

The background of the cover is a medieval-style painting. It depicts a town square or street scene. In the foreground, several people are on horseback, moving from left to right. They are dressed in rich, colorful medieval attire. Behind them, a group of people is walking. The background is filled with multi-story buildings with various architectural features, including balconies, arched windows, and a prominent dome on the left. The overall style is reminiscent of a medieval manuscript illumination or a painting from the 15th or 16th century.

Microeconomics  
Behavior, Institutions  
and Evolution

SAMUEL BOWLES

---

Microeconomics

---

### The Roundtable Series in Behavioral Economics

The Roundtable Series in Behavioral Economics aims to advance research in the new interdisciplinary field of behavioral economics. Behavioral economics uses facts, models, and methods from neighboring sciences to establish descriptively accurate findings about human cognitive ability and social interaction and to explore the implications of these findings for economic behavior. The most fertile neighboring science in recent decades has been psychology, but sociology, anthropology, biology, and other fields can usefully influence economics as well. The Roundtable Series publishes books in economics that are deeply rooted in empirical findings or methods from one or more neighboring sciences and advance economics on its own terms—generating theoretical insights, making more accurate predictions of field phenomena, and suggesting better policy.

#### Colin Camerer and Ernst Fehr, editors

*Behavioral Game Theory: Experiments in Strategic Interaction* by Colin F. Camerer

*Microeconomics: Behavior, Institutions, and Evolution* by Samuel Bowles

*Advances in Behavioral Economics*, edited by Colin F. Camerer, George Loewenstein, and Matthew Rabin

#### The Behavioral Economics Roundtable

Henry Aaron  
George Akerlof  
Linda Babcock  
Colin Camerer  
Peter Diamond  
Jon Elster  
Ernst Fehr  
Daniel Kahneman  
David Laibson

George Loewenstein  
Sendhil Mullainathan  
Matthew Rabin  
Thomas Schelling  
Eldar Shafir  
Robert Shiller  
Cass Sunstein  
Richard Thaler  
Richard Zeckhauser

---

# Microeconomics

BEHAVIOR, INSTITUTIONS, AND EVOLUTION

*Samuel Bowles*

RUSSELL SAGE FOUNDATION  
NEW YORK  
PRINCETON UNIVERSITY PRESS  
PRINCETON AND OXFORD

Copyright © 2004 by Russell Sage Foundation

Requests for permission to reproduce materials from this work should be sent to  
Permissions, Princeton University Press

Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540  
In the United Kingdom: Princeton University Press, 3 Market Place, Woodstock,  
Oxfordshire OX20 1SY  
And the Russell Sage Foundation, 112 East 64th Street, New York, New York 10021  
All Rights Reserved

Library of Congress Cataloging-in-Publication Data

Bowles, Samuel.

Microeconomics : behavior, institutions, and evolution / Samuel Bowles.

p. cm. — (The roundtable series in behavioral economics)

Includes bibliographical references and index.

ISBN 0-691-09163-3 (alk. paper)

1. Microeconomics. 2. Institutional economics. 3. Evolutionary economics.

I. Title. II. Series.

HB172.B67 2003

338.5—dc21 2003049841

British Library Cataloging-in-Publication Data is available

This book has been composed in Sabon

Printed on acid-free paper. ∞

[www.pupress.princeton.edu](http://www.pupress.princeton.edu)

[www.russellsage.org](http://www.russellsage.org)

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

List of Credits

Quotation from “Mending Wall” by Robert Frost used with permission of Henry Holt  
and Company.

Quotation from “16 tons” by Merle Travis used with permission of Warner-Chappell  
Music, a division of Warner Brothers.

Map of Italy in the fifteenth century (Figure 13.1) adapted from *Atlas of Medieval Europe*  
by Donald Matthew, used with permission of Andromeda, Oxford, Ltd.

Chapter 13 makes substantial use of work co-authored with Jung-Kyoo Choi and Astrid  
Hopfensitz that appeared in the *Journal of Theoretical Biology* (2003) 23:2, pp. 135–  
47, and is used here with permission from Elsevier.

---

*For Libby and for Herb*



## Contents

---

<i>Preface</i>	ix
<i>Prologue: Economics and the Wealth of Nations and People</i>	1
<b>PART I: Coordination and Conflict: Generic Social Interactions</b>	<b>21</b>
CHAPTER ONE Social Interactions and Institutional Design	23
CHAPTER TWO Spontaneous Order: The Self-organization of Economic Life	56
CHAPTER THREE Preferences and Behavior	93
CHAPTER FOUR Coordination Failures and Institutional Responses	127
CHAPTER FIVE Dividing the Gains to Cooperation: Bargaining and Rent Seeking	167
<b>PART II: Competition and Cooperation: The Institutions of Capitalism</b>	<b>203</b>
CHAPTER SIX Utopian Capitalism: Decentralized Coordination	205
CHAPTER SEVEN Exchange: Contracts, Norms, and Power	233
CHAPTER EIGHT Employment, Unemployment, and Wages	267
CHAPTER NINE Credit Markets, Wealth Constraints, and Allocative Inefficiency	299
CHAPTER TEN The Institutions of a Capitalist Economy	331
<b>PART III: Change: The Coevolution of Institutions and Preferences</b>	<b>363</b>
CHAPTER ELEVEN Institutional and Individual Evolution	365

CHAPTER TWELVE	
Chance, Collective Action, and Institutional Innovation	402
CHAPTER THIRTEEN	
The Coevolution of Institutions and Preferences	437
PART IV: Conclusion	471
CHAPTER FOURTEEN	
Economic Governance: Markets, States, and Communities	473
<i>Problem Sets</i>	502
<i>Additional Readings</i>	529
<i>Works Cited</i>	537
<i>Index</i>	571

## Economic Governance: Markets, States, and Communities

Which of these systems [*central planning or competition*] is likely to be more efficient depends on the question under which of them can we expect that fuller use will be made of the existing knowledge. And this, in turn, depends on whether we are more likely to succeed in putting at the disposal of a single central authority all the knowledge which ought to be used but which is initially dispersed among many different individuals, or in conveying to the individuals such additional information as they need in order to enable them to fit their plans in with those of others.

—F. A. Hayek, “The Use of Knowledge in Society” (1945)

Lawgivers make the citizen good by inculcating habits in them, and this is the aim of every lawgiver; If he does not succeed in doing that, his legislation is a failure. It is in this that a good constitution differs from a bad one.

—Aristotle, *Nicomachean Ethics* (350 B.C)

The man of systems . . . imagines that he can arrange the different members of a great society with as much ease as the hand arranges the different pieces upon a chess-board; he does not consider . . . that in the great chess-board of human society, every single piece has a principle of motion of its own.

—Adam Smith, *Theory of Moral Sentiments* (1759)

IN SOME Chicago neighborhoods, adults admonish youngsters’ skipping school, creating a disturbance, or decorating walls with graffiti. Residents are also willing to intervene in public meetings to maintain neighborhood amenities such as a local firehouse threatened with budget cuts. These are all examples of what Sampson, Raudenbush, and Earls (1997) term *collective efficacy*. Where neighbors express a high level of collective efficacy, violent crime is markedly lower, controlling for a wide range of community and individual characteristics, including past

The first epigraph is from Hayek (1945:521), the second from Aristotle (1962:103), the third from Smith (1976:234).

crime rates. In other neighborhoods, residents adopt a more hands-off approach. Sampson, Raudenbush, and Earls found considerable variation in the neighborhood levels of collective efficacy, with examples of rich and poor, and black and white neighborhoods exhibiting both high and low levels. Remarkably, ethnic heterogeneity was considerably less important in predicting low collective efficacy than were measures of economic disadvantage, a low rate of home ownership, and other indicators of residential instability. Chicago's neighborhoods illustrate the informal enforcement of community norms.

The Toyama Bay fishing cooperatives in Japan illustrate another aspect of community problem solving (Platteau and Seki 2001). Faced with variable catches as well as the high level and changing nature of skills required, some fishermen elect to share income, information, and training. One coop, which has been highly successful since its formation in the mid 1960s, consists of the crews and skippers of seven shrimp boats. The boats share income and costs, repair damaged nets in common, and pool information about the changing location and availability of shrimp. Elder members pass on their skills, and the more educated younger members teach others the new electronic methods of locating fish and navigating. The coop's income- and cost-pooling activities allow its boats to fish in much riskier and higher yield locations, and the skill and information sharing raises profits and reduces productivity differences among the boats. Fishing, off-loading the catch, and marketing by individual boats are synchronized to increase the transparency of the sharing process and make opportunistic cheating on the agreement easy to detect.

The success of the Toyama Bay shrimp cooperatives and of collective efficacy in Chicago neighborhoods are examples of *community governance*. The worker-owned plywood coops described in chapter 10 are another example. By community I mean a group of people who interact directly, frequently, and in multi-faceted ways. People who work together are usually communities in this sense, as are some neighborhoods, groups of friends, professional and business networks, gangs, and sports leagues. Connection, not affection, is the defining characteristic of a community. The evolutionary models in chapter 7 showed how the repeated and multi-faceted nature of social interactions in communities, the relatively small numbers of people involved, and, as a result, the availability of information on one's associates may support high levels of what is sometimes referred to as social capital: trust, concern for one's associates, and a willingness to live by the norms of one's community and to punish those who do not.

These other-regarding behaviors were recognized as essential ingredients of good governance among classical thinkers from Aristotle to

Thomas Aquinas, Jean-Jacques Rousseau, and Edmund Burke. Nicolo Machiavelli's *The Prince* (1513) and Thomas Hobbes's *Leviathan* (1651) represented a sharp break with the Aristotelian tradition. These founding works of modern political philosophy took self-interest as a fundamental behavioral assumption and asked how the potentially destructive consequences of the autonomous pursuit of individual gain might be constrained by the authority of a sovereign ruler.

The more radical notion that selfish motives could be harnessed for public good was the key contribution of Bernard Mandeville's *Fable of the Bees*, first published in 1705 (Mandeville 1924). The subtitle of the 1714 edition of the *Fable* announced that the work contained "several discourses to demonstrate that human frailties . . . may be turn'd to the advantage of civil society, and made to supply the place of moral virtues." In place of the Aristotelian view that good laws make good citizens, Mandeville proposed the more modern notion that the right rules of the game governing social interactions might harness selfish motives to promote general well-being. This radical conjecture was given economic content by Adam Smith's invisible hand argument. Thus, most political theorists and constitutional thinkers since the late eighteenth century have taken the self-interested *Homo economicus* as their fundamental assumption about behavior, and partly for this reason, have stressed competitive markets, well-defined property rights, and efficient, well-intentioned states as the critical ingredients of governance. Good rules of the game thus came to displace good citizens as the *sine qua non* of good government.

The contending camps that emerged in the nineteenth and early twentieth centuries advocated laissez faire on the one hand or comprehensive state intervention on the other as *the* ideal form of economic governance.<sup>1</sup> The debate in the 1920s and 1930s on the economic feasibility of centralized planning was emblematic of the truncation of the constitutional menu to state versus laissez faire. Ludwig von Mises and others (Hayek 1935) advanced the view that the rational economic calculation entailed by planning required the knowledge of prices reflecting true scarcity (i.e., measuring social marginal costs and benefits), and that this information could be obtained only by the extensive use of decentralized allocation through markets. Oskar Lange (Lange and Taylor 1938), Enrico Barone (1935), Abba Lerner (1944), and others countered that prices are implicit in any optimizing problem (whether or not markets

<sup>1</sup> Outside of academic circles, the menu of options was considerably broader, including the "mixed" economy models pioneered by Nordic social democrats and market socialist models initiated by Oskar Lange. Dahl and Lindblom (1953) is exemplary, but rare, in avoiding the polarization of the planning versus markets debate.

exist). These implicit (or shadow) prices, they claimed, could either be computed directly or extracted from observations of competitive behaviors in an economy using markets to implement the allocations determined by central planners. This being the case, the planner could implement any allocation achieved by decentralized competitive markets but could do better than this in cases in which missing markets or impediments to competition gave rise to allocational inefficiencies.

By the 1940s the debate was all but over. Even the arch-opponent of socialism, Joseph Schumpeter, had conceded: “Can socialism work? Of course it can. . . . There is nothing wrong with the pure theory of socialism” (Schumpeter 1942:167, 172). He was echoing another opponent of socialism, Vilfredo Pareto (1896), who much earlier had affirmed the feasibility of rational economic calculation in what he termed a “collectivist regime.” In a section of his famous *Manuel d’Economie Politique* labeled “An argument in favor of collectivist production,” Pareto (1909:364) had concluded that “pure economics does not give us a truly decisive criterion for choosing between the organization of society based on private property and a socialist organization.”

What then *was* wrong with socialism? And what was wrong with the economic theory that so inadequately captured the economic shortcomings of centralized allocations and vindicated socialist planning in the debate?

A striking feature of the debate had been that both sides deployed the Walrasian model on behalf of their arguments. Hayek soon appreciated the error. His “The Uses of Information in Society” (quoted above) reframed the debate in terms of the costs and limited availability of information, concepts absent from the Walrasian paradigm. The problem with socialism, according to Hayek, was that the information needed by the planner is privately held by millions of economic actors, and they have neither the will nor in many cases even the way to transfer it to a central authority. By contrast, Hayek continued, decentralized markets make effective use of dispersed information, each actor knowing his own preferences and responding to a price vector that under ideal circumstances is both known to the individual actor and reflects the real social scarcities of the goods in question. We now know (chapter 6) that there is no even remotely realistic model of market competition for which these ideal conditions hold, in part because many of the relevant prices simply do not exist, others do not reflect social scarcities, and still others (prices of future goods for example) are unknowable. But by focusing attention on which institutions more effectively utilize the information that *is* available, Hayek’s paper, like Mandeville’s *Fable*, counts as a landmark work in the theory of economic institutions.

In formalizing a major shortcoming of centralized planning, Hayek

also pointed to the deficiencies of the Walrasian paradigm, namely, the assumption of complete information. To Ronald Coase, the debate had revealed an inconsistency, one that was to prompt him to study the theory of the firm. At the beginning of his career, he recalls wondering:

How did one reconcile the views expressed by economists on the role of the pricing system and the impossibility of successful central economic planning with the existence . . . of these apparently planned societies, firms, operating within our own society. (Coase 1992:715)

Shortly after the fall of Communism, Stiglitz (1994:10) wryly observed that “if the neoclassical model of the economy were correct, market socialism would have been a success [and] centrally planned socialism would have run into far fewer problems.” Long before either neoclassical economics or Communism, John Stuart Mill (1976) had provided a critique of the problems of a hypothetical socialist economy—worker motivation, reduced innovation, lack of appropriate property rights—far more searching than any produced within the neoclassical paradigm. Commenting on the role of U.S. economic advisors in the ex-Communist transition economies of the 1990s, Coase remarked: “Without the appropriate institutions no market economy of any significance is possible. If we knew more about our own economy, we would be in a better position to advise them”(Coase 1992:714).

Along with its failure to illuminate the problems of underdevelopment, the inability of Walrasian economics to understand either the economic disabilities of Communism or the appropriate institutions for a transition to a market-based economy is a striking indictment of the approach.

In this chapter, I adopt a post-Walrasian approach to address the contemporary challenges of economic governance. I use results from previous chapters to explore the ways that markets, states, and communities jointly may provide solutions to the coordination problems studied in the previous chapters. (I do not evaluate these solutions from the standpoint of distributive justice but rather focus on their implications for allocative efficiency.) I single out three generic governance structures—communities, states, and markets—for the distinctive ways that they coordinate joint activities and allocate claims on goods and services.<sup>2</sup> Unavoidably, the treatment will be suggestive, not exhaustive. A first task, suggested by the inappropriate use of Walrasian assumptions in the planning versus *laissez faire* debates, is to pull together the strands

<sup>2</sup> The family might be considered a fourth governance structure. Families share many of the characteristics of communities but differ in that roles are assigned by age, sex, and kinship.

of the post-Walrasian perspective. I do this in the next section. I then identify the distinctive capacities and shortcomings of markets, states, and communities. I conclude with a reconsideration of Mandeville's radical conjecture.

#### ECONOMICS AND EVOLUTIONARY SOCIAL SCIENCE

The Walrasian paradigm provides the only fully worked out, economy-wide model of the way that the actions of large numbers of autonomous actors support aggregate social outcomes. Some of the shortcomings of this model have been identified in the previous pages, and some alternative formulations have been suggested. To synthesize the main features of the Walrasian approach, I will characterize the Walrasian *paradigm* by what its students are *taught* rather than by the impossibly heterogeneous union of the distinct contributions by scholars representative of this paradigm. This necessarily will involve some discrepancies between the representation of the paradigm and the state-of-the-art knowledge in the field. For an example, consider the uniqueness and stability of general equilibrium: students are regularly taught as if this were true—consider the standard supply and demand diagram—even though (as was pointed out in chapter 6) the assumptions required to demonstrate either uniqueness or stability are exceptionally restrictive. I use the term *evolutionary social science* to refer to the alternatives to the characteristic Walrasian paradigm. There is no unified paradigm of this name, but rather a disjointed set of approaches, many of which are rather rudimentary and most of which have been introduced in the previous pages. Whether in the years to come these approaches will be unified in a coherent replacement for the Walrasian paradigm remains to be seen. (The hunch on which this book is based is that they will.)

Table 14.1 summarizes the contrasting approaches. It would be redundant to comment on each row. But the last line in the table, concerning *reductionism* and *methodological individualism* is worth comment. Reductionism is an approach to science that prefers explanations based on lower-level entities (cells, for example) rather than simply positing the higher-level entities that they make up (multi-cellular organisms, for example). Methodological individualism is an expression of reductionism in social science that insists that explanations of group-level phenomena such as institutions or aggregate output must be built up from the actions of individuals. The approach taken in this book is consistent with methodological individualism in that it has focused on the causal mechanisms connecting what individuals do to aggregate social outcomes. But, as the discussion of endogenous preferences and

TABLE 14.1  
The Walrasian paradigm and some alternative

	<i>Walrasian economics (as taught)</i>	<i>Evolutionary social science (in prospect)</i>
<i>Social interactions</i>	Complete and enforceable claims exchanged on competitive markets	Direct (noncontractual) relationships in noncompetitive settings are common
<i>Technology</i>	Exogenous production functions with non-increasing returns	Generalized increasing returns in both (endogenous) technology and social interactions (positive feedbacks)
<i>Updating</i>	Forward-looking individuals instantaneously update based on knowledge of entire system	Backward-looking (experienced-based) individuals update using local information
<i>Outcomes</i>	A unique stable equilibrium based on stationarity of individual actions	Many equilibria; aggregate outcomes may be long term averages of non-stationary lower level entities
<i>Time</i>	Comparative statics	Explicit dynamics
<i>Chance</i>	Relevant only to risk-taking and insurance	Essential component of evolutionary dynamics
<i>Domain</i>	The economy as a self-contained self-regulating entity: exogenous preferences and institutions	The economy as embedded in a larger social and ecological system: coevolving preferences and institutions
<i>Preferences</i>	Self-regarding preferences, defined over outcomes	Self- and other-regarding preferences, defined over outcomes and processes
<i>Prices and quantities</i>	Prices allocate resources; actors are not quantity constrained	Quantity constraints; wealth-dependent contractual opportunities
<i>Method</i>	Reductionist (methodological individualism)	Non-reductionist; selection on individual and higher order entities

cultural evolution has made clear, the effect of aggregate outcomes on individuals is no less important.

The conventional concept of equilibrium in economics expresses the methodological individualism of the discipline. It is standard practice—and one that has been frequently used in the pages above—to define an equilibrium as a state such that none of the individuals involved has a

reason to alter his behavior. The aggregate properties of the equilibrium — an economy-wide allocation of resources, for example — are then derived by aggregation of the equilibrium individual behaviors. The aggregate properties are stationary because the individual behaviors are stationary. But as the model of general equilibrium by Foley (1994), described in chapter 6, demonstrates, stationarity of aggregate properties does not require stationarity of lower-level entities. Foley's model shows that for average prices not to change, it is not necessary that trade cease. The analysis of attendance at his favorite bar in Santa Fe by Brian Arthur (1994a) conveys a similar message. Nobody wants to go there when it's too crowded, and people estimate how many will attend based on past experiences. Arthur simulates an adaptive learning process, the result of which is that about sixty people show up at the El Farol each Thursday. But this does not require that the same people show up, or that the beliefs of those showing up about how many others will show up are accurate or stationary.

In this and many other applications, lower-level entities are nonstationary in ways that average out, producing no change in the aggregate property. The evolutionary analysis in chapters 11 through 13 adopted this method. Stochastically stable states (chapter 12) are not stationary outcomes; rather, they describe long-term average behavior of a system. The models in chapters 11 and 13 described populations constantly in motion, propelled by deliberate nonbest-response actions by collectives of individuals, other idiosyncratic behavior, genetic drift, and institutional innovation. The results of the agent-based simulations were long-term averages reflecting all of these influences.

Methodological individualism is also evident in a common approach to the analysis of economic institutions. Schotter (1981:20) provides an example:

If economics . . . is going to study the rise and evolution of social institutions, a very simple methodological approach is suggested. We should start our analysis in a Lockean state of nature in which there are no social institutions at all, only agents, their preferences, and the technology they have at their disposal. . . . The next step would be to study when, during the evolution of this economy, such institutions as money, banks, property rights, competitive markets, insurance contracts, and the state would evolve.

There is no question that Schotter's method is interesting, and that it has proven insightful. But if instead one takes technologies and preferences as endogenous, it would be equally insightful to violate the methodological individualist precepts. One could, for example, posit a set of institutions and then ask what kind of preferences and technologies would evolve. The approach adopted here (in chapters 11 through 13,

especially) represents individual preferences and group level institutions as coevolving, *thereby not privileging either the lower- or higher-order entities.*

Whether the group or individual or both (or more) processes need to be modeled depends on the analytical problem at hand and practical considerations of tractability. For most social science applications nothing is lost and much simplicity gained by not modeling the cellular interactions within individuals. But this would be a poor strategy for understanding cancer. Where group characteristics can be taken as given, modeling at the individual level is a reasonable approach. Correspondingly, if we can abstract from within-group variation, the group-level selection process can be the focus of attention, as it is in models of competition among firms. Richard Dawkins (1989b:3), a strong proponent of reductionism in biology, rightly observed that its usually more informative to explain cars in terms of carburetors than quarks.

From this perspective, positing a pristine institution-free environment is a curious way to investigate the historical evolution of real institutions. The reason is that since the advent of biologically modern humans, and even among other primates, social conventions and property rights of various kinds have almost certainly provided an institutional environment for our interactions. Locke, Hobbes, and other philosophers used the state of nature as a hypothetical inquiry into what might *justify* property, the authority of the state, and the like, not as part of an explanation of how these institutions evolved historically. (Recall Hobbes' deliberately fanciful metaphor of the state of nature in the epigraph to chapter 3: people were "sprung out of the earth . . . like mushrooms.")

In the prologue I distinguished the evolutionary method from a social engineering approach to public policy. By the latter I mean the view that social outcomes are determined by the autonomous actions of public-spirited officials, more or less as the chessmen in the Smith epigraph might be moved around the board. Nobody believes this literally (least of all public officials), but many fail to appreciate the extent to which this view misrepresents the process by which outcomes are determined. While I have given no attention to questions of public policy, the models developed here suggest a quite different approach. This is to apply the same behavioral assumptions to state officials as we routinely do to those engaged in private exchange, namely, that their actions are best responses based on their preferences not subject to complete contracting. Adopting an early version of this approach, Jeremy Bentham advocated constitutional arrangements which would structure incentives so that public servants' "duties" would coincide with their "interests." But this objective can rarely be met.

In the evolutionary view, aggregate outcomes are the result of the

interactions of the public officials' actions and the best responses of all the individuals involved. This does not suggest that governmental interventions are ineffective, but rather that to be effective in the intended ways requires an understanding of the dynamical system in which one is intervening. For example, the policies required to displace a socially undesirable equilibrium in favor of some other outcome may be entirely different if the system producing the outcomes is characterized by a single equilibrium or if there are many stable equilibria, and the job of public policy is to displace one equilibrium in favor of another. The example of child labor below illustrates this.

A final comment does not concern the contrasting paradigms directly but rather is directed to the normative concerns that are never absent when discussing institutional alternatives. "Utility" is a heavily freighted term: economists commonly use it to refer to motives, behaviors, and well-being. The convenience of collapsing these three distinct usages into a single term is considerable. But it requires the implicit assumption of *substantive rationality*, namely, that people act so as to get what they want, which in turn contributes to their well-being as gauged by some independent evaluation of the relevant outcomes. By contrast, the *formal rationality* explicitly assumed by most economists imposes only consistency requirements (such as transitivity) on behaviors, without any requirements for the individual's hedonistic or other subjective reasons for acting, the reasonableness of the means adopted in pursuit of some outcome, or the consequences for the individual's well being. A consistent masochist is not irrational.

To be of practical or moral relevance, economic reasoning about institutions and policies requires the substantive concept of rationality. If, for example, one believes that third parties should not intervene in transactions voluntarily engaged in by adult economic actors, it is not sufficient to know that they have complete and transitive preferences. We also must have confidence that their choices will not be grossly or irreversibly destructive of their well-being. The same is true of the common interpretation of Pareto efficiency in terms of the "well-being" of individuals. Formal rationality alone does not provide the motivation for preferring Pareto-superior outcomes, except in the minds of extreme libertarians. An allocation preferred by two masochists might not be endorsed by others.

But the assumption of substantive rationality is based on strong empirical claims about why people do what they do, and on the consequences of their actions. These claims are generally false. Extensive empirical evidence suggests that by the standard of well-being, people are bad choosers. We are myopic, fail to predict the preferences we will have when the relevant consequences of our actions take place, do not

accumulate accurate information about the hedonic aspects of past experiences, act inconsistently in intertemporal choice situations, and commonly violate the expected utility hypothesis (Kahneman 1994, Camerer 2000). The subjects in experiments and real world settings showing this would find it strange to hear their behaviors termed irrational. They include students at the most selective universities, Harvard professors, and New York City cab drivers.

If preferences are to explain behaviors, they cannot unassisted also do the work of evaluating outcomes. This is true because some common reasons for behavior—weakness of will and addiction, for example—induce behaviors that few would condone. The disjuncture between the reasons for behaviors and the standards by which a liberal and democratic polity should evaluate outcomes raises profound challenges, ones that are sure to pit liberal against utilitarian and paternalistic values. For example, if loss aversion is a powerful subjective reaction among most people, should it be taken into account in evaluating public policies? Doing so would affect a substantial shift in favor of the status quo, as the costs borne by losers would now be at double counted or more. But addressing these questions would take us far afield.

#### MARKETS AND STATES: A POST-WALRASIAN COMPARISON

Given that the rhetoric of the debate on planning versus *laissez faire* was highly polarized, a remarkable conclusion was that markets and states are difficult to distinguish from an allocative standpoint. F. M. Taylor's 1928 Presidential Address to the American Economic Association opened with:

In the case of a socialist state, the proper method of determining what commodities shall be produced would be in outline the same as . . . [u]nder the present economic order of free competitive enterprise. (Taylor 1929:1)

This unexpected similarity in systems of allocations results from the complete information and complete contracting assumptions of most of the participants in the debate. If everyone knew the same things (and what they knew were admissible in court), and if there were no other impediments to contracting, institutional differences would matter less. You have already encountered Samuelson's affirmation (in the epigraph to chapter 10) of the *Walrasian equivalence of worker-run and capitalist-run firms*: if contracting is complete it indeed does not matter who hires whom.<sup>3</sup> This equivalence means that to understand the operational

<sup>3</sup> This equivalence was shown formally by Sertel (1982), Fehr (1993), and Dow (1996).

differences between conventional and worker-owned firms like the plywood cooperatives mentioned in chapter 10, one must analyze the differing problems of contractual incompleteness that they encounter, and their differing capacities to surmount them. The same conclusion holds for comparisons of markets and states. As a result, the relevant comparisons are among imperfect institutional configurations. This attention to the relative advantages and shortcomings of flawed institutions is a hallmark of the institutional economics of Ronald Coase and Oliver Williamson (1985) and goes back to Pareto, who, immediately after having shown the equivalence of competitive and collectivist allocations in a highly abstract model, introduced the idea of transactions costs: “A second approximation will take account of the expense of putting the mechanism of free competition in full play, and will compare this expense with that necessary for establishing some other new mechanism society may wish to test” (Pareto 1896:500).

Which combination of market, state, and community is most successful in addressing a given coordination problem depends on the underlying technological and social facts that give rise to interdependence among actors. For example, strong increasing returns in a production process make both solitary production and market competition not only inefficient (because marginal cost pricing is not feasible) but also difficult to sustain (because of the positive feedbacks generated by increasing returns and the resulting winner-take-all aspect of the competitive process). Institutions will affect four aspects of economic interactions. First, institutions influence the distribution of information, the way in which information can be acquired, hidden, shared, and used to enforce contracts. Second, institutions in conjunction with a given distribution of wealth differ in the assignment of decision-making power and residual claimancy status among those participating in an interaction. Third, differing institutions and wealth distributions give rise to distinct patterns of conflict of interest among parties to transactions. Finally, the institutions governing a particular interaction will affect the preferences and beliefs of the participants.

A capsule overview of the argument is the following: institutional differences have important allocative consequences where conflicts of interest exist among actors whose interdependence is not governed by complete contracts. The coordination failures that arise in these situations may be attenuated by institutions that accomplish one or more of the following desiderata. First, they may more closely align rights of control and residual claimancy so that individuals own the results of their actions, reducing the degree of effective interdependence. Second, they may reduce the conflict of interest over noncontractible aspects of a transaction among affected parties. Third, they may reduce the extent

or importance of private information, allowing for more complete contracting and more efficient bargaining.

Using these ideas to compare institutions (including communities) will occupy the remainder of the chapter. What are the distinct capacities of markets, governments, and communities that might serve these ends?

Adam Smith's appreciation of the value of competitive markets is distinctively modern: markets make collusion difficult when competition is socially beneficial. "People in the same trade seldom meet together," he wrote, "even for merriment and diversion, but the conversation ends in a conspiracy against the public; or in some contrivance to raise prices" (Smith 1937:128). If such conspiracies are to be effective in a market setting, large numbers of actual and potential participants must cooperate in what is a public goods game. As we saw in chapter 13, sustaining cooperation in these situations through the threat of subsequent retaliation and related strategies becomes exceptionally difficult as the numbers of participants rise. Thus, by increasing the number of "conspirators" necessary to affect prices, competitive markets impede collusion in a situation in which collusion is not socially beneficial.

The first attractive feature of markets is thus a result of the noncooperative interactions that result from large numbers interactions. Market competition is a means of inducing agents to make public the economically relevant private information they hold. It is often said that in markets people vote with their money, which is correct if what is meant is not that markets are democratic but rather that it is costly to express a preference in a competitive market system. Indeed, the only way to register a preference in a market is to make a purchase, and the price at which one is willing to purchase a good conveys what would otherwise be private information, namely, that the good is worth at least as much as the price paid.

Similarly, in a market interaction it is rewarding to reveal a productive capacity and costly to misrepresent the true costs of production. In a competitive market equilibrium with non-increasing returns, profit-maximizing producers will make goods available at their private marginal cost of production, thereby revealing an important and otherwise private piece of information. Those who "misrepresent" their productive capacities by offering goods at prices not equal to the marginal cost will make lower profits than those whose prices convey the true costs. In effect, market competition turns the pricing problem into an  $n$ -person prisoner's dilemma in which the  $n$ -producers have a common interest in restricting output and "overstating their costs" by setting  $p > mc$ . But if  $n$  is large, each firm has an incentive to defect by undercutting its rivals, thereby revealing its true production conditions.

By contrast to markets, in centralized nonmarket systems, producers typically have an incentive to understate their productive capacities to secure a lower production quota. Consumers similarly have an incentive to overstate their needs hoping to establish a superior claim on goods and services.

Second, where residual claimancy and control rights are closely aligned, market competition provides a decentralized and relatively incorruptible disciplining mechanism that punishes the inept and rewards high performers. Markets are a way of increasing what biologists call selection pressure: they have the effect of reducing the variance of performance and hence (under suitable conditions) increasing average performance. The substantial differences observed between high and low performers (chapter 2) suggest that the process works imperfectly, but also that when it does work, the resulting effects on productivity can be significant.

The disabilities of markets are related to their strengths. Markets, it is said, impose hard budget constraints on the relevant actors, but they do this only when decision makers own the results of their decisions. However, because contractual opportunities are dependent on wealth and for other reasons, residual claimancy and control are often misaligned; as a result, the disciplining process is often poorly targeted. A job well done need not benefit an employee who is paid a fixed wage. A plant closing, to take another example, will eliminate the job rents of hundreds of workers; but it need not punish those responsible for the losses that induced the shut-down. Moreover, even where control over noncontractible actions and residual claimancy over a project's income stream are unified, environmental externalities and other external effects carry the consequences of actions taken by the decision maker far beyond the reach of contracts.

By contrast to markets, states may attenuate coordination failures by their ability to allow and often compel individuals to interact cooperatively in situations where noncooperative interactions are inefficient. The comparative advantage of governments is in the production of rules: states alone have the power to make and enforce universal compliance with the rules of the game that govern the interaction of private agents. Where individuals face prisoners' dilemma-like situations or other coordination problems in which the autonomous pursuit of individual objectives leads to an undesirable outcome, the state may provide or compel the coordination necessary to avert this outcome. Services that governments can perform well that communities and markets often cannot include the definition, assignment, and enforcement of property rights, the provision of public goods, the regulation of environmental and other external or "spillover" effects, the regulation of natural monopolies, the provision of some forms of insurance, and macroeconomic

regulation. Less obvious cases include equilibrium selection: where multiple equilibria exist, a one-time state intervention may be able to implement the socially desirable equilibrium. Basu and Van (1998), for example, show that a one time ban on child labor could displace an equilibrium constituting a kind of poverty trap and induce a movement to another equilibrium in which the children and their families would all be better off.

The state addresses prisoners' dilemmas in a manner diametrically opposed to that of markets. Competitive markets hinder the formation of cartels and other forms of collusion by providing incentives for defection, while the state can induce cooperation by impeding defection. Since both defection and cooperation are desirable under different circumstances, markets and states serve complementary roles in solving coordination problems. The state prevents defection by compelling participation in exchanges that would not be voluntarily chosen by economic agents acting singly—for example, cooperating in a prisoner's dilemma situation. This capacity to force compliance can contribute to the solution of coordination problems even where individuals have information that is private and therefore inaccessible to the state.

An example involving the availability of some kinds of insurance illustrates this principle. Before they have learned the capacities, health status, and the special risks they face as individuals, all members of a population might prefer to purchase insurance. But after they have learned their own special position, those with a low probability of collecting on the insurance will not be willing to purchase it since they would be subsidizing those with a high probability of collecting. Thus the low-risk people would drop out of the market and the price of the insurance would be too high for the high-risk people. Since before obtaining specific knowledge of their own risk position, all would have been willing to purchase the insurance, and since it is unavailable on the market, there is a clear market failure. By providing the insurance and compelling all agents to pay for it, the state overcomes this market failure.

Other examples have been provided in the previous chapters. Implementing the social optimum curfew in chapter 6 may require the city planner to set a curfew (and then let the Deadhead and the sleepyhead use their private information to make Pareto improvements over this mandated curfew through Coasean bargaining). In chapter 9 we saw that starting from an assignment of property rights in competitive equilibrium, a redistribution of wealth by governmental fiat could enhance both Pareto-efficiency and technical efficiency. Other, less transparent examples can be given: by conferring the right to collectively bargain on employees, the under-provision of on-the-job amenities and the oppor-

tunities for arbitrary use of short side power demonstrated in chapter 10 may be attenuated.

The state, however, has several weaknesses as a governance structure. The first is state officials' lack of access to private information held by producers and consumers. The second is the mirror image of the first: the lack of access by voters and citizens (assuming a democratic polity) to the private information held by state officials. In this case, the agent (the state) is only weakly accountable to the principals (the citizens). The same arguments showing that first-best solutions are generally unattainable in principal-agent relationships in private exchange apply here as well. The third shortcoming of the state as a governance structure is that there exists no ideal system of making decisions that are binding on large numbers of people. Because there is no consistent democratic way to aggregate individual preferences into consistent social choice criteria, the results of majority rule and other voting mechanisms depend critically on who controls the voting agenda. Moreover, unlike markets, voting schemes have difficulty representing the intensity of preferences for different goods or social outcomes. Finally, where government intervention suppresses market outcomes, economic actors privileged by the intervention earn rents—incomes above their next best alternative. Thus groups will engage in rent-seeking behavior, attempting to influence it to intervene on their behalf rather than for another group or the public at large, thereby wasting resources and distorting policy outcomes.

As in the case of markets, these weaknesses derive from the state's unique capacities. To compel while preventing exit requires that the state be universal and unchallenged in some spheres. This universality of the state makes it difficult to render the state accountable by subjecting it to the competitive delivery of its services. Moreover, the inability of voting schemes to aggregate preferences in a consistent manner requires that nonelectoral ways of influencing collective decision making—including interest group activities—must be available as correctives. But it is difficult to regulate the rent-seeking activity directed toward these nonelectoral processes without corrupting democratic procedures. Of course, states can be made more accountable by fostering competition among local governments, other public agencies, and private bodies, by ensuring competition among autonomous parties and civil liberties so as to foster the careful monitoring of the actions of state officials, by subjecting elected and administrative positions within the state to well-designed incentives, and by limiting the state's actions to those that cannot be regulated in a more accountable manner by some other governance structure.

## COMMUNITY GOVERNANCE

To Marx and other nineteenth-century modernists, “community” was the antithesis of markets, representing an anachronistic remnant of feudal times, destined to be swept away by the requirements of economic progress or, as Marx and Engels (1978:475) put it, by the “icy water of egotistic calculation.” The inertial character of community governance was affirmed by economic historians who, like Marx, pointed to the restrictions placed on individual initiative and the poorly defined property rights associated with the collective decision making required by the open field system of agriculture that prevailed in England and many parts of early modern Europe. Agricultural productivity, according to this view, was held back until the common lands were enclosed and assigned to private owners, as they were in England in the late eighteenth century. But this staple of economic instruction has been overturned by quantitative economic historians during the past generation. A leading contributor to the new literature, Robert Allen (2000:43, 50) writes:

[T]he open fields were an efficient institution for meeting the needs of small scale, grain growing farmers. These needs included diversification against . . . risk . . . and increasing agricultural productivity . . . Enclosure explains neither the productivity advantage that England enjoyed over other countries *c.* 1800 nor the rise in efficiency that had occurred since the middle ages.

The communities governing the open field system used local information and peer pressure to foster innovation and solve the allocational problems arising through the unavoidable interdependence of the farmers. In contrast to the farmers in Palanpur whose inability to coordinate an optimal early planting of their crops provided the introduction to chapter 1, in Taston, England, in 1703 “three fieldmen were chosen on the first of each year to establish the dates when [crops] would be planted, when animals would graze and to enforce the maintenance provisions” (Allen 2000:58).

Recent historical research has also demonstrated the importance of community-based governance in handling the incentive problems associated with incomplete credit contracts in nineteenth-century German banking (Banerjee, Besley, and Guinnane 1994). Community-based governance plays a central role in many sectors of the modern economy, from the development and distribution of open source software to the role of ethnic networks in the mobilization and allocation of credit among motel owners in the United States. Thus, far from a vestigial remnant of

the past, community governance has survived due to its ability to attenuate the incentive problems arising in contemporary economies.

Communities sometimes solve problems that both states and markets are ill-equipped to address, especially where the nature of the social interactions or the goods and services being transacted preclude complete contracting. An effective community monitors the behavior of its members, thereby making them accountable for their actions. Community governance relies on dispersed private information that is often unavailable to states, employers, banks, and other large formal organizations to apply rewards and punishments to members according to their conformity with or deviation from social norms. In contrast to states and markets, communities effectively foster and utilize the incentives that people have traditionally deployed to regulate their common activity: trust, solidarity, reciprocity, reputation, personal pride, respect, vengeance, and retribution, among others.

Several aspects of communities account for their unique capacities as governance structures. First, in a community, the probability that members who interact today will interact in the future is high, and thus there is a strong incentive to act in socially beneficial ways now to avoid retaliation in the future. Second, the frequency of interaction among community members lowers the cost and raises the benefits associated with discovering more about the characteristics, recent behavior, and likely future actions of other members. The more easily acquired and widely dispersed this information, the more will community members have an incentive to act in ways that result in collectively beneficial outcomes. Third, communities overcome free-rider problems because their members directly punish anti-social behaviors. Monitoring and punishment by peers in work teams, credit associations, partnerships, local commons situations, and residential neighborhoods is often an effective means of attenuating incentive problems that arise where individual actions affecting the well-being of others are not subject to enforceable contracts.

But how might communities enforce such norms in the absence of the state's judicial apparatus? Recall that Alchian and Demsetz (1972) suggest that residual claimancy should be assigned to an individual designated to monitor team members' inputs, thus providing incentives for the (noncontractible) activity of monitoring itself, while addressing the members' incentive to free ride by the threat of dismissal by the monitor. (I explained in chapter 10 the underlying assumptions underlying this argument.) Another well-known solution is provided by Hölmstrom (1982) who models a principal multi-agent relationship in which efficiency or near-efficiency is achieved through contracts that make in-

dividual team members residual claimants on the effects of their actions without conferring ownership rights on them. Hölmstrom's solution is infeasible, however, when there are significant stochastic influences on the level of performance of the team, team members have limited wealth, and capital and insurance markets are imperfect.

These explanations have in common that individuals are treated as self-interested. By contrast, many behavioral scientists outside of economics have sought to explain communities by relations of altruism, affection, and other non-self-regarding motives. Many of these approaches, however, have treated the community organically without investigating whether or not the problem-solving capacities claimed for communities are consistent with the fact that individual members are pursuing their own interests (whether self-regarding or not). As a result, some treatments—like Marx's—represent community governance as an anachronism based on collectivist behavioral habits that will be eroded over time and replaced by individual choice. However, we saw in chapters 3 and 4 that motives of reciprocity, shame, generosity, and other social preferences can provide the behavioral foundations of a model of mutual monitoring that avoids these shortcomings. The public goods with punishment experiment and the model presented there indicate that under favorable institutional circumstances and with sufficiently many members motivated by social preferences, high levels of voluntary provision of public goods can be sustained.

Like markets and governments, communities also fail. The personal and durable contacts that characterize communities require them to be of relatively small scale. A preference for dealing with fellow members therefore often limits their capacity to exploit gains from trade on a wider basis. Moreover, the tendency for communities to be relatively homogeneous may make it impossible to reap the benefits of economic diversity associated with strong complementarities among differing skills and other inputs. Neither of these limitations is insurmountable. By sharing information, equipment, and skills, for example, the Japanese fishermen (mentioned above) exploited economies of scale unattainable by less cooperative groups, and reaped substantial benefits from the diversity of talents among the membership. Similarly, cooperation in the local business networks in what is called "the third Italy" along with their associated local governments allow otherwise unviably small firms to benefit from economies of scale in marketing, research, and training, allowing their survival in competition with corporate giants. But compared to bureaucracies and markets, which specialize in dealing with strangers, the limited scope of communities often imposes inescapable costs.

A second community failure is less obvious. Where group membership is the result of individual choices rather than group decisions, the composition of groups is likely to be more culturally and demographically homogeneous than any of the members would like, thereby depriving people of valued forms of diversity. The model of residential segregation in chapter 2 showed that if individuals sort themselves among communities, there will be a strong tendency for communities to end up segregated by race even if this is an outcome that no individual prefers. In cases such as this, integrated communities would make everyone better off, but they will prove unsustainable if individuals are free to move.

When insider-outsider distinctions are made on divisive and morally repugnant bases such as race, religion, nationality, or gender, community governance may contribute more to fostering parochial narrow-mindedness and ethnic hostility than to addressing the failures of markets and states. The problem is endemic. Communities work because they are good at enforcing norms, and whether this is a good thing depends on what the norms are. The recent resistance to racial integration by the white residents of Ruyterwacht (near Cape Town), is as gripping an account of social capital in action as one can imagine (Jung 2001). Even more striking is Dov Cohen's (1998) study of U.S. regional differences in the relationship between violence and community stability. Nisbett and Cohen (1996) described a "culture of honor" that often turns public insults and arguments into deadly confrontations among white males in the U.S. South and West, but not in the North. Cohen's research confirms that in the North, homicides stemming from arguments are less frequent in areas of higher residential stability, measured by the fractions of people living in the same house and the same county over a five-year period. But this relationship is *inverted* in the South and West, residential stability being positively and significantly related to the frequency of these homicides where the culture of honor is strong.

Thus, over some range of governance problems, communities contribute to the desiderata outlined above: aligning control and residual claimancy through the punishment of those inflicting costs on other group members, making information less private by providing incentives to establish reputations through consistent behavior, and reducing the degree of conflict of interest over noncontractible aspects of exchange through the provision of division rules and other norms capable of working even when property rights are not well defined. These reasons may help explain why communities, long dismissed by social scientists as anachronistic remnants of an earlier era, have not been eclipsed by markets and the state.

The ability of communities to address coordination problems depends on the types of property rights in force and their distribution among the population. Where community members are not residual claimants on the results of their actions, there may be little incentive to engage in the forms of sanctioning and reputation building we have stressed. Among the Chicago neighborhoods mentioned at the outset, for example, where most residents are renters rather than home owners collective efficacy was significantly lower. This may be due to the fact that, if some members of a group are vastly more wealthy than others, shared norms may be difficult to maintain, and the punishment of noncooperative actions may lack effectiveness or credibility. For similar reasons, the distinctive capacities of communities are likely to be undermined where the costs of exit are very asymmetrical, for instance when some members have attractive outside options and others do not. In short, the effectiveness of communities depends on the assignment of property rights and on individuals' outside options.

In this respect communities are not unlike markets. The allocational efficiency advantage of the decentralization of control rights (either the extensive use of markets or community-based governance systems) lies in placing decision making in the hands of those who have relevant information that others lack. For this to be beneficial, the holders of private information must be residual claimants on the results of their actions. On efficiency grounds, decentralization to individuals through use of markets is favored over decentralization to communities in cases for which contracts are relatively complete and enforceable at low cost and hence in which interests may conflict without generating coordination failures. Decentralization to communities is favored where complete contracting is precluded but where low levels of conflict of interest within the community and other aspects of community structure facilitate the transmission of private information and mutual monitoring among community members. William Ouchi (1980) suggests that where neither complete contracting nor informal community-based enforcement is possible and where conflicts of interest are significant, bureaucratic organization results, the modern conventional firm being his example. Thomas Schelling (1960:20) put the same point more colorfully:

[W]hen trust and good faith are lacking and there is no legal recourse for breach of contract . . . we may wish to solicit advice from the underworld, or from ancient despotisms, on how to make agreements work.

Most economic interactions are governed by a heterogeneous set of formal and informal rules reflecting aspects of markets, states and communities. Some combinations work better than others.

## INSTITUTIONAL COMPLEMENTARITIES AND CROWDING OUT

For concreteness I will begin with two examples.

The lobster fishermen on the coast of Maine have for decades regulated their catch by limiting access to a defined fishing territory. Only those belonging to a particular so-called harbor gang—those fishing from a particular harbor who have been granted membership—are by local custom allowed to set their traps in the territory (Acheson 1988). Boundary violators are likely to find the buoys cut from their traps, which are then impossible to locate. Intruders have been fired upon. Infringements of environmental regulations or violations of the norms of the gang are also sanctioned by other gang members. In recent years, the State of Maine has formalized the gang system by recognizing the territories of the harbor gangs and setting up democratically elected councils with powers to regulate limits on number of traps and numbers of days fishing. State officials occasionally intervene when conflicts exceed the enforcement capacities of the local communities, as they did during the near collapse of the fishery during the 1920s, or when violence between gangs erupts. But the State employs only six officers to enforce environmental regulations along the entire 4342-mile coastline and to oversee the fishing of 6,800 lobstermen. In recent years, fishing yields have grown and the lobstermen have prospered.

The relationship between the harbor gangs and the state of Maine illustrate *institutional complementarity*. The effectiveness of the state's regulations is greatly enhanced by their informal enforcement by the gangs, while the gang's effectiveness is conditioned on the availability of the state as the enforcer of last resort. Another example of institutional complementarity are the symbiotic effects of trade unions (regulating labor effort) and macroeconomic regulation (reducing volatility of demand for labor) in underpinning the Pareto-improving effort-wage bargains modeled in chapter 8.

The mismanagement of the Himalyan forests of Kamaun and Garhwal districts in Uttar Pradesh, India, provides a sharp contrast to the success of the harbor gangs.<sup>4</sup> Before the twentieth century, large well-defined tracts of forests were considered the exclusive property of each village. Access was regulated by the village *panchayats*; should unauthorized outsiders remove forest products, fighting might break out or fines be levied. To this point forestry management resembled the decentralized regulation by Maine's harbor gangs. But during the First World War, the British colonial administration took over the forestry manage-

<sup>4</sup> This account is based on Sethi and Somanathan (1996) and Somanathan (1991).

ment, seeking to meet the demand for railroad ties and other wood products. The colonial intervention disrupted the regulation by the local communities and evoked incendiary protests that destroyed large stands of pine. The government, in retreat, awarded access to the less valuable forests to “all *bona fide* residents of Kumaun” thereby obliterating the traditional boundaries of village forests and making local regulation virtually impossible. For example, in 1932 a group of villagers from Papedev prevented their neighbor, Jeet Lal, from harvesting grass from the forest, because he had not contributed to the construction of fencing for the grass preserve. Jeet Lal took his neighbors to court and *they* were fined, the punishment being upheld on appeal because, according to the new regulations, Jeet Lal had an unconditional right of access.

The government’s destruction of the community’s capacity to regulate access illustrates the opposite of complementarity, namely, *institutional crowding out*. This occurs when the presence of one institution undermines the functioning of another. Another example of crowding out comes from nearby Palanpur (also in Uttar Pradesh) where the extension of the labor market (and increased geographical mobility) appears to have reduced the costs of exit and hence the value of one’s reputation, with the effect that the informal enforcement of lending contracts has been undermined (Lanjouw and Stern 1998:570). The counter-productive imposition of fines to deter tardiness at the daycare centers in Haiffa (chapter 3) is another example of crowding out: using a market mechanism (the fine) seems to have undermined the parents’ sense of personal obligation to avoid inconveniencing the teachers.

Experiments confirm that crowding out may be a common problem. To explore the effects of explicit incentives, Fehr and Gaechter (2000a) designed a gift exchange game in which principals (employers) make a wage offer with a stipulated desired level of effort on the part of the agent (worker). The agent may then choose an effort level, with costs to the agent rising in effort. In the “trust” treatment, the interaction ends there, but in the “incentive” treatment, following the agent’s choice of an effort level, the employer may fine the worker if the worker’s effort level is thought to be inadequate. By contrast with the trust treatment, the incentive treatment links pay to performance and hence represents a more complete contract. In this experiment, the total surplus from the interaction is the principal’s profits plus the agent’s wage minus the cost of effort (and the fine, where applicable).

In the trust treatment, a self-regarding agent would choose the minimum feasible level of effort irrespective of the principal’s wage offer, and, anticipating this, a self-regarding principal would offer the minimum wage. Experimental subjects did not conform to this expectation: employers made very generous offers and workers’ effort levels were

strongly conditioned on these offers, high wages being reciprocated by high levels of effort. The introduction of explicit incentives had a dramatic effect: average effort levels by agents were substantially *lower*. Only for very low wage offers did the explicit incentives elicit (marginally) higher levels of work. For relatively generous wage offers, the effort performed with explicit incentives was about a third the level performed in their absence.

The experiment was constructed so that had subjects responded optimally on the basis of self-regarding preferences, the surplus would have been more than twice as great under the incentive treatment as under the trust treatment. But the total surplus was higher in the trust treatment, by 20 percent in those cases in which the principal offered a contract such that the expected fine for shirking exceeded the cost of working (so that the no-shirking condition was fulfilled), and by 53 percent where the principal's contract did not meet the no shirking condition.

A striking result of this experiment emerges if we compare the distribution of the surplus under the trust treatment and the incentive treatment. In the incentive treatment (confining our attention to the cases in which the principal's contract fulfilled the no shirking condition), profits are more than double the profits in the trust treatment, while the net payoffs to the agent are less than half. The incentive treatment allowed employers to save enough in wage costs to offset the reductions in work effort. Summarizing this result, Fehr and Gaechter (2000a:17) write, "The incentive opportunities in the incentive treatment allow principals to increase their profits relative to the trust treatment, but . . . this is associated with an efficiency loss."

Similar results occurred in a field experiment in Colombia conducted by Juan Camilo Cardenas (Cardenas, Stranlund, and Willis 2000). The experiment, a variant of the public goods game, captured the logic of a common pool resource extraction problem—degradation of a nearby forest—faced by the rural people who participated. In the absence of explicit incentives the subjects selected extraction levels not far above the social optimum and much less than the Nash equilibrium level based on individual optimization with self-regarding preferences. But when monitoring of the subjects' extraction levels (by the experimenter) and the prospect of a fine for over-extraction were introduced, subjects extracted more rather than less. After a few rounds, their extraction levels approximated the Nash equilibrium level (taking account of the fine). Like the fine imposed on the tardy Haifa parents, the effect of "improving" the incentive structure apparently was to diminish the salience of the other-regarding motives that had been in force in the absence of the incentives.

A final experiment may provide some insight into how crowding out

works (Frohlich and Oppenheimer 1992). Subjects played five-person public goods games under two conditions: one group played the standard contribution game and the other played a modified (veil of ignorance) game in which a randomized assignment of payoffs made it optimal to contribute the maximal amount to the public good. Half of the subjects (in each treatment) were allowed to engage in discussion before each play (of course, the discussion should have had no effect on the outcome of the standard game, as the dominant strategy is to contribute nothing). After eight rounds of play, another eight rounds were conducted, this time with the same groups but with all playing the standard game. Among those who had been permitted discussion, those who had experienced the incentive-compatible (veil of ignorance) game contributed significantly less in the final eight rounds, and (in subsequent questionnaires) expressed less concern with questions of fairness.

The authors' explanation is that the incentive-compatible mechanism rewarded those contributing to the public good, thus making self-interest a good guide to action, while those experiencing the standard game gained high payoffs only to the extent that they evoked considerations of fairness as a distinct motive among their group-mates. They conclude:

The failure of the . . . [incentive-compatible] mechanism to confront subjects with an ethical dilemma appears to lead to little or no learning in ethical behavior in the subsequent period. . . . It is an institution, like other incentive compatible devices, which can generate near optimal outcomes. . . . However from an ethical point of view it is not only unsuccessful as pertains to subsequent behavior; it appears to be actually pernicious. It undermines ethical reasoning and ethically motivated behavior. (Frohlich and Oppenheimer 1992:44)

Fehr and List (2002) offered a different interpretation of counter-productive incentives found in their trust experiments with Costa Rican businessmen. They found that the highest level of trustworthiness was elicited when the principal was *permitted* to fine the agent for untrustworthy behavior, but *declined to use it*, evidently a signal by the principal of trusting behavior that was then reciprocated by the agent. By contrast “explicit threats to penalize shirking backfire by inducing less trustworthy behavior.” They conclude that: “the psychological message that is conveyed by incentives—whether they are perceived as kind or hostile—has important behavioral effects.”

Experiments by psychologists have demonstrated conditions under which extrinsic rewards (to use their terminology), such as monetary payment for performance of a task, may diminish one's intrinsic motivation to do the task (Deci, Koestner, and Ryan 1999). These crowding-out effects appear for interesting rather than boring tasks and when

the reward is expected in advance and closely tied to the task performance. One may conclude that performance-based pay in workplaces may diminish employee's motivation to do tasks that they initially found intrinsically interesting or challenging. But the evidence is also consistent with an important role for explicit (extrinsic) incentives in motivating individuals to do tasks in which they have little intrinsic interest (namely, a great many jobs in modern economy).

#### CONCLUSION: MANDEVILLE'S MISTAKE

Implementation theory is a branch of economics that studies the ways in which privacy-preserving contracts and decision rules—in short, constitutions—can lead individuals with self-regarding preferences to implement (as a Nash equilibrium) an outcome not sought by any of the individual participants but which by some measure is socially valued. The methods of contemporary implementation theory are new, but the idea goes back to Mandeville's radical conjecture (in the epigraph to chapter 2) that interactions could be structured so that “The worst of all the Multitude Did something for the Common Good.” This objective of harnessing indifferent motives to elevated ends has been central to constitutional thinking ever since. Recall that David Hume (in the epigraph for chapter 3) recommended the maxim that “in contriving any system of government . . . every man ought to be supposed to be a *knave* and to have no other end, in all his actions, than private interest.” But the above examples of institutional complementarity and crowding out suggest the effectiveness of policies and laws may depend not solely on their capacity to recruit selfish ends to social purposes but also on the preferences they induce or evoke. Albert Hirschman (1985:10) pointed out that economists propose

to deal with unethical or antisocial behavior by raising the cost of that behavior rather than proclaiming standards and imposing prohibitions and sanctions. The reason is probably that they think of citizens as consumers with unchanging or arbitrarily changing tastes in matters civic as well as commodity-related behavior. . . . A principal purpose of publically proclaimed laws and regulations is to stigmatize antisocial behavior and thereby to influence citizens' values and behavioral codes.

As we have just seen, raising the cost of an anti-social behavior and other explicit incentive-based devices may actually do harm. There is thus a norm-related analogue to the second best theorem of welfare economics: *where contracts are incomplete (and hence norms may be important in attenuating market failures), more closely approximating*

*idealized complete contracting markets may exacerbate the underlying market failure (by undermining socially valuable norms such as trust or reciprocity) and may result in a less efficient equilibrium allocation. A constitution for knaves may produce knaves.*

The fact that institutions and preferences coevolve suggests an important (if difficult) extension of implementation theory and a modification of the Humean maxim. In seeking to implement a socially desired outcome, one must check that the preferences necessary to implement the outcome are sustainable under the policies, contracts, or rules used in the implementation. The problem is more difficult than Hume suggested, not only because preferences are endogenous but also because, as we saw in chapter 3, populations are heterogeneous and individuals are versatile. The problem, then, is not to find a way to induce a homogeneous population of self-regarding individuals to implement a socially desirable outcome. Rather, it is to devise rules such that in cases in which cooperation is socially desirable, individuals with other-regarding preferences will have opportunities to express their pro-sociality in ways that induce all or most to cooperate, as in the public goods with punishment experiments discussed in chapter 3. And in situations in which competition rather than cooperation is essential to socially valued outcomes, the task is exactly the opposite.

Providing practical guidance on how this might be done is one of the major challenges to contemporary studies of economic institutions and behavior. Modern microeconomics has demonstrated the important contribution that well-defined property rights can make in meeting this challenge. In his Nobel Prize lecture Ronald Coase expressed this position succinctly:

It is obviously desirable that these rights should be assigned to those who can use them most productively and with incentives that lead them to do so and that, to discover (and maintain) such a distribution of rights, the costs of their transference should be low, through clarity in the law and by making the legal requirements of such transfers less onerous. (Coase 1992:718)

But modern microeconomics also shows that well-defined and easily transferred property rights are unattainable in important arenas of economic interaction—in labor and credit markets, in neighborhoods, in adherence to socially valuable norms including the rule of law, and in the production and distribution of information, for example. In these cases, the government can contribute to economic performance through the direct assignment of property rights, rather than simply by facilitating their private exchange. Robert Frost’s “Good fences make good neighbors” is the epigraph for the chapter entitled “Utopian Capitalism.” But the New England poet’s point was quite the opposite, namely,

that his curmudgeonly neighbor's embrace of well-defined property rights may be unwarranted. Here is some of the rest of the poem:

He only says, "good fences make good neighbors."  
 . . . Why do they make good neighbors? Isn't it  
 Where there are cows? . . .  
 Before I build a wall I'd ask to know  
 What I was walling in or walling out,  
 And to whom I was like to give offense.  
 Something there is that doesn't love a wall, That wants it down.  
 He moves in darkness as it seems to me,  
 Not of woods only and the shade of trees.  
 And he likes the thought of it so well  
 He says again, "Good fences make good neighbors."

(Frost 1915:11–13)

Neighborliness may also be necessary where good fences fail. From Mandeville to Arrow and Debreu, economic thinkers have sought to devise property rights and other rules that would induce self-regarding individuals to implement socially desirable aggregate outcomes. Of particular interest has been the question, under what conditions will the competitive exchange of well-defined property rights among self-regarding individuals result in an outcome that is in some sense optimal. In light of the importance of self-interest in human motivation, the insights produced by this three-century-long tradition are a major contribution to science and to public policy. But as we now know, thanks to the Fundamental Theorem, the invisible hand requires complete contracting and nonincreasing returns, and these do not describe, even approximately, any known economy.

The project that began with the *Fable of the Bees* may be of even less practical relevance in the future. The reason is that the technologies and social interactions of the modern economy increasingly depart from these canonical assumptions. Direct noncontractual interactions with positive feedbacks arise increasingly in modern economies, as information-intensive team production replaces assembly lines and other technologies more readily handled by contract, and as difficult to measure services usurp the preeminent role — as both outputs and inputs — once played by measurable quantities like kilowatts of power and tons of steel. Danny Quah (1996) calls the modern system of production "the weightless economy." The key characteristics of the information-intensive economy are generalized increasing returns, with near zero marginal costs in many cases, along with the fact that most information is either not subject to complete and enforceable contracts, or will be inefficiently allocated if it is. Kenneth Arrow (1999:162, 156) writes that

information is a fugitive resource . . . we are just beginning to face the contradictions between the systems of private property and of information acquisition and dissemination. . . . [We may see] an increasing tension between legal relations and fundamental economic determinants.

The information-intensive economy of the future may more closely resemble the economy of the mobile foraging band in human prehistory, rather than the economy of grain and steel that displaced it. Pursuing good ideas with practical applications is a costly and uncertain project, much like hunting large game. Success is rare, but its fruits are immensely valuable. The private appropriation of the prize is both difficult to accomplish and socially wasteful, for the foregone benefits to those excluded from access to the prize far outweigh the gains to the individual appropriator to be had by excluding others. A new drug or a new software application is not so different in this respect from an antelope. Thus it is not surprising that the system of prestige and norms of sharing in some parts of the modern information-intensive economy—those involved in open source software, for example—in many ways parallel the culture of the foraging band.

The challenges laid out by Arrow are not likely to be addressed simply by greater precision in the definition of private property rights. It appears equally utopian to think that national governments would (or even could) devise centralized solutions to these problems. A complementary configuration of market, states, and community governance may be the best hope for mobilizing the heterogeneous and versatile capacities and motives of people to address these dilemmas, to better harness the potential of expanding knowledge to the objective of human betterment, and thereby to make good what Alfred Marshall a century ago identified as the promise of economic studies.