Levi 1988a, p. 52). Taxpayers, according to Levi, will adopt a strategy of quasi-voluntary compliance when they have

> confidence that (1) rulers will keep their bargains and (2) the other constituents will keep theirs. Taxpayers are strategic actors who will cooperate only when they can expect others to cooperate as well. The compliance of each depends on the compliance of the others. No one perfers to be a "sucker." (Levi 1988a, p. 53)

Levi stresses the *contingent* nature of a commitment to comply with rules that is possible in a repeated setting. Strategic actors are willing to comply with a set of rules, Levi argues, when (1) they perceive that the collective objective is achieved, and (2) they perceive that others also comply. Levi is not the first to point to contingent behavior as a source of stable, long-term cooperative solutions. Prior work, however, had viewed contingent behavior as an *alternative* to coercion; see, for example, Axelrod (1981, 1984) and T. Lewis and Cowens (1983). Levi, on the other hand, views coercion as an *essential condition* to achieve quasi-voluntary compliance as a form of contingent behavior. In her theory, enforcement increases the confidence of individuals that they are not suckers. As long as they are confident that others are cooperating and the ruler provides joint benefits, they comply willingly to tax laws. In Levi's theory, enforcement is normally provided by an external ruler, although her theory does not preclude other enforcers.

To explain commitment in these cases, we cannot posit external enforcement. CPR appropriators create their own internal enforcement to (1) deter those who are tempted to break rules and thereby (2) assure quasi-voluntary compliers that others also comply.[39] As discussed in Chapter 2, however, the normal presumption has been that participants themselves will not undertake mutual monitoring and enforcement because such actions involve relatively high personal costs and produce public goods available to everyone. As Elster (1989, p. 41) states, "punishment almost invariably is costly to the punisher, while the benefits from punishment are diffusely distributed over the members." Given the evidence that individuals monitor, then the relative costs and benefits must have a different configuration than that posited in prior work. Either the costs of monitoring are lower or the benefits to an individual are higher, or both.

The costs of monitoring are low in many long-enduring CPRs as a result of the rules in use. Irrigation rotation systems, for example, usually place the two actors most concerned with cheating in direct contact with one another. The irrigator who nears the end of a rotation turn would like to extend the time of his turn (and thus the amount of water obtained). The next irrigator in the rotation system waits nearby for him to finish, and would even like to start early. The presence of the first irrigator deters the second from an early start, the presence of the second irrigator deters the first from a late ending. Neither has to invest additional resources in monitoring activities. Monitoring is a by-product of their own strong motivations to use their water rotation turns to the fullest extent. The fishing-site rotation system used in Alanya has the same characteristic that

cheaters can be observed at low cost by those who most want to deter cheaters at that particular time and location.

Many of the ways that work teams are organized in the Swiss and Japanese mountain commons also have the result that monitoring is a natural by-product of using the commons. Institutional analysis that simply posits an external, zero-cost enforcer has not addressed the possibility that the rules devised by appropriators may themselves have a major effect on the costs, and therefore the efficiency, of monitoring by internal or external enforcers.

Similarly, it is apparent that personal rewards for doing a good job are given to appropriators who monitor. The individual who finds a rule-infractor gains status and prestige for being a good protector of the commons. The infractor loses status and prestige. Private benefits are allocated to those who monitor. When internal monitoring is accomplished as part of a specialized position accountable to the other appropriators, several mechanisms increase the rewards for doing a good job or exposing slackards to the risk of losing their positions. In the Spanish *huertas*, a portion of the fines is kept by the guards; the Japanese detectives also keep the saké they collect from infractors.[40] All of the formal guard positions are accountable to the appropriators, and thus the monitors can be fired easily if discovered slacking off. Because the appropriators tend to continue monitoring the guards, as well as each other, some redundancy is built into the monitoring and sanctioning system. Failure to deter rule-breaking by one mechanism does not trigger a cascading process of rule infractions, because other mechanisms are in place.

Consequently, the costs and benefits of monitoring a set of rules are not independent of the particular set of rules adopted. Nor are they uniform in all CPR settings. When appropriators design at least some of their own rules (design principle 3), they can learn from experience to craft enforceable rather than unenforceable rules. This means paying attention to the costs of monitoring and enforcing, as well as the benefits that accrue to those who monitor and enforce the rules.

In repeated settings in which appropriators face incomplete information, appropriators who undertake monitoring activities obtain valuable information for themselves that can improve the quality of the strategic decision they make. In most theoretical models, where contingent strategies are shown to lead to optimal and stable dynamic equilibria, actors are assumed to have complete information about past history. They know what others did in the last round of decisions and how those choices affected outcomes. No consideration is given to how this information is generated. In the

settings we have examined in this chapter, however, obtaining information about behavior and outcomes is costly.

If the appropriators adopt contingent strategies – each agreeing to follow a set of rules, so long as most of the others follow the rules – each one needs to be sure that others comply and that their compliance produces the expected benefit. Thus, a previously unrecognized "private" benefit of monitoring in settings in which information is costly is that one obtains the information necessary to adopt a contingent strategy. If an appropriator who monitors finds someone who has violated a rule, the benefits of that discovery are shared by all who use the CPR, and the discoverer gains an indication of compliance rates. If the monitor does *not* find a violator, previously it has been presumed that private costs are involved without any benefit to the individual or the group. If information is not freely available about compliance rates, then an individual who monitors obtains valuable information from monitoring. The appropriator-monitor who watches how water is distributed to other appropriators not only provides a public good for all but also obtains information needed to make future strategic decisions.

By monitoring the behavior of others, the appropriator-monitor learns about the level of quasi-voluntary compliance in the CPR. If no one is discovered breaking the rules, the appropriator-monitor learns that others comply and that no one is being taken for a sucker. It is then safe for the appropriator-monitor to continue to follow a strategy of quasi-voluntary compliance. If the appropriator-monitor discovers a rule infraction, it is possible to learn about the particular circumstances surrounding the infraction, to participate in deciding the appropriate level of sanctioning, and then to decide whether or not to continue compliance. If an appropriator-monitor finds an offender who normally follows the rules but in one instance happens to face a severe problem, the experience confirms what everyone already knows: There will always be instances in which those who are basically committed to following the set of rules may succumb to strong temptations to break them.

The appropriator-monitor may want to impose only a modest sanction in this circumstance. A small penalty may be sufficient to remind the infractor of the importance of compliance. The appropriator-monitor might be in a similar situation in the future and would want some understanding at that time. Everyone will hear about the incident, and the violator's reputation for reliability will depend on complying with the rules in the future. If the appropriator-monitor presumes that the violator will follow the rules most of the time in the future, the appropriator-monitor

can safely continue a strategy of compliance. The incident will also confirm for the appropriator-monitor the importance of monitoring even when most others basically are following the rules.

A real threat to the continuance of quasi-voluntary compliance can occur, however, if an appropriator-monitor discovers individuals who break the rules repeatedly. If this occurs, one can expect the appropriator-monitor to escalate the imposed sanctions in an effort to halt future rule-breaking by such offenders and any others who might start to follow suit. In any case, the appropriator-monitor has up-to-date information about compliance and sanctioning behavior on which to base future decisions about personal compliance.

Let us also look at the situation through the eyes of someone who breaks the rules and is discovered by a local guard (who will eventually tell everyone) or another appropriator (who also is likely to tell everyone). Being apprehended by a local monitor after having succumbed to the temptation to break the rules will have three results: (1) It will stop the infraction from continuing and may return contraband harvest to others. (2) It will convey information to the offender that someone else in a similar situation is likely to be caught, thus increasing confidence in the level of quasi-voluntary compliance. (3) A punishment in the form of a fine, plus loss of reputation for reliability, will be imposed. A large monetary fine may not be needed to return an occasional offender to the fold of those who are quasi-voluntary compliers with the rules. A large monetary fine imposed on a person facing an unusual problem may produce resentment and unwillingness to conform to the rules in the future. Graduated punishments ranging from insignificant fines all the way to banishment, applied in settings in which the sanctioners know a great deal about the personal circumstances of the other appropriators and the potential harm that could be created by excessive sanctions, may be far more effective than a major fine imposed on a first offender.

If quasi-voluntary compliance is contingent on the compliance rate of others, then the question is, What rate must be maintained to ensure that the commitment to comply will continue over time? Previous theoretical work has assumed that 100% is needed; but also see M. Taylor (1987, pp. 89–90), who posits less than 100%. It is assumed that any infraction (or error) will trigger a relentless process: Everyone will resolutely punish the offender (and themselves) by breaking their previous agreement. Although these trigger-strategy models have the attractive theoretical property of stable equilibria, they do not describe the behavior observed in our case studies (or any of the other cases I have read or observed in the field). Acceptable quasi-voluntary compliance rates that will lead appropriators to

continue their own quasi-voluntary compliance will differ from one setting to another and will depend on economic or other circumstances within the CPR. Tolerance for rules infractions may be very high during a depression, so long as the higher rate appears temporary and not threatening to the survival of a CPR. This appears to have happened in one of the Japanese villages studied by McKean during the depression of the 1930s:

> Almost all the villagers knew that almost all the other villagers were breaking the rules: sneaking around the commons at night, cutting trees that were larger than the allowed size, even using wood-cutting tools that were not permitted. This is precisely the behavior that could get a tragedy of the commons started, but it did not happen in Yamanaka. Instead of regarding the general breakdown of rules as an opportunity to become full-time free riders and cast caution to the winds, the violators themselves tried to exercise self-discipline out of deference to the preservation of the commons, and stole from the commons only out of desperation. Inspectors or other witnesses who saw violations maintained silence out of sympathy for the violators' desperation and out of confidence that the problem was temporary and could not really hurt the commons. (McKean 1986, pp. 565–6)

In other situations, the harm that a single infraction can inflict on others may be so substantial, and the potential for private gain so great, that 100% compliance is essential. McKean (1986, p. 565) describes a situation in the village of Shiwa when it suffered a severe drought. The temptation to break the dikes, in order to obtain water illegally, was so great for those serving as guards, as well as for the remaining farmers, that all adult males patrolled the dikes every night in mutual surveillance until the emergency was over.

The fourth and fifth design principles – monitoring and graduated sanctions – thus take their place as part of the configuration of design principles that can work together to enable appropriators to constitute and reconstitute robust CPR institutions. Let me summarize my argument to this point. When CPR appropriators design their own operational rules (design principle 3) to be enforced by individuals who are local appropriators or are accountable to them (design principle 4), using graduated sanctions (design principle 5) that define who has rights to withdraw units from the CPR (design principle 1) and that effectively restrict appropriation activities, given local conditions (design principle 2), the commitment and monitoring problem are solved in an interrelated manner. Individuals who think that a set of rules will be effective in producing higher joint benefits and that monitoring (including their own) will protect them against being suckered are willing to make a contingent self-commitment[41] of the following type:

> I commit myself to follow the set of rules we have devised in all instances except

dire emergencies if the rest of those affected make a similar commitment and act accordingly.

Once appropriators have made contingent self-commitments, they are then motivated to monitor other people's behaviors, at least from time to time, in order to assure themselves that others are following the rules most of the time. Contingent self-commitments and mutual monitoring reinforce one another, especially when appropriators have devised rules that tend to reduce monitoring costs. We are now ready to discuss the sixth design principle.

Conflict-resolution mechanisms

6 Appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials.

In theoretical models of rule-governed behavior, the rules that structure the strategies available to participants are unambiguous and are enforced by external, all-knowing officials. In field settings, applying the rules is never unambiguous, even when the appropriators themselves are the monitors and sanctioners. Even such a simple rule as "each irrigator must send one individual for one day to help clean the irrigation canals before the rainy season begins" can be interpreted quite differently by different individuals. Who is or is not an "individual" according to this rule? Does sending a child below age 10 or an adult above age 70 to do heavy physical work meet this rule? Is working for four hours or six hours a "day" of work? Does cleaning the canal immediately next to one's own farm qualify for this community obligation? For individuals who are seeking ways to slide past or subvert rules, there are always various ways in which they can "interpret" a rule so that they can argue they have complied with the rule, but in effect subverting its intent. Even individuals who intend to follow the spirit of a rule can make errors. What happens if someone forgets about a labor day and does not show? Or what happens of the only able-bodied worker is sick, or unavoidably in another location?

If individuals are going to follow rules over a long period of time, there must be some mechanism for discussing and resolving what constitutes an infraction. If some individuals are allowed to free-ride by sending less able workers to a required labor day, others will consider themselves to be suckers if they send their strongest workers, who could be using that time to produce private goods rather than communal benefits. Should that continue over time, only children and old people would be sent to do work

that would require strong adults, and the system would break down. If individuals who make honest mistakes or face personal problems that occasionally prevent them from following a rule do not have access to mechanisms that will allow them to make up for their lack of performance in an acceptable way, rules may come to be viewed as unfair, and conformance rates may decline.

Although the presence of conflict-resolution mechanisms does not guarantee that appropriators will be able to maintain enduring institutions, it is difficult to imagine how any complex system of rules could be maintained over time without such mechanisms. For those cases discussed earlier, such mechanisms sometimes are quite informal, and those who are selected as leaders are also the basic resolvers of conflict. In some cases – such as the Spanish *huertas* – the potential for conflict over a very scarce resource is so high that well-developed court mechanisms have been in place for centuries.

Minimal recognition of rights to organize

7 The rights of appropriators to devise their own institutions are not challenged by external governmental authorities.

Appropriators frequently devise their own rules without creating formal governmental jurisdictions for this purpose. In many inshore fisheries, for example, local fishers devise extensive rules defining who can use a fishing ground and what kind of equipment can be used. Provided the external governmental officials give at least minimal recognition to the legitimacy of such rules, the fishers themselves may be able to enforce the rules themselves. But if external governmental officials presume that only they have the authority to set the rules, then it will be very difficult for local appropriators to sustain a rule-governed CPR over the long run. In a situation in which one wishes to get around the rules created by the fishers, one may go to the external government and try to get local rules overturned. In Chapter 5 we shall examine several cases in which this design principle is not met.

Nested enterprises

8 Appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises.

All of the more complex, enduring CPRs meet this last design principle. In

the Spanish *huertas*, for example, irrigators are organized on the basis of three or four nested levels, all of which are then also nested in local, regional, and national governmental jurisdictions. There are two distinct levels in the Philippine federation of irrigation systems. The problems facing irrigators at the level of a tertiary canal are different from the problems facing a larger group sharing a secondary canal. Those, in turn, are different from the problems involved in the management of the main diversion works that affect the entire system. Establishing rules at one level, without rules at the other levels, will produce an incomplete system that may not endure over the long run.

In the last part of this chapter I have identified a set of design principles that characterize the long-enduring CPR institutions described in the first part. I have also attempted to examine why individuals utilizing institutional arrangements characterized by these design principles will be motivated to replicate the institutions over time and sustain the CPR to which they are related. We shall continue to discuss these design principles throughout the remainder of this study. In the next chapter we shall examine how individuals supply themselves with new institutions to solve CPR problems.

4

Analyzing institutional change

In the preceding chapter I examined institutions for governing CPRs in which appropriators have devised governance systems that have survived for long periods of time in environments characterized by considerable uncertainty and change. Although the particular problems involved in governing mountain commons vary from those involved in governing irrigation systems, all of these long-enduring institutional arrangements have shared commonalities. These cases clearly demonstrate the feasibility (but obviously not the likelihood) of robust, self-governing institutions for managing complex CPR situations, but the origins of these systems are lost in time. It is not possible to reconstruct how earlier users of Swiss alpine meadows, Japanese mountain commons, the Spanish *huertas*, or the Philippine *zanjeras* devised rules that have survived such long periods. We do not know who originated or opposed various proposals, or anything about the process of change itself.

A study of the origins of institutions must address the problem of supply raised in Chapter 2. As Bates (1988) points out, the presence of collective benefits as a result of designing new institutions is itself a second-order collective dilemma. A proposed new institution "is subject to the very incentive problems it is supposed to resolve" (Bates 1988, p. 395). Many questions need to be addressed. How many participants were involved? What was their internal group structure? Who initiated action? Who paid the costs of entrepreneurial activities? What kind of information did participants have about their situation? What were the risks and exposures of various participants? What broader institutions did participants use in establishing new rules? These questions are rarely answered in the extensive case-study literature describing behavior within ongoing institutional arrangements. Once a set of rules is in place, the incentives facing

appropriators are entirely different from the incentives that faced an earlier set of appropriators when confronted with severe appropriation or provision suboptimalities.

In this chapter, the origins of a set of institutions to manage a series of groundwater basins located beneath the Los Angeles metropolitan area are examined. Louis Weschler and I did extensive fieldwork in these areas during the late 1950s and early 1960s, when many changes were occurring (E. Ostrom 1965; Weschler 1968). We attended meetings, read internal memoranda, and interviewed participants to obtain information about the strategies of groundwater producers to organize voluntary associations, to undertake litigation, to create special districts, and to constitute a complex public-private governance system to regulate their basins. Recently, William Blomquist (1987a, 1988a–e) has expanded the number of groundwater basins studied and updated the available information. For these groundwater basins, we have a good understanding of the processes involved in changing the rules, and sufficient time has elapsed to allow us to evaluate the stability and efficiency of the results obtained in using these rules to govern and manage these basins. In this chapter we examine the processes of changing the rules in three basins (Raymond, West, and Central) that have relied on negotiated settlements of water rights as a key element in the transformation of their situation.[1] See Figure 4.1 for a map of the area.

THE COMPETITIVE PUMPING RACE

The setting

In an earlier geologic era, rivers and streams draining the mountains surrounding what has now become the Los Angeles metropolitan area laid down wide and deep bands of sand and gravel that were then partially overlaid by hard layers of clay. The former streambeds are now deep, water-bearing strata that can be thought of as underground reservoirs. These reservoirs are replenished by the rains that fall in the foothills and upper valleys and, to a more limited extent, by precipitation and drainage on the flat coastal plain itself.

In a semiarid region such as Los Angeles, groundwater basins are extremely valuable when used in conjunction with surface supply systems. First, they are sources of inexpensive and high-quality water, as compared with the cost of importing water from long distances. In 1985, the Metropolitan Water District charged $240 per acre-foot (the volume of water that would cover one acre of land with one foot of water) as the wholesale price for imported water from northern California and from the Colorado

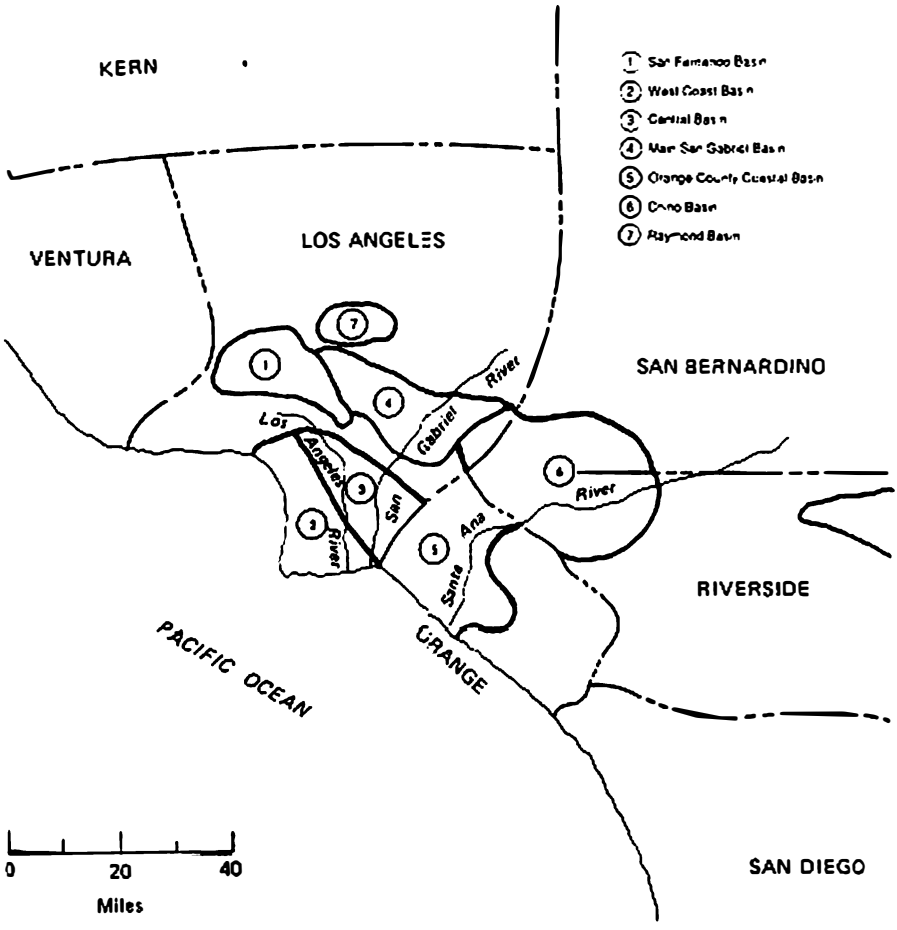


Figure 4.1. Groundwater basins underlying the south coastal plain in California. (Adapted from Lipson 1978.)

River. The cost of pumping groundwater in the Los Angeles area averaged around $134 per acre-foot – a saving of more than $100 per acre-foot. If the 282,458 acre-feet of groundwater that were pumped in 1985 from the three basins discussed in this chapter had been replaced with surface water, it would have cost the industrial users, the urban households, and the irrigators at least $28 million more per year.[2]

The value of the basins as sources of water supply is overshadowed, however, by their even greater value as natural storage vessels that can retain water for use during periods of peak demand.[3] Every surface-water system must have available some type of short-term storage so that it can rapidly meet the accelerated demands of water users that occur at regular intervals during each day and each week, and during the course of a year. The current construction costs for a water tower in the Los Angeles area average around $57,500 per acre-foot (Blomquist 1987a). The minimum amount of short-term storage recommended by the relevant engineering standards is 16% of the total water used in an area. In the area of the West Basin, with an annual demand for water of 327,435 acre-feet, storage reservoirs that could hold 52,400 acre-feet would be required if the basin were not available for this purpose. The replacement costs for this single basin would be about $3.01 billion. The loss of all the groundwater basins underlying the Los Angeles metropolitan area would be an economic disaster of major proportions.

Groundwater basins can be destroyed by overextraction and/or pollution. If more water is withdrawn per year than the average level of replenishment (referred to as the safe yield of a basin), eventually the gravel and sand in the water-bearing strata will compact so that they cannot hold as much water as they formerly did. If a groundwater basin is located near the ocean, and its water level is drawn down below sea level, saltwater intrusion will occur along the coast. Wells along the coastline must be abandoned. If intrusion is not halted, eventually the entire basin will no longer be usable as a source of supply or for its storage capacity. Overextraction threatened all of the groundwater basins in this region until institutional changes were initiated by those affected.

The logic of the water-rights game

Overextraction was the logical outcome of the way groundwater rights were defined prior to the institutional changes described in this chapter. Water rights in California had been defined on the basis of whether a producer owned the overlying land and used the water on that land (an overlying landowner) or used the water to serve areas other than land

owned by the water producer (an appropriator). Under the common law, an overlying landowner held a riparian right to the "full flow" of the water supply underlying his or her land (Nunn 1985). In a region of extreme scarcity of water, the common law does not provide secure rights for an overlying landowner. Water underlying any parcel of land (e.g., parcel A) can be siphoned to a neighbor's land if the neighbor withdraws water more rapidly than does the owner of parcel A. In *Katz v. Walkinshaw* [141 Cal. 116, 74 P. 766 (1903)], the doctrine of "correlative rights" was developed to replace the strict interpretation of riparian rights. That doctrine held that in times of shortage, if the court was called on to adjudicate among competing interests, the court would treat all overlying owners as correlative and coequal owners. In times of scarcity, each would gain a *proportionate* share of the water rather than an *absolute* share of the water. That doctrine was modified somewhat in *San Bernardino v. Riverside* [186 Cal. 7 (1921)], in which overlying landowners were limited to taking only water that they could put to "beneficial" use.

Thus, overlying landowners facing only other overlying landowners knew that if they went to court to settle a dispute over water rights during a time of shortage, they would all share proportionately in any cutback in the total water available to them. In most groundwater basins, however, overlying landowners faced other water users called "appropriators," whose claim to water was on a different basis than that of an overlying landowner. Appropriators pumped groundwater to be used on land not owned by those withdrawing the water. Most private and public water companies were legally classified as appropriators, because the water they pumped was used by their customers, not by the water companies themselves. Nonoverlying landowners were allowed, if not encouraged, by the appropriative-rights doctrines made part of the statutory law in 1872 to withdraw "surplus water" or water that was not being put to beneficial use by the overlying landowners. The key elements in defining the rights of an appropriator had to do with

1 when the appropriator began to withdraw water from the source,
2 how much water was actually put to beneficial use, and
3 whether or not the use was continuous.

Under the doctrine of "first in time, first in right," appropriators acquired rights depending on their history of use. Among appropriators, a court-resolved conflict over a scarce supply would exclude use by the most junior appropriator, and then the next most junior appropriator, and so forth. The most senior appropriators would be fully protected against encroachment on their rights by more junior appropriators. However, the rights of

the most senior appropriators were potentially subordinate to those of overlying landowners.

The simultaneous existence of the doctrines of correlative and appropriative rights in the same state introduced considerable uncertainty about the relative rights of one groundwater producer against others. The uncertainty was compounded by the presence of a third common-law doctrine that enabled groundwater producers to gain rights through "adverse use" or prescription. In regard to land, prescriptive rights are relatively straightforward: If one person occupied someone else's land in an open, notorious, and continuous manner for a set period of time (five years in California), and the owner makes no effort to eject the occupier, the original owner loses the right to the land.

In regard to groundwater, possession of water was not enough to establish open and *adverse* use. Any junior appropriator could legally use any water that was surplus water. Surplus water was defined as a part of the "safe yield" of a basin that was not of beneficial use to overlying landowners or senior appropriators. The safe yield of a basin is the average, long-term supply of water to the basin. If that quantity of water was put to beneficial use, no surplus was available to others. An appropriator had to take nonsurplus water openly and continuously for more than five years to perfect prescriptive rights. Once perfected, prescriptive rights were superior to those of overlying owners and appropriators. The same actions of an appropriator – openly taking water continuously from a basin – could lead to the acquisition of rights *superior* to those of overlying landowners or, alternatively, to the *inferior* rights of a junior appropriator relative to an overlying landowner in time of scarcity. The key difference between these outcomes was whether the court ruled that a surplus did or did not exist for the five-year period prior to litigation. Given that all producers suffered from lack of information concerning the safe yield of a basin and the pumping rates of other producers, no one knew at the time of making such decisions what the pumping rates were or whether or not a surplus existed.

The situation in these basins can be characterized as an open-access CPR for which clear limits have not been established regarding who can withdraw how much water. In such situations, two strong pressures encourage pumpers to adopt inefficient strategies. The first is a pumping-cost externality. The second is a strategic externality (Negri 1989). Pumping costs increase as the pumping lift increases, because of falling water levels, and therefore each person's withdrawals increase the pumping costs for others. No one bears the full cost of personal actions. Each pumper is consequently

led toward overexploitation. The strategic externality involved in an open-access groundwater basin is aptly described by Negri (1989, p. 9).

> With property rights undefined and access nonexclusive, the "rule of capture" governs the "ownership" of the reserve stock. The rule of capture grants [pumpers] exclusive rights to that portion of the groundwater that they pump. What an operator does not withdraw today will be withdrawn, at least in part, by rival[s]. The fear that [pumpers] cannot capture tomorrow what they do not pump today undermines their incentive to forgo current pumping for future pumping.

The two incentives reinforce one another to aggravate the intensity of the pumping race. Without a change of institutions, pumpers in such a situation acting independently will severely overexploit the resource. Overexploitation can lead to destruction of the resource itself.

Current institutions affect not only the intensity of a pumping race but also the relative incentives of different participants to initiate institutional change. Given the legal structure of rights in California, overlying landowners were more motivated than appropriators to launch court action so as to keep appropriators from obtaining prescriptive rights. The decision about when to start litigation, however, involved high risks of being too soon or too late. The overlying owner faced two possibilities:

> (1) If he went to court before all "surplus" water had been appropriated, and the court ruled that the water being diverted by the defendant was indeed surplus water, the overlying owner would suffer the costs of the litigation and receive no remedy;
> (2) If he waited too long to go to court, the overlying owner might find that the defendant had perfected a prescriptive right if the court ruled that the water being diverted was non surplus water. There was, in other words, no way for the overlying owner, on whom the burden of initiating litigation rested, to succeed in protecting his right until it had been invaded, and yet within a short time after the right had been invaded, the overlying owner would have lost the right he sought to protect due to prescription.
>
> (Blomquist 1988a, p. 19)

The uncertainty of the competing water doctrines was compounded by the uncertainty shared by all water producers about the actual supply of water to a basin and the quantity of water withdrawn by all of the parties. It was essential to know the quantities supplied and demanded from a basin to determine the presence or absence of a surplus. Both types of information were costly to obtain. Both could be obtained at the time of litigation

by asking the court to appoint a watermaster to make a geologic survey of the basin and determine its water supply and to obtain information about the past water uses of all producers. When determined in this manner, the cost would be shared by all producers involved in the litigation. But that did not solve the problem of uncertainty prior to the initiation of litigation. In past cases, signs of potential problems – such as falling water tables – had not been accepted by the court as sufficient evidence of a water shortage to declare a lack of surplus and uphold the rights of overlying owners as against junior appropriators [*San Bernardino v. Riverside*, 186 Cal. 7 (1921)].

Given these compound uncertainties, it is easy to explain the behavior of groundwater pumpers in the Los Angeles metropolitan area during the first 50 years of this century. To obtain any kind of water right, one needed to show continuous withdrawal of water and application to beneficial use. In that environment of legal uncertainty, attorneys advised producers to pump as much as they needed and to defend later (Krieger 1955). A pumping race occurred in each of the groundwater basins underlying the Los Angeles area.

Given those incentives, many water producers and local government officials during the 1940s and 1950s worried that all of the basins would be severely overdrawn and that those basins located adjacent to the ocean – West Basin and Central Basin – would be lost to the sea. By the 1960s, however, the pumping race had been halted in all of the coastal basins. Water rights were eventually established in all the basins, except in Orange County, which continues to rely on a pump tax for regulation.[4]

Special water districts have been established throughout the area to obtain surface water, to levy pump taxes on water production, and to replenish the basins through a variety of artificial means. A series of injection wells has been constructed along the coast to create a barrier of fresh water against the sea, enabling the coastal districts to regulate the uses of their basins in a manner similar to the use of a surface reservoir. In other words, diverse private and public actors have extricated themselves from the perversity of the pumping race and transformed the entire structure of the incentives they face. Public arenas were involved in many stages of these developments. The initial steps were taken in the shadow of a court order. Elections and public hearings were held at key stages. The solutions to the pumping race, however were not imposed on the participants by external authorities. Rather, the participants used public arenas *to impose constraints on themselves*. Because litigation to gain defined water rights was involved in all of the basins, except in Orange County, we first discuss this strategy to transform the pumping race.

THE LITIGATION GAME

The Raymond Basin negotiations[5]

The Raymond Basin is a small basin, with a surface area of 40 square miles, located inland and thus protected from saltwater intrusion. The area was already highly developed by the turn of the century. Later studies have revealed that the safe yield of the basin was steadily exceeded from 1913 onward. The cities of Pasadena, Sierra Madre, Arcadia, Altadena, La Cañada–Flintridge, South Pasadena, San Marino, and Monrovia are located on the surface of the basin. The city of Alhambra lies on its borders and appropriates water from the basin for use within its boundaries. The city of Pasadena was by far the largest producer of water from the basin – its production equaled the production of the other 30 producers combined. Pasadena thus approached, but did not reach, the position of a dominant actor in a privileged group (Olson 1965). According to Olson's model, if the Raymond Basin producers had been a privileged group, the city of Pasadena would have borne all of the costs associated with stopping the pumping race. The prediction one derives from Olson's model is consistent with some, but not all, of the activities pursued by the city of Pasadena.

The city of Pasadena for some years adopted the strategy of the dominant player in a privileged group. From 1914 to 1923, for example, the city replenished the basin by capturing floodwaters and spreading them on the gravel areas located at the feet of the San Gabriel Mountains. The water that percolated into the basin was then available for capture by the city of Pasadena as well as by other groundwater producers. In the late 1920s, the city of Pasadena was a leading participant in the formation of the Metropolitan Water District of Southern California, which would eventually construct an aqueduct to bring water 250 miles to the Los Angeles area from the Colorado River.

During the 1930s, however, the city of Pasadena was no longer willing to undertake independent actions that were substantially benefiting others who were not contributing to the costs. The city tried unsuccessfully to negotiate a voluntary settlement with the other producers whereby all producers would jointly reduce the amounts of water they were withdrawing from the basin. In 1937 Pasadena initiated legal proceedings against the city of Alhambra and 30 other producers.[6] The case was referred to the Division of Water Resources of the California Department of Public Works for determination of the geologic structure of the basin, the safe yield of the basin, and whether or not there was a surplus.

That referral procedure was time-consuming and costly. The draft report of the referee was not completed until March of 1943 and cost about

$53,000. The referee found that the yearly withdrawals from the basin were 29,400 acre-feet, whereas the safe yield of the basin was 21,900, leading to an annual overdraft of 8,500 acre-feet per year. The referee recommended that the parties curtail their pumping to the safe yield of the basin.

The parties then shared a single, authoritative "image" of the problem they faced. They also would confront a new "default condition" (E. Ostrom 1986a) if they could not agree on their own solution. Prior to litigation, the failure to agree would simply mean a return to the pumping race. Once the court took jurisdiction, an absence of agreement would mean that the judge would decide which parties had to bear the brunt of the cutback. It was not at all clear what the judge would decide. The judge might, for example, assign preeminent rights to the overlying owners and then assign the remainder of the 21,900 acre-feet as a "surplus" to the appropriators to be apportioned according to their seniority. Or the judge might decide that there was no surplus. In that case, senior appropriators might be granted prescriptive rights, and overlying landowners would bear the major brunt of the cutback.

A simplified picture of the bargaining problem that the producers faced is shown in Figure 4.2. If we assume that the overlying owners were withdrawing 12,000 acre-feet and the appropriators (who might become prescriptors) were withdrawing 18,000 acre-feet, the total withdrawals prior to a decision would be 30,000 acre-feet. Everyone accepted the fact that a cutback to 22,000 acre-feet would occur. A worst-case analysis done by the overlying landowners would assume that the judge would declare that there had been no surplus for more than five years prior to litigation. Thus, the appropriators would be given superior rights to all that they had withdrawn. They would be assigned 18,000 acre-feet, leaving only 4,000 acre-feet for the overlying landowners. Point B marks the worst possible solution that the overlying owners could face.

Similarly, the appropriators could do a worst-case analysis and assume that the judge would assign firm rights of 12,000 acre-feet to the overlying landowners and then assign the "surplus" of 10,000 acre-feet to the appropriators according to their seniority. Point A is the worst possible solution from the perspective of the appropriators. For all participants, the range of variation between complete protection and major loss (the line connecting A and B) would be considerable. Further, a fully contested trial would last a long time, given the conflicting legal doctrines, and the costs of litigation would be extremely high.

At the instigation of the city of Pasadena, the parties held some serious negotiations in the shadow of the court. Within six months they had

drafted a stipulated agreement signed by all but 2 of the 32 parties involved in the litigation. The negotiation process was furthered in that instance by the unusual fact that one attorney, Kenneth Wright, represented 16 of the parties. After another six months, one of the holdouts also agreed to the stipulation. The other – the California-Michigan Land and Water Company – never agreed to the stipulation and challenged the final court decision based on the stipulation.

The signatories agreed that the safe yield had been exceeded for a long time and that it was necessary to cut back to the safe yield of the basin. They stated that each producer's withdrawal of groundwater had been open, continuous, and notorious and was, because of the overdraft, adverse to the claims of all of the others. Thus, each producer had prescribed against all of the others. The term "mutual prescription" has been used to describe the concept used by these parties as the foundation for their negotiated settlement. The signatory parties agreed to *share the cutback proportionately* instead of pursuing further legal procedures to determine whose rights took precedence.[7] The proportional division of the cutback is represented by point D in Figure 4.2.[8] They further guaranteed each other's proportional shares of the safe yield (if it were to change in the future) and established an arrangement to enable those most adversely affected by the

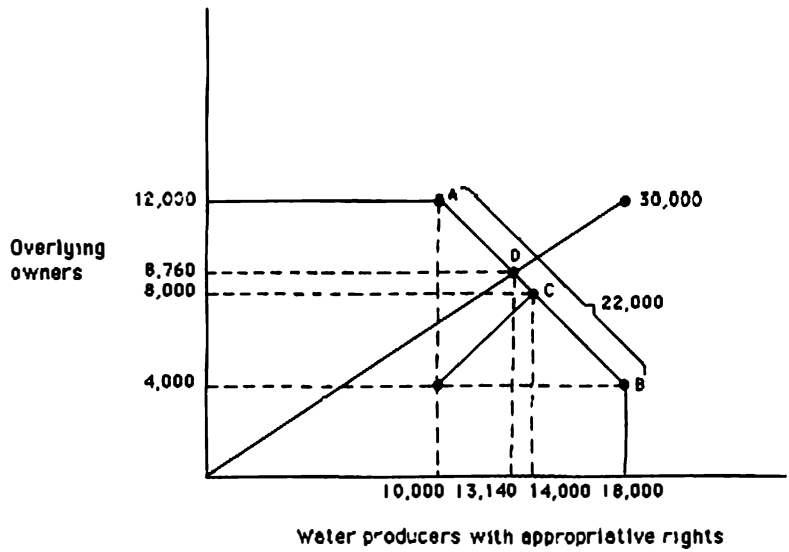


Figure 4.2. The bargaining situation faced by overlying owners and appropriators.

cutback to obtain exchange rights from others willing to sell their rights on an annual basis.

A short trial was held to hear the objections of the California-Michigan Land and Water Company and to assign the Division of Water Resources of the California Department of Public Works to serve as the watermaster–an official monitor – to supervise the agreement. Rather than imposing his own solution,[9] the judge, after considerable reflection, issued a final judgment on December 23, 1944,[10] based on the stipulated agreement. The final judgment declared all of the decreed rights to be of equal standing in any future dispute and enjoined all parties from taking more than their decreed rights. The judgment continued the role of the watermaster to enforce the provisions of the judgment and to supervise the exchange pool they had developed. In addition to the leasing arrangements of the exchange pool, decreed rights could be leased or sold outright so long as the transfers were recorded by the watermaster. Two-third of the costs of the watermaster were to be paid for by the parties, and the state of California would pay for the remaining costs of monitoring the agreement. The case was appealed to the California Supreme Court, and the decision was upheld.[11] The United States Supreme Court declined to review the case.

By negotiating their own agreement, the parties had ended the pumping race faster and at a lower cost than they could have through a court proceeding.[12] They also had gained firm and marketable rights to defined shares of the safe yield of the basin. A market for those water rights developed, and most of the smaller right-holders have sold their rights to the water companies, for whom the rights have a higher value. There are now 17 active producers from the basin, and they are almost all municipal or private water companies. Only three overlying landowners continue to produce water from the basin. The areas within the basin that did not have access to imported water formed a municipal water corporation in 1953 and started receiving imported water in 1955.

The West Basin negotiations

West Basin, with a surface area of 170 square miles, is a much larger area than Raymond Basin. Located immediately adjacent to the ocean, it extends from the city of Inglewood to the Palos Verdes Peninsula. The situation in West Basin was not as favorable for negotiations as that facing the Raymond Basin producers. The major advantage for the West Basin producers was that their upland neighbors had already borne the costs of innovation and had developed a formula for reaching a negotiated settlement within the California legal environment that was considered to be fair

by many potential litigants. The disadvantages faced in West Basin included (1) a large number of producers (around 500 parties were named in the litigation), (2) the absence of a single dominant producer, and (3) considerable asymmetry in the risks regarding saltwater intrusion (those near the sea would lose their wells long before those pumping inland). The problem of the size of the group was offset to some extent by the concentrated nature of the groundwater production in the basin: 19 producers accounted for about 85% of the total quantity of water withdrawn from West Basin.[13]

The overdraft came a decade later to West Basin than to Raymond Basin. The heavy industrialization that occurred during World War II exacerbated the already growing overdraft, particularly because there were many oil companies located in the area whose water production had increased steadily. In the early 1940s, wells located along the coast began to show increasing salinity. Many water producers in the basin continued to believe, however, that the salinity in those wells was symptomatic of only a "local" problem immediately along the coast, not a more general problem that could affect their own situation in the future. During 1943, nine of the coastal municipalities met several times to discuss the importance of the increasing salinity of their wells. They agreed that more information was needed to gain a realistic and common image of the structure of the groundwater basin. Those cities signed a cooperative agreement with the United States Geological Survey and the Los Angeles County Flood Control District to undertake an initial study of the problem of groundwater supply in the basin.[14]

The report, completed in 1944, painted a grim picture. Wells all along the coast had been invaded by seawater. The investigators had found no natural barrier at any point in the basin to halt the advance of the sea. The entire basin was threatened with destruction. The report provided a common image of the general boundaries of the basin and the extent of the problem without providing an exact picture of the safe yield and current levels of water production. It was no longer possible, however, for producers to maintain that the salinity in the coastal wells was strictly a local problem.

In December of 1944, all of the major water producers met and established an ad hoc committee to consider what should be done next. That committee had three major recommendations:

1 that a permanent association be created of all interested water producers so that they could continue to discuss their mutual problems and possible joint actions,

2 that a technical survey be made of alternative sources of water for the area, and
3 that water producers consider initiating legal action similar to the action just completed in Raymond Basin to reduce total pumping and to ration the limited water supply in West Basin among all water producers (Ways and Means Committee 1945, p. 16).

All three recommendations were followed. The West Basin Water Association was created within a few months.[15] The association provided a continuous open forum[16] for discussion of all major steps taken in West Basin by producers and representatives of various local, regional, and state public agencies. The resources of the association frequently were used to obtain and make available the best possible technical information about the basin. Extensive minutes were kept for all West Basin Water Association meetings, as well as the meetings of the Executive Committee and most of the working committees of the association. Those files were open to all members, as well as to others interested in gaining information about past decisions, technical data, and studies of the benefits and costs of alternatives. A weekly newsletter was dispatched to all members from 1946 through 1954. The motto of the newsletter, according to its editor, was "let there be no surprises, either pleasant or unpleasant" (Fossette and Fossette 1986, p. 57). The practice of obtaining the best information available and disseminating it widely increased the degree of understanding and level of cooperation among the participants.

The first official act of the association, in March 1946, was to retain a renowned engineer, Harold Conkling, to examine the possibility of finding alternative sources of supplemental water for the basin. Conkling recommended the creation of a municipal water district to import water from the Metropolitan Water District of Southern California. Obtaining surface water eventually would mean that the groundwater basin would no longer serve as the major source of water for the area, but the question who would obtain rights to use the reservoir capacities of the basin had to be resolved.

Three appropriators initiated the West Basin litigation in October 1945: the California Water Service Company, the city of Torrance, and the Palos Verdes Water Company. Kenneth Wright, who had served as the attorney for the city of Pasadena in the Raymond Basin litigation, was the attorney for the California Water Service Company and had made several presentations to West Basin producers concerning the mutual-prescription concept used in the Raymond Basin case. Although the initiators of the litigation, and many other water producers in the basin, strongly supported the

concept of proportionate cutbacks by all water producers, several major water producers vigorously opposed such a plan.

The Dominguez Water Corporation, a senior appropriator with overlying rights as well, was one strong opponent. Because Dominguez was the largest producer from the basin,[17] it was unlikely that others would agree to a curtailment without the participation of the Dominguez Corporation. The city of Inglewood initially opposed the litigation and all of the actions proposed within the context of the West Basin Water Association. Inglewood's lawyers had advised city officials that its status as senior appropriator would protect it from having to cut back production. Inglewood's position changed, however, after the Raymond Basin decision had been sustained in the California Supreme Court.[18] Inglewood, which owned some wells near the sea, was to become an active participant in the effort to find solutions.

The city of Hawthorne, on the other hand, was located inland, and its people believed that their water supply was protected. Hawthorne adopted a hold-out strategy for many years. Thus, whereas the Raymond Basin case was a guiding model in the minds of the initiators of the litigation, it was not at all certain that the water producers of West Basin would achieve the high level of agreement needed to negotiate their own settlement. Once litigation had been initiated, however, the court could impose its own judgment if the water producers could not reach an agreement on their own. Thus, again the default rule had been changed by the initiation of litigation.

The case was referred to the Division of Water Resources of the California Department of Public Works. The difficult task of ascertaining the production levels for more than 500 producers and determining the geologic structure and inflow levels for a large and complex basin took four years. By the time the referee's report was completed, the decision of the trial court based on the stipulated agreement in the Raymond Basin case had been upheld by the California Supreme Court. Therefore, the West Basin water producers knew that the mutual-prescription concept could withstand a legal challenge by a private company.

The referee's findings and recommendations, however, came as a written bombshell. The referee found that overdrafts had been occurring since 1920 and that the safe yield of the basin was 30,000 acre-feet per year. The referee recommended a curtailment to the safe yield. By 1952, total groundwater withdrawals had reached 90,000 acre-feet per year. Even the supporters of mutual curtailment vigorously opposed a two-thirds reduction in groundwater production. Imported water had just begun to trickle

into the basin. Many water suppliers would not be able to meet their customers' demands if they were to reduce the quantity of water pumped by two-thirds. Early experiments with injection wells provided some encouraging indications that the supply of water to the basin could be increased. An increase in the supply would reduce the necessity of cutting back to the safe yield. The default condition, however, had again been changed. If the water producers were unable to arrive at their own settlement, they could expect the court to order a two-thirds cutback.

The West Basin Water Association provided a forum for serious negotiations about a settlement. The association established a Legal Settlement Committee composed of six attorneys and five engineers. The creation of the Legal Settlement Committee within the association changed the structure of the bargaining situation in subtle but important ways. Although the 11 committee members continued to represent the interests of their own firms, they were accountable to the members of the association as well. The association charged the committee with the responsibility for achieving *timely* curtailment of water production. The committee had to report quarterly to the full membership. The committee members would be subject to public criticism by respected colleagues if they simply pursued recalcitrant strategies and failed to find sources of agreement on which progress toward settlement could be based.[19] The members of this committee were expected to achieve an agreement whereby *all* parties would curtail their withdrawals. The first question to be resolved concerned how much curtailment.

The negotiators had to find a method to reduce withdrawals below 90,000 acre-feet and above the 30,000 acre-feet recommended by the referee. If the negotiated settlement was not above the referee's recommendation, some litigants would prefer to contest the matter in court in the hope that a judge would give their claims precedence over those of others. The engineers on the committee were asked to determine the maximum cutback that the parties could undertake in the near future without grave economic damage. The engineers concluded that a reduction of 25% to 30% could occur without serious economic harm to any water producer, if an exchange pool similar to the one devised in Raymond Basin were established.

Next the committee searched for a particular formula, based on the concept of mutual prescription, that would enable them to achieve a proportional cutback of 25% to 30%. Because 340 additional parties had been added to the case in 1949, one possibility was to use 1949 for determining shares, rather than the 1944 water year that immediately preceded the initiation of litigation. Using the referee's historical findings, the committee

compiled estimates of each party's "prescriptive rights" based on the 1944 water year versus the 1949 water year.[20] Their estimates totaled 44,387 acre-feet for 1944 and 63,728 for 1949. The committee proposed to use the 1949 data as the basis for negotiating an *interim* agreement that the parties could ratify immediately in order to achieve an actual cutback within a short time. One member of the committee reasoned that

> with the present usage in the amount of 90,000 acre-feet and . . . with the historical usage of 1949 amounting to 63,000 acre-feet or one-half way back to where the Division wanted the curtailment to go, a cutback to 1949 might be more acceptable at the present time . . . . [T]he parties would have enough water left under this arrangement to meet peak demands and it would afford a period in which to adjust to curtailment and . . . no one would be giving up any prescriptive rights already acquired. (West Basin Water Association, Legal Settlement Committee minutes, February 25, 1953, p. 4)

The interim agreement was drafted as a contingent contract. In other words, a water producer who signed the agreement and thus promised to curtail production to his own "Prescriptive Rights, 1949" was not committed to curtail until producers representing at least 80% of the total "Prescriptive Rights, 1949" had signed and the agreement was presented to and approved by the court. A signatory was committed to undertaking this "cooperative action" only if most of the other large water producers were also committed to the action. Thus, no one would be a "sucker," and the joint impact of their curtailments would make a substantial difference. By November 1954, agencies representing 82.5% of the total "Prescriptive Rights, 1949" had signed the agreement, and it was filed with the court. The court appointed the referee to continue as the official watermaster to ensure that the provisions of the agreement were followed.

It had taken two full years of negotiations and the threat of court action[21] to achieve this interim agreement, but at last a major change in the basic rules affecting the use of West Basin had occurred. Water levels in the basin rose immediately and continued to rise for several years, except in a water trough underlying the city of Hawthorne, which refused to sign the agreement.

The interim agreement was used for seven years while the water producers pursued other strategies to enhance the local water supplies, to replenish the basin, and to try to convince nonsignatories to agree to the curtailment. Two major parties did not sign. The first was the California Water Service Company, which had been one of the three initiators of the litigation and had borne a large share of the cost of the litigation. That company had not increased its water production after 1944, presuming that the litigation had protected its interests and that it could afford to take

an independent action to conserve water supplies. The choice of 1949, rather than 1944, as the date for determining rights meant that some of the water producers who had increased production during the four-year period gained somewhat proportionately, while California Water Service slipped behind a little in its proportionate share.[22]

Although it refused to sign the interim agreement, California Water Service Company voluntarily limited its own groundwater production. It did not pump any more water than allocated to it under the interim agreement. Consequently, the effect of the company's refusal to sign the interim agreement was to shift the burden of the cost of watermaster services back onto those who had gained proportionately more rights under the Agreement. The company's actions imposed no physical harm on others. Further, the company clearly did not plan to challenge efforts to make the interim agreement the basis for a final settlement.

On the other hand, the city of Hawthorne increased its withdrawals. By 1960, Hawthorne pumped more than 2,250 acre-feet in excess of its allocation under the interim agreement. During the period of the interim agreement, Hawthorne saved at least $100,000 by pumping more groundwater per year than it had been allotted. As Hawthorne's production increased, the pumping trough beneath the city continued to drop. The watermaster's report for 1960–1 (plate 4) shows that the 1961 water levels below Hawthorne averaged 30–40 feet below those for surrounding territories (California, State of, 1960–1). Nearby producers were harmed substantially.[23] The economic costs of Hawthorne's action were spread generally among all signatories who paid higher costs for imported water while Hawthorne continued to utilize the least expensive source of water.

From the perspective of Hawthorne's leaders, however, the problem seemed different. Instead of viewing the basin as something jointly owned by all water producers, Hawthorne viewed its needs to serve a municipality with water as superior to the needs of industry in the area. Hawthorne saw the interim agreement as favoring the industrial producers, an effort to take away water rights that should be devoted to public use. Hawthorne looked to other cities for support for its position. However, the beach communities had already suffered severe hardship because of saltwater intrusion. According to Hawthorne officials, those communities were willing to see any basis used to curtail production from the basin and slow down the saltwater intrusion. Hawthorne argued that the beach cities were giving away their rights.

During most of 1957 and 1958, the Legal Settlement Committee met weekly and sometimes biweekly trying to prepare a final agreement. The

technical problems of tracing all water-rights transactions for such a large group delayed the process substantially. The substantive problems were also considerable. Some signatories of the interim agreement opposed any final agreement that would not achieve curtailment down to the safe yield. Others preferred to wait until experiments with the saltwater barrier could establish the feasibility of being protected against the sea. The lack of *total* agreement to the interim agreement disturbed many who feared that Hawthorne or others might appeal. Given the experience of a costly appeal to the California Supreme Court in the Raymond Basin case, negotiators hoped to avoid an appeal of their settlement. Because the interim agreement afforded partial physical protection, many signatories believed that they had time to work out an agreement satisfactory to all parties.

Envoys were sent to the city of Hawthorne to urge city officials to reconsider their previous stance in regard to the interim agreement.[24] In 1958, the association appointed the mayor of Hawthorne to its Executive Committee in the hope that he would be able to change the attitude of other city officials. However, the attempts to reach an agreement with Hawthorne were unsuccessful. A final draft of a proposed "Agreement and Stipulation for Judgment" was presented to a meeting of the West Basin Water Association (WBWA) in February 1960. The Dominguez Water Corporation, as the largest water producer from the basin, gave the final agreement its full support by bringing signed copies of the agreement to the meeting. The city of El Segundo and Chanslor-Canfield Midway Oil Company joined Dominguez in this effort to show immediate support (WBWA minutes, February 25, 1960, p. 8). Three months later, 20 parties representing 32.5% of the total adjudicated rights had signed the agreement (WBWA minutes, May 26, 1960, p. 15). Obtaining the remaining signatures took one more year. By early summer of 1961, producers holding 82% of the adjudicated rights had signed the agreement. The Legal Settlement Committee indicated that it was unlikely that further signatures could be gained.

On July 21, 1961, 16 years after the litigation was initiated, a short trial was held, and the proposed judgment was presented to the court. The judgment was entered in August of 1961 substantially as presented to the court. As of October 1, 1961, all entities included as parties in the case were "perpetually enjoined and restrained from pumping or otherwise extracting from the Basin any water in excess of said party's Adjudicated Rights" (Judgment, *California Water Service Company et al. v. City of Compton et al.*, Civil Case No. 506806, Superior Court of the State of California in and for the County of Los Angeles, Sec. 5). Ninety-nine parties were found to

have adjudicated rights of 64,065 acre-feet.[25] The city of Hawthorne, like all nonsignatories, was placed under legal order to reduce its groundwater production to that stipulated in the agreement.

At the association meeting following the trial court's decision, a city councilman from the city of Hawthorne rose to congratulate the group on their "victory," but he warned them that his city planned to fight the decision "through every court in the land." The city of Hawthorne backed up that threat by retaining a firm of highly respected attorneys specializing in water law. At first, the association assumed the financial responsibility for supporting the judgment against the Hawthorne appeal. In 1962, when a Replenishment District had been formed, as described later, the new district undertook financial responsibility for defending the judgment, while the association maintained direct relations with the attorneys. After hearing the Hawthorne appeal, the District Court of Appeals concluded that the trial court had acted properly and affirmed its decision. After the California Supreme Court declined to review the decision made by the District Court of Appeals, the California Water Service case closed 18 years after it had opened.

No one really knows the exact costs involved in the West Basin litigation, given the large number of parties and the length of time involved, but the best available estimate is $3 million (Blomquist 1987a, p. 39). On the one hand, that was 10 times as expensive as the Raymond Basin negotiations. On the other hand, it was one-tenth of costs that would be involved in replacing the short-term storage capacities of the basin when used in conjunction with a surface supply. Amortizing the costs of the litigation over a 50-year period (as one would do for the construction of a major physical facility), the adjudication in Raymond Basin amounted to an annualized cost of 50 cents per acre-foot of water rights allocated, whereas the adjudication costs in West Basin amounted to an annualized cost of $2.50 per acre-foot of water rights (Blomquist 1987a, p. 39). In 1985, the annual costs of monitoring these water rights were $3.00 per acre-foot in Raymond Basin and $2.40 per acre-foot in West Basin.[26]

Adjudicating the water rights in West Basin was only one of a long series of steps taken by water producers to regulate their basin. Some of the subsequent steps are discussed later. The Raymond Basin and West Basin experiences were closely watched by water producers located in Central Basin. These producers also used court litigation as the setting in which to negotiate settlements of their individual rights to water. Central Basin is larger and more diverse than West Basin. Considerable effort was expended there to learn from the difficulties of the West Basin case and to adopt the process so as to reduce both the length of time needed to achieve

agreement and the high costs. The effort was successful on both counts. We now turn to a brief discussion of the Central Basin negotiation process.[27]

The Central Basin litigation

Overdraft conditions in Central Basin occurred much later than they had in Raymond Basin and West Basin. Central Basin is quite large (277 square miles of surface area) and was being used by around 750 owners of wells in the 1950s. The overdraft in Central Basin began in 1942. Most of Central Basin is located inland, and it is protected from the ocean on its western border by its downstream neighbor, West Basin. Thus, water producers in Central Basin are able to draw down their water levels farther than are those in West Basin without immediate adverse consequences. On the other hand, Central Basin does have a small southern exposure to the sea, and saltwater intrusion did begin to occur along that boundary as early as 1950.

At the prodding of their downstream neighbors, Central Basin water producers formed the Central Basin Water Association in 1950 using an organizational structure similar to that of West Basin. The part-time executive director of the West Basin Water Association became the part-time executive director of the newly formed Central Basin Water Association.[28]Some West Basin water producers were also active in Central Basin. Discussions immediately focused on the importance of achieving a negotiated settlement of the water rights in Central Basin. Central Basin producers, however, wanted to avoid the long delay and high cost of using the court-ordered reference procedure and avoid involving all of the very small water producers in the basin.

Consequently, instead of moving immediately toward the initiation of a suit, the Central Basin Water Association employed the services of a private engineering firm, well known for its expertise in the area of groundwater basins, to conduct an initial survey of conditions in the basin and of past water use. Further, a considerable effort was made to achieve a general agreement about the type of negotiated settlement they would reach before they actually went to court in 1962. An interim agreement, signed by parties holding 79% of the water rights, was approved by the court just 10 months after the litigation was initiated. The producers agreed to cut back production on a proportional basis by 20% and to establish a set of working rules modeled on the West Basin agreement, but also reflecting the particular circumstance in Central Basin. Watermaster services were initiated in October 1962. A voluntary cutback of approximately 45,000 acre-feet was initiated immediately (Fossette and Fossette 1986, p. 182). The

final settlement, signed by parties holding over 75% of the rights, was approved by a judge in October 1965 and went into effect in October 1966 – four years after initiation. The estimated costs of the Central Basin litigation were $450,000.

The litigation and negotiation processes in these three basins involved different problems and followed different paths. In Raymond Basin, the number of pumpers was relatively small, and one participant – the city of Pasadena – was more dominant than was any participant in West Basin or Central Basin. Pasadena withdrew about one-half of all water obtained from the basin. The city could not ignore the action of the other pumpers, because their actions could adversely affect joint outcomes. However, Pasadena had such a large stake in seeing that the basin was preserved that the city was willing to invest heavily in achieving a settlement. After failing to obtain a voluntary agreement to curtail pumping, the city initiated legal action and bore more than its proportionate share of litigation costs. By initiating efforts to obtain an external water supply and to control pumping from the basin before pumpers had become accustomed to withdrawal far in excess of safe yield, it was physically possible for all pumpers to cut back their water withdrawals and still serve the growing urban population settling in the area. The major asymmetry of interest faced by the litigants in Raymond Basin was their legal status as overlying owners or appropriators. By devising a new legal concept of mutual prescription, the parties found a basis to share the costs of curtailing groundwater production equitably. All pumpers could continue to use the basin for peaking purposes or could sell water rights, which had been well defined, to those who placed a higher value on acquiring such rights.[29]

Negotiators in West Basin faced three disadvantages not faced in Raymond Basin: (1) the large number of parties involved, (2) the absence of a dominant party, and (3) the asymmetrical risks faced by inland pumpers versus coastal pumpers. The negotiation process took longer, was more expensive, and involved a major conflict between coastal and inland pumpers. By using mutual prescription as the basis for an agreement, the parties reduced other potential asymmetries of interest that could have exacerbated the conflict. Once a final judgment was reached, all parties shared proportionately in the cost of curtailment. The process took sufficient time, however, that it became difficult to cut back to the safe yield and still serve the urban population, which had increased in the years following World War II. As discussed later, water producers in the area had to turn to other mechanisms to increase the supply of water to the groundwater basins, because the control over demand that they achieved did not bring the basin into balance.

Pumpers in Central Basin had several advantages in this process. Because the basin was very large and had only a small coastal section, pumpers could safely delay resolution of their water rights while they watched the process in the other basins. They were prodded into action by their downstream neighbors, who feared that lack of action in Central Basin might eventually negate the benefits of conservation in West Basin. By acting before it would be necessary to cut back much more than 20%, and using a private firm to gather much of the information before they went to court, Central Basin pumpers saved themselves considerable time and money in achieving a negotiated settlement based on the same principle that had been used in Raymond Basin and West Basin.

Conformance of parties to negotiated settlements

Forty-five years have passed since the judgment was entered in the Raymond Basin case, and 35 and 27 years have passed since the interim agreements were signed in West Basin and in Central Basin. Thus, the parties to these three agreements have had many occasions to decide whether or not to comply. Given the value of groundwater, the temptation not to comply must have been relatively great for all producers at one time or another in the combined 107 years of water use that have elapsed. However, the level of infractions has been insignificant during that time.

The watermaster in each basin has extensive monitoring and sanctioning authority. Monitoring activities are obvious and public. Every year, each party reports total groundwater extractions and receives a report listing the groundwater extractions of all other parties (or anyone else who has started to pump). The reliability of these records is high. Several agencies cross-check the records. The watermaster is authorized to calibrate all meters, thereby reducing the probability of one form of cheating. Given the accuracy of the information and its ease of access, each pumper knows what everyone else is doing, and each knows that his or her own groundwater extractions will be known by all others. Thus, the information available to the parties closely approximates "common knowledge," so frequently a necessary assumption for solutions to iterated dilemma games (Aumann 1976).

Instead of perceiving itself as an active policing agency, the watermaster service tries to be a neutral, monitoring agency. Because anyone who possesses a legal water right can initiate a court action to enforce compliance to the judgments, the watermaster does not need to initiate punitive actions against nonconformers. As expressed by an official of the watermaster service in 1960,

it is our policy not to take any affirmative actions against any party since this would place us in the position of being an active party in the action. Our policy has been to inform the active parties of any infringements and leave affirmative action up to them. We want to stay as neutral as possible in order to gain as much voluntary cooperation as possible.[30]

In the early years of the West Basin agreement, for example, the Moneta Water Company began to withdraw more than its allocation. After a couple of years, it was obvious that the overextractions were not accidental. In addition to listing Moneta's annual withdrawals in the tabular material included in all reports, the watermaster devoted several pages in an annual report to the recent activities of the company. The company began to comply with the judgment soon after the publication of those facts. Other than a few isolated incidents, handled in the same manner, the original litigants have complied with the curtailments without formal sanctions being imposed. Even the city of Hawthorne has curtailed its withdrawals to the stipulated amounts of the final judgment. It has been necessary, however, to initiate legal action against new pumpers who have attempted to withdraw groundwater without first purchasing water rights. Charges have been filed and defendants enjoined from groundwater production other than under the rights they eventually acquired by purchase.[31]

The levels of quasi-voluntary compliance with the final judgments in all of these court decisions have been extremely high. Although each pumper might be tempted from time to time to withdraw more water than legally allowed, each pumper wants total withdrawals from the basin constrained so that access to the storage and flow values of the resource will be continued over the long run. Given the active, reliable, and neutral monitoring of the watermaster service, no pumper can expect to overextract without everyone else learning about any noncompliance at the end of the next water year. Because everyone is organized and communicating with one another about joint strategies, continued noncompliance is likely to bring legal sanctions, as well as loss of reputation and the application of informal sanctions. Because a pumper is constrained, and almost all pumpers voluntarily agreed to the initial allocation of rights, the basic system is perceived to be fair by most participants. Further, participants continue to have control over the monitoring system to ensure that it continues to be active, fair, and reliable. Two-thirds of the watermaster's budget is paid for by those possessing water rights, and they can petition the court to appoint a different watermaster if they are not satisfied with performance.[32]

THE ENTREPRENEURSHIP GAME

Immediately after the interim agreement was signed in West Basin, and before litigation was initiated in Central Basin, West Basin water producers recognized that litigation was not a sufficient means to achieve long-term regulation of their basin. They took steps that culminated five years later in the creation of a new public enterprise and a series of agreements with surrounding public enterprises to manage West Basin and Central Basin as interconnected basins. The process of problem-solving and negotiation involved in the establishment of this new district and the series of agreements with existing agencies illustrates how public entrepreneurship can be used as a strategy to transform the structure of incentives facing those jointly using a CPR. The process of putting together the necessary components of a new enterprise was immensely complicated. Only a sketch can be presented here, but I try to present the problems the water producers faced, as they saw them, and the steps they took to try to solve them in the political environment they faced.

The litigation had left several unresolved questions. First, producers had been unwilling to cut back production to the safe yield. Although the cutbacks immediately improved water conditions, they were insufficient to achieve a final regulation of the basins. Either the replenishment of both basins had to be accelerated or further cutbacks in production were needed. If the replenishment rate could be increased, then it would be possible to use the underground storage capacity in a manner somewhat analogous to the use of surface storage facilities, whereby one draws down and then refills the facility repeatedly.[33]

A second unresolved, and related, problem was the specific danger that saltwater intrusion posed along the long western border of West Basin and the short southern border of Central Basin. Early in the 1960s, water engineers from West Basin and the Los Angeles County Flood Control District began to experiment with the concept of building a freshwater barrier against the sea. An initial experiment, funded in part by local sources and in part by the state of California, proved that it was both technically and economically feasible to construct a series of wells along the coast that could be used to inject fresh water under pressure into a groundwater basin. The resulting cone of fresh water would prevent further saltwater intrusion. Most of the fresh water would then be available at a later juncture to be withdrawn when needed. If such a barrier could be constructed along the entire coastline, the artificial recharge of the basin would be greatly enhanced, and the threat of the sea would be eliminated. Once the technical and economic feasibility had been established, the

question of exactly who would pay for the barrier, and how, remained to be resolved.

A third delicate question centered on the relevant boundary for managing West Basin and Central Basin. That question had not arisen in regard to Raymond Basin, which was an upland basin and relatively self-contained. Once water producers in West Basin reduced their pumping levels, while water producers in Central Basin continued heavy production, water from West Basin began to flow eastward into Central Basin, instead of westward from Central Basin into West Basin. That change in the direction of the "natural" water flow led producers in both basins to recognize how closely interconnected their two basins were. A barrier erected along the coast would afford protection not only for West Basin but also for Central Basin. Further, an open porous area in Central Basin could be used to replenish far more efficiently than any area located in West Basin. Water spread at that location could raise water levels in Central Basin, which, combined with a cutback in production in Central Basin, would increase the flow of water into West Basin.

No existing public agency had the authority or the appropriate boundaries to address these questions. Water producers in West Basin and Central Basin were reluctant to turn to any of the large-scale agencies currently in existence for fear that they would lose control of the decisions being made and might end up worse off.[34] In the fall of 1954, the president of the West Basin Water Association suggested to the president of the Conservation Association of Southern California that there was a need for representatives from all segments of the California water industry to meet and discuss potential legislation for solving critical groundwater problems. As a result, 45 agencies were invited to meet in September of 1954 "to draft equitable and effective ground water legislation for introduction at the 1955 Legislature" (letter from W. S. Rosecrans to the Central Basin Water Association). From that group, the "Committee of Twelve" was formed to draft legislation.[35]

Discussion of the proposed legislation centered on two types of changes. The first was designed to expedite future groundwater adjudications. The referee in the West Basin case had taken seven years to prepare a report because of lack of information concerning the historical patterns of water use by a large number of the producers. To correct that situation for the future, the committee drafted legislation to require all those who produced at least 25 acre-feet of groundwater per year to file notices of their annual extraction with a state agency. That legislation could not expedite the West Basin case, but it did provide some of the information necessary for relatively rapid adjudications in Central Basin and San Gabriel Basin.

The second area of concern in the proposed legislation involved the authorization of a new type of district empowered to undertake broad replenishment responsibilities financed primarily by a "pump tax" or an assessment on the groundwater production within the boundaries of a district. The legislation as finally drafted was different from what producers in either West Basin or Central Basin might have drafted on their own.[36] By taking into account the needs of other areas and existing water service agencies, West Basin and Central Basin producers were able to gain rapid approval of both pieces of legislation when they were submitted to the state legislature in 1955.

The new Water Replenishment District Act authorized citizens located in southern California to create a new district after they had (1) obtained signatures from at least 10% of the registered voters residing within the boundaries of the proposed district, (2) proposed specific limits on the taxing power of the new district, (3) received agreement from the Department of Water Resources that the area included within the boundaries of the district would be benefited by inclusion, and (4) received a majority of positive votes in a special election held to consider the creation of the new district. A district, once created, was given a wide diversity of powers to raise revenue through a pump tax and, to a limited extent, through a property tax and to undertake actions to replenish a groundwater basin. The organic legislation included a unique provision intended "to avoid duplication of similar operations by existing agencies and replenishment district" (*California Water Code*, sec. 60231). The provision stated that

> in the event an existing agency has facilities available and adequate to accomplish any part of the purposes of a district . . . the district shall investigate and determine the cost of contracting for the accomplishment of such purposes through such existing agency. (*California Water Code*, sec. 60231)

In other words, a new replenishment agency would be expected to investigate the costs of contracting to have services provided, rather than immediately creating its own production staff for any activity it wanted to undertake.[37] The legislation provided a general "constitution" for a new district. Water producers in any specific area could then use that general framework to create a particular "constitution" for their own district. At first, West Basin producers presumed that they would go it alone and created a working committee within the association to draft a specific proposal to create a district.

The basic issues that had to be resolved in that constitutional process were (1) the source of water for the barrier, (2) the exact boundaries of the new district, (3) how the internal electoral boundaries would be drawn,

and (4) the extent of taxing powers to be authorized. West Basin water producers had hoped that it would be possible to purchase reclaimed water from the Hyperion water-treatment plant – a sewage-disposal facility operated by the city of Los Angeles – located on the coast. If that had proved technically feasible, they would have had a source of low-cost water and would not have had to negotiate with the powerful Metropolitan Water District of Southern California (MWD). After considerable experimentation, it proved technically infeasible to use the reclaimed water, at least in the short run. The committee had to start over again. Members of both associations were appointed "to approach the Metropolitan Water District . . . to see whether a firm commitment of a sufficient quantity of water could be obtained . . and to request a certain amount of engineering and costs estimates which would be needed" (report by Allan Harris, West Basin Water Association, minutes, March 22, 1956, p. 6).

Once West Basin producers realized that they would have to use MWD water for the barrier, they began more intensive talks with their Central Basin neighbors concerning the creation of one large district to include both basins. When the two associations first started discussions, the differences between the two basins were quite apparent. West Basin was smaller in area, population, and assessed valuation. In a joint district, it could be dominated by Central Basin. In 1955, West Basin producers had signed the interim agreement and had limited their production to 60,000 acre-feet per year, whereas Central Basin producers were pumping 110,000 acre-feet and still increasing their annual rates of withdrawal. At a meeting of the West Basin Water Association, the chairman of the committee devising the proposal outlined the reasons for and against forming a district to include both basins.

Reasons for forming a district to include both basins

1 The purpose would be the same in both basins: replenishment of the groundwater supply.
2 Greater financial resources would be available; hence, the tax rate and amount of pumping assessment could be lower.
3 A large district would have greater political strength and would be more effective in dealing with the Upper San Gabriel Valley District and various state bodies.
4 The Long Beach harbor area offers a potential route for intrusion of seawater into West Basin and probably would be included. It is doubtful that any of Long Beach could be included in a district comprising only the West Basin.

5 The flow of groundwater across the fault from Central Basin to West Basin probably would be greater under the replenishment program of a larger district.
6 Administration costs would be less in a larger district.

Reasons against forming a district to include both basins

1 The injection of replenishment water would be unique and necessary to West Basin. Central Basin would control that program in the West Basin if a large district were formed and might not want to continue the well-injection method along the coast.
2 Pumping was curtailed in West Basin, but not in Central Basin.
3 The degrees of ultimate curtailment might not be the same in the two basins.
4 Control of the local tax rate and amount of pumping assessment would be relinquished by West Basin.
5 A local district could initiate proceedings to ensure financial replenishment from Central Basin.[38]
6 Extensive recharge of Central Basin might contribute free water to West Basin (West Basin Water Association, minutes, November 17, 1955, pp. 9–10).

The West Basin water producers were physically disadvantaged because they were at the end of the groundwater "pipeline." They were concerned that their physical disadvantage could be exaggerated by the creation of a new public agency in which they would be politically dominated. The hopes and fears of West Basin producers were summarized in a letter written by the chairman of the West Basin committee to his committee:

> In the event a water replenishment district to include both Basins is decided upon, it appears desirable that a statement of policy morally binding on the new district board of directors should be adopted. The policy should provide assurance of an effective salt water barrier program for West Basin, a curtailment of pumping in Central Basin to insure continued ground water flow into West Basin, and an arrangement of the five divisions of the new district so that territory of both Basins would be included in each such division to prevent West Basin versus Central Basin representation on the board of directors.
>
> (letter from R. R. Thorburn to the Replenishment District Boundary Committee, October 27, 1955, p. 2)

Soon thereafter, members of both associations came to a working agreement that the benefits of a larger district would outweigh the costs. Assurances were given to West Basin producers that they would not be

dominated by their eastern neighbors. Next, a joint committee of both associations began a series of relatively tough negotiations with all of the public agencies that might eventually be involved in managing these two basins.[39]

The results of those negotiations were formalized in a seven-page proposal that the committee submitted to the two associations for approval in August of 1958. The proposal set forth the essential factors for constituting the new enterprise. The statement proposed that a new replenishment district would be formed to (1) repel saltwater intrusion, (2) recharge the groundwater basins, and (3) reduce pumping in the basins to safe limits (West Basin Water Association and Central Basin Water Association, "Proposal Submitted by the Joint Committee on Water Replenishment District," mimeograph, July 30, 1958, p. 1). The proposal stated that to accomplish those purposes, "the district will have responsibility for financing the purchase of water used in halting the intrusion of sea water and in replenishing the groundwater supply" (ibid., p. 3).

The proposal then clarified the future relationships of the replenishment district to all of the existing agencies that might consider the replenishment district to be a potential competitor. It then outlined the amount of water that would be purchased from MWD and spread or injected by the Los Angeles County Flood Control District. The proposal stated that the new district "would have no authority to purchase replenishment water with *ad valorem* tax derived funds, and the petition for the formation of the district will clearly set forth this limitation on its taxing power" (ibid., p. 6). It was estimated that a levy of $6.00 per acre-foot would be necessary to raise the necessary funds to purchase 165,000 acre-feet of water from MWD – an amount equal to the average annual overdraft. The proposal concluded by stating that the new district would be "an administrative agency operated by a five-member board of directors with a minimum staff" (ibid., p. 7).

The proposal was in effect a "constitution" for a multiple-agency management system to operate a coordinated program. Constitutional documents do not need to carry the formal name "constitution" to serve the purpose of determining the decision rules to be used for making future collective choices about some specified physical domain. The proposal was attached to the formal petitions presented to the Los Angeles County Board of Supervisors (to gain approval for the special election) and to the California Department of Water Resources (to gain approval for the boundaries), and in that way had formal recognition as a type of constitutional document. Once that constitution had been approved by the two private associations, all formal steps outlined within it were achieved within a few months, and the Central and West Basin Water Replenishment District was

supported in the election of November of 1959 by a vote of 4 to 1 (*Los Angeles Times*, November 18, 1959).

THE POLYCENTRIC PUBLIC-ENTERPRISE GAME

The creation of the Central and West Basin Water Replenishment District in 1959 dramatically transformed the structure of incentives facing water producers and their representatives. It was an enterprise created by the water producers (and approved by the citizens living in the area), with public powers to tax, to sue, and to engage in the provision of collective goods. Whereas the replenishment district took over the *active* role of managing West Basin and Central Basin, the two private water associations continued to have strong input into all policy decisions.

Further, the replenishment district is only one public enterprise among a half dozen agencies that are actively involved in the management program. Thus, instead of one central governmental authority, a polycentric public-enterprise system has emerged to achieve a very sophisticated management system. This polycentric system has restored water levels throughout both basins, has completed a freshwater barrier along the exposed coasts of both basins, and is now engaged in focused efforts to eliminate pumping troughs and other physical impediments that inhibit the effective use of the basins in conjunction with a surface supply.

The overall costs of this system are quite low.[40] In Table 4.1, the amortized and annual costs (in constant dollars) of the management systems in these basins, as computed by Blomquist, are contrasted with the amortized costs of replacing the basins with surface storage. Total costs are substantially lower in each basin than they would be if the basins had been destroyed. Total costs would, of course, be lower if water producers had been able to negotiate a settlement of their water rights at an earlier juncture and had not had to pay the high costs of prolonged negotiations. The water producers of Central Basin, however, learned from the experiences of their colleagues in Raymond Basin and West Basin and thus were able to achieve a settlement at lower costs.

In this discussion I have tried to focus more on the origins of these institutions than on their current operations, because it is so difficult to find documentation about the origins of institutions. I do think it is important, however, to describe briefly the types of polycentric relationships that exist among the public enterprises that currently manage West Basin and Central Basin.

At the core of each of those relationships is the Central and West Basin Water Replenishment District. This district receives the funds assessed on

Table 4.1. *Basin management costs and savings per acre-foot resulting from basin management in the three basins (dollars)*

Cost	Raymond	West	Central
Basin management cost per acre-foot of ground-water extraction, 1985	3.50	77.40	73.77
Average cost of an acre-foot of water with basin management	184.65	235.71	224.85
Estimated cost of an acre-foot of water if all groundwater were replaced by imported water	748.68	739.30	739.94

Source: Adapted from Blomquist (1987a, Figure 9).

all water pumped in the district and thus has the power to take collective action for both basins. To get the water into the basins, however, the replenishment district must relate to several other public districts. Until the late 1960s, the replenishment district depended on a monopoly supplier of water, the MWD, for its replenishment water. In 1966, MWD unilaterally announced a change in its pricing structure that would substantially increase the cost of replenishment water. The replenishsment district and both associations bargained hard, but unsuccessfully, for a reconsideration. The replenishment district then opened negotiations with the Los Angeles County Sanitation Districts to obtain a reliable supply of water at lower cost from a specially constructed reclamation plant.[41] Opening this alternative source of water supply has meant that the replenishment district has assured itself of a continuing supply, and at a cost well below that of imported water. In 1987, for example, the district was seeking approval from the relevant regulatory agencies to increase its purchase of reclaimed water from 30,000 acre-feet per year to 50,000 acre-feet per year, at an average cost of $8.00 per acre-foot, as compared with the $153 it has to pay MWD for replenishment water (Central and West Basin Water Replenishment District 1987, pp. 44–56).

In regard to the actual operation of the replenishment works, the replenishment district entered into an exchange agreement with the Los Angeles County Flood Control District (reorganized in 1987 to be the Los Angeles County Department of Public Works). Thus, the replenishment