Readings in Public Choice and Constitutional Political Economy

Edited by Charles K. Rowley
and Friedrich G. Schneider

Springer

Readings in Public Choice and Constitutional Political Economy

Charles K. Rowley · Friedrich Schneider

Readings in Public Choice and Constitutional Political Economy

Charles K. Rowley
Locke Institute
Fairfax, VA 22030
USA

and

Department of Economics
George Mason University
Fairfax, Virginia 22030
USA
crowley@gmu.edu

Friedrich G. Schneider
Universität Linz
Inst. Volkswirtschaftslehre
Abt. Volkswirtschaftspolitik
Altenbergerstr. 69
4040 Linz
Austria
friedrich.scheider@jku.at

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

We dedicate *Readings in Public Choice and Constitutional Political Economy*
to the memory
of Duncan Black
23 May 1908 to 14 January 1991
The Founding Father of Public Choice

Preface

Readings in Public Choice and Constitutional Political Economy provides a detailed and comprehensive account of the subject known as *public choice*. However, the title would not convey sufficiently the breadth of the contents, which can be summarized better as the fruitful interchange of economics, political science and moral philosophy on the basis of an image of man as a purposive and responsible actor who pursues his own objectives as efficiently as possible.

This fruitful interchange between the fields mentioned above existed in the late eighteenth century, during the brief period of the Scottish Enlightenment when such great scholars as David Hume, Adam Ferguson and Adam Smith contributed to all these fields, and more. However, as intellectual specialization gradually replaced broad-based scholarship from the mid-nineteenth century onwards, it became increasingly rare to find a scholar making major contributions to more than one field.

Once Alfred Marshall defined economics in neoclassical terms, as a narrow positive discipline, the link between economics, political science and moral philosophy was all but severed and economists redefined their role as that of 'the humble dentist' providing technical economic information as inputs to improve the performance of impartial, benevolent and omniscient governments in their attempts to promote the public interest. This indeed was the dominant view within an economics profession that had become dominated by the economics of John Maynard Keynes and Paul Samuelson, immediately following the end of the Second World War.

Even during this 'dark age' for political economy, however, a little known Scot named Duncan Black was sowing the seeds for a renaissance that would enable a reunion between economics and political science. Black launched the public choice research program in 1948 with a seminal paper on the rationale of group decision-making and in so doing earned later fame as the founding father of public choice.

Black's seminal contribution was extended in 1951 by Kenneth Arrow in his famous monograph entitled *Social Choice and Individual Values*. A further major extension occurred in 1957, when Anthony Downs published his seminal book entitled *An Economic Theory of Democracy*. In 1962, James Buchanan and Gordon Tullock, in their famous book *The Calculus of Consent*, extended the perspective of

public choice by shifting attention away from direct elections and parliamentary democracy, to outline a rational choice approach to the logical foundations of a constitutional republic. In 1965, Mancur Olson opened up the discussion of interest group behavior to rational choice analysis in his famous book entitled *The Logic of Collective Action*. In 1971 William A. Niskanen opened up the discussion of bureaucratic behavior to rational choice analysis in his book entitled *Bureaucracy and Representative Government*.

These six contributions constitute the foundations of the public choice research program. Two other books also contributed to the early public choice tradition, namely the 1951 monograph by Black and Newing entitled *Committee Decisions with Complementary Valuation* and the 1962 masterpiece by William Riker entitled *The Theory of Political Coalitions*. All these works are as relevant to scholars of public choice now as they were several decades ago when they were written.

Since public choice first emerged during the years of reconstruction from the devastation of the Second World War, the world's political environment has evolved and changed dramatically. The Marshall Plan enabled Western Europe to eliminate its dictatorships and to establish and/or to reinforce democracy. The European colonial powers eased themselves out of their imperial roles, releasing their former colonies into independence, albeit in many cases an independence that rapidly deteriorated into the one party state, outright dictatorship or even kleptocracy. Even Latin-America has slowly eased itself into democracy, albeit in many cases of a fragile and unstable nature.

The United States utilized its economic strength and its political resilience to confront and contain the USSR throughout the Cold War and eventually to defeat it, thus opening up Eastern Europe and even Russia itself to varying forms of democratic or semi-democratic government. The remaining communist dictatorships, notably The People's Republic of China, Cuba and North Korea, clearly are endangered species, unlikely to survive the first decade of the new century. The last bastions of non-communist, non-sub-Saharan African dictatorship, mostly located in the Middle East, are finding it increasingly costly and difficult to fend off the democratic desires of their down-trodden and mostly impoverished subjects. For the first time in the history of the world, a majority of individuals now live under conditions of democracy, a state that public choice is uniquely qualified to analyze.

Given the enormity of the political changes outlined above, it is very reassuring to discover, not least through the contributions to this book, that public choice has retained its ability to explain and predict the behavior of all actors in political markets—even the behavior of al-Qaeda terrorists—within the framework of the rational choice approach.

Many of the contributions to this volume are revised versions of essays published in Volume 1 of *The Encyclopedia of Public Choice* also edited by Charles K. Rowley and Friedrich Schneider (Kluwer 2004). The volume has been significantly re-organized and several new essays have been added. Great care has been taken throughout to make the language of the Readings as non-technical and comprehensible as possible. For this reason, the *Encyclopedia* should be accessible to all

scholars, all graduate and undergraduate students of economics, political science, and public choice as well as to most scholars and students of such closely related disciplines as law, philosophy, sociology and psychology. The volume should be an indispensable companion to all practitioners of public policy.

The editors have made every effort to present a well-balanced and comprehensive body of public choice scholarship from the early beginnings of the discipline to its current flourishing state. By and large, we believe that we have achieved this goal. However, as always, the proof of the pudding is in the eating. We trust that you will enjoy the rich banquet that is set before you.

Fairfax, VA   *Charles K. Rowley*
Linz-Auhof, Austria   *Friedrich Schneider*

References

Arrow, K.F. (1950). "A difficulty in the concept of social welfare." *Journal of Political Economy*, 58: 328–346.

Arrow, K.J. *Social Choice and Individual Values*. New York: Wiley.

Black, D. (1948). "On the rationale of group decision-making." *Journal of Political Economy*, 56: 23–34.

Black, D. and Newing, R.A. (1951). *Committee Decisions with Complementary Valuation*. London: W. Hodge.

Blaug, M. (2000). *Who's Who in Economics*. Cheltenham, UK and Northampton, USA: Edward Elgar Publishing.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Niskanen, W.A. (1971). *Bureaucracy and Representative Government*. New York: Aldine-Atherton.

Olson, M. *The Logic of Collective Action*. Cambridge: Harvard University Press.

Riker, W. (1962). *The Theory of Political Coalitions*. New Haven: Yale University Press.

Acknowledgments

Contents

Part V   Applications

Part I
Introductory Overviews

Chapter 1
Public Choice and Constitutional Political Economy[1]

Charles K. Rowley

Introduction

Public choice – or the economics of politics – is a relatively new science located at the interface between economics and politics (Rowley 1993, Mueller 1997, Shughart and Razzolini 2001). It was founded in 1948 by Duncan Black, who died in 1991 without ever achieving full recognition as the Founding Father of the discipline (Tullock 1991). Its practitioners seek to understand and to predict the behavior of political markets by utilizing the analytical techniques of economics, most notably the rational choice postulate, in the modeling of non-market decision-making behavior.

Public choice thus defined, is a positive science concerned with what is or what conditionally might be. Its dedicated journal is *Public Choice*, introduced by Gordon Tullock in 1966 and now ranked among the thirty most important journals in social science, worldwide. Its intellectual home is The Center for Study of Public Choice, now located in The James M. Buchanan Center for Political Economy at George Mason University in the Commonwealth of Virginia.

The public choice research program was launched in 1948 by Duncan Black's paper on the rationale of group decision-making. This paper demonstrated that, under certain conditions, at most one motion is capable of securing a simple majority over every other motion. Specifically, if voter preferences are single-peaked over a single-dimensional issue space, a unique equilibrium exists in the motion most preferred by the median voter. For Black (1948), this result was the political science counterpart of the competitive market equilibrium in his own discipline of economics. However, Black was by no means convinced that the median voter theorem would hold in practice. His paper clearly identifies conditions in which majority voting would cycle across pair-wise choices.

In 1950, Arrow, without recognizing Black's insight about cycling, demonstrated that when Black's condition of single-peaked preferences does not hold, the unique vote equilibrium will not hold and voting cycles may prevail (Arrow 1950). Arrow incorporated this insight into his famous 1951 book, *Social Choice and Individual Values,* outlining a difficulty in social welfare (Arrow 1951). These papers fundamentally challenged Black's initial theoretical notion of political stability and offered an alternative theoretical viewpoint that political markets are inherently unstable. In so doing, Arrow seized (perhaps unfairly) for himself the reputation of first establishing the instability theorem. In any event, these alternative viewpoints would be subjected to extensive empirical evaluation throughout the first half a century of the public choice research program.

In 1957, Anthony Downs moved public choice from its early beginnings in analyzing committee decisions and direct elections in an environment essentially devoid of institutions to its subsequent preoccupation with the institutions of democracy and representative government (Downs 1957). In a far-reaching contribution, he laid the foundations for an ambitious research program that would apply rational choice theory to every aspect of the political market place. Without apparently having read Black's (1948) contribution, and having no clear concept of the importance of the median, as distinct from the mean or the mode, as a measure of central tendency, Downs utilized the spatial economic theory of Harold Hotelling (1929) to emphasize the predictable dominance of the middle voter in two party democracies, thus offering a falsifiable theory of democracy that would attract a large volume of high quality empirical research.

However, even while re-establishing the notion that political markets, under favorable circumstances, may reflect the preferences of the middle voter and even while forcing the rational choice analysis of economists down the throats of reluctant political scientists, Downs sowed the seeds of doubt that subsequently generated fruitful research on public choice. He noted that, in an environment where information is complex, costly to acquire, and offering little economic returns to those who acquire it, members of the electorate may economize in its acquisition, relying on ideology, as represented by political party brand images, to direct their voting decisions. He also noted that members of the electorate might rationally abstain from voting in situations where they could not distinguish between the policy positions of rival candidates or political parties.

Such doubts notwithstanding, Downs (1957) essentially replicated the theoretical work of Black (1948), reinforcing the notion that political markets are inherently stable and reflect the preferences of the middle voter. Down's truly original contributions consist of extending the 1948 insight of Black to the real world institutions of politics and in introducing the individual self-seeking, rational choice approach to the analysis of politics.

The classics of public choice reviewed so far focused attention exclusively on voting in unconstrained majority-vote democratic environments as exemplified by the United Kingdom. As such they were only of limited significance, arguably, for a constitutional republic, such as the United States of America, a republic that deliberately was not designed to be a democracy as is usually defined. In 1962,

Buchanan and Tullock ingeniously shifted the public choice perspective away from the environment of parliamentary democracy as envisaged by Downs (1957) to reflect the institutions of a constitutional republic as envisaged by the authors of *The Federalist* (most especially James Madison) almost two centuries earlier.

The Calculus of Consent (Buchanan and Tullock 1962) differed sharply from earlier contributions in the emphasis provided by Buchanan and Tullock on methodological individualism and universal consent. More important for public choice and constitutional political economy, however, was the insight provided by Buchanan and Tullock's constitutional economic theory. The authors were able to demonstrate that at the constitutional stage, an individual rationally would choose to abide by a vote ratio that minimized the sum of his expected external costs and his expected decision-making costs from collective action. Whether this vote ratio is some minority vote, a simple majority or some supra-majority vote depends on the minimum point of the aggregated function. This result was a direct challenge to political scientists who at that time almost universally endorsed the majority rule.

The Calculus of Consent also challenged the new welfare economics of Paul Samuelson and Kenneth Arrow that systematically paved the way for government intervention in free markets on the grounds of widespread market failure. Buchanan and Tullock noted that all categories of market failure – monopoly power, public goods, externalities, limited and asymmetric information and moral hazard – were evident much more in the political than in the ordinary markets, not least because of the ubiquitous presence of indivisibilities in the political markets.

With this insight, Buchanan and Tullock leveled the playing field in the debate over the respective merits of the political and economic markets (Goetz 1991). By directing attention to the difference between 'choices over rules' and 'choices subject to rules', the book also provided the logical foundations for the constitutional political economy research program.

Although Downs (1957) and Buchanan and Tullock (1962) discussed the role of interest groups in political markets, neither of them analyzed interest group behavior from the perspective of rational choice theory. This lacuna was filled by the fifth and final founding father of public choice, Mancur Olson, whose book *The Logic of Collective Action* (1965) fundamentally challenged the conventional political science view of interest group behavior.

Whereas political science viewed interest groups as reflective of underlying voter preferences and as suppliers of relevant information to political markets, Mancur Olson offered a radically different interpretation. Because the objectives pursued by interest groups have profound characteristics of 'publicness', rational choice predicts that their efforts, typically, will be eroded by free-rider problems, so that groups will be difficult to form and to motivate. Predictably, the outcome is the under-provision of interest group pressures in political markets.

However, such difficulties are not uniform across groups. Existing groups have decisive advantages over potential groups in the competition for political favors; groups offering concentrated benefits are more effective than groups offering dispersed benefits; small groups are more effective than large groups; groups that can coerce supply (e.g. professional associations and trade unions) are more

effective than those that cannot; and successful large groups must rely on providing selective (private) benefits to members in order to attract support for policies with public good/bad characteristics (for a trenchant critique of this view, however, see Stigler 1974).

Thus the logic of collective action suggests that competition among interest groups does not simply reinforce an underlying voter-directed political equilibrium. Rather, it predictably distorts the underlying political equilibrium in favor of policies favored by the more effective interest groups, policies that typically provide concentrated benefits for the few financed by dispersed taxes on the many.

Alternative Perspectives in Public Choice

Like all successful intellectual innovations, public choice has given birth to a new generation of scholars, journals and research institutions, offering a diversity of approaches and methods, not all of which correspond to those adopted by the 'founding fathers' (Mitchell 1988, 1989, 2001). Four schools currently dominate the 'public choice' landscape, namely Rochester, Chicago, Virginia and Political Economics, each worthy of a brief discussion.

Rochester

The intellectual entrepreneur of the Rochester school of positive political theory was William Riker, who began to consider the applicability of the rational choice approach and game theory in political science during the late 1950s (Riker 1962). In 1964, he strengthened this presence by transforming his introductory text on American government into the first book on rational choice aimed at undergraduate majors in political science (Riker 1964).

By rejecting the then fashionable behavioral school in favor of the rational choice theory, Riker indicated that he was stepping outside conventional political science in order to embrace the challenge from economics on its own terms. By employing game theory, Riker indicated that conflict and conflict resolution were an integral part of public choice, a view that was not universally shared by the leading Virginian scholars at that time (Buchanan and Tullock 1962).

By 1973, Riker and Ordeshook felt that they were able to define politics as 'the mystery of how social science evolves out of individual preferences' (Riker and Ordeshook 1973, p. 6). Their book demonstrated that the mystery would be resolved by mathematical political science buttressed by the use of rigorous statistical methods. Once again, Buchanan and Tullock, the joint leaders of the Virginia School, were uncomfortable with this choice of scientific method.

The Rochester School encompasses well-known scholars such as Riker, Aranson, Mckelvey and Banks (all now deceased), Ordeshook, Brams, Enelow, Hinich, Munger, Aldrich, Schofield, Fiorina, Ferejohn, Shepsle, Weingast, Romer and

Austin-Smith. It consistently applies positive political science to the study of elections, party strategies, voting agenda manipulation, interest groups, coalition formation, legislative behavior and bureaucratic behavior. The rational choice approach is deployed unremittingly in this research program.

Until the early 1980's, with the notable exceptions of Riker and Aranson, the Rochester School focused primarily on abstract theoretical analysis, largely ignoring institutional details. In part, this reflected a reaction against the institutionalising of conventional political science. In part, it reflected the preoccupation of the Rochester scholars (Riker excepted) with spatial voting models (see, for example, Enelow and Hinich 1984). As empirical work in public choice analysis gradually eroded confidence in the 'vote motive' as a primary determinant of political market behavior, and as Virginia School interest-group theories began to play an ever more important role, the research program of the Rochester School appeared to be in significant decline.

The program was rescued during the early 1980s by such scholars as Kenneth Shepsle and Barry Weingast who shifted direction and initiated influential research into the institutions of the US legislature and the federal bureaucracy. Drawing heavily on recent research findings in the new institutional economics, these scholars have blended political science with economics to the extent that it is now extremely difficult to unravel the primary focus. Initially, the Rochester program was chauvinistic, directed almost exclusively at US institutions and surprisingly narrow, ignoring the complex interactions between the separate branches of a compound republic. More recently, it has extended its focus to the international arena and has begun to model very effectively the interactive behavior of the separate branches of the US government.

The Rochester program, for the most part, eschews normative discussion. Its practitioners, whatever their personal philosophies, report neutrally on such matters as cyclical majorities, log-rolling, interest-group politics, legislative stability, bureaucratic discretion, coalition formations and the like. Some, like Shepsle (1982) are skeptical about constitutional reforms. Others like Fiorina (1983) are hostile to studies that find fault with the federal bureaucracy. Riker and Aranson were notable exceptions to this apolitical neutrality. However, they are no longer with us.

Chicago

The Chicago political economy research program (CPE) was a relatively late starter, launched by George Stigler's 1971 article on economic regulation (Stigler, 1971). Like so much of Chicago scholarship, this program largely ignored preceding non-Chicago work in the field and still fails to cite much of such work in its own publications. In rebuilding the wheel, however, it has made distinctive contributions to public choice literature.

Although Stigler retained the mantle of leadership of the Chicago Political Economy Program until his death in 1991, leading Chicago economists such as

Gary Becker, Sam Peltzman and William Landes and leading legal scholars such as Richard Posner quickly joined the program. Although the Chicago School itself has a lengthy pedigree in normative as well as in positive analysis – including Frank Knight, Jacob Viner, Henry Simons and Milton Friedman – CPE, under the arguably deconstructive influence of Stigler, was overtly positive, asserting for the most part that 'what is, is efficient'. Although economists could observe, explain and predict, attempts to change the course of history, by and large were deemed to be futile and wasteful uses of scarce resources (Rowley 1992, pp. 38–41).

CPE is a body of literature that analyses government from the perspective of the rational choice theory and the neoclassical price theory (Mitchell 1989, Tollison 1989). It views government primarily as a mechanism utilized by rational, self-seeking individuals to redistribute wealth within the society. Homo economicus is modeled almost exclusively as an expected wealth-maximizing agent (Reder 1982). From this perspective, 'fresh-water economics' mocks the 'salt-water economics' of the east coast academies for their adherence to the public interest theory of government: 'Get your heads out of the sand you hay-bags!'

Ironically, however, CPE ends up with a view of the political process that is not far distant from that of the school of public interest. Specifically, political markets are viewed as technically efficient mechanisms for satisfying the preferences for redistribution of individual citizens working through efficient pressure groups. This interpretation of the political process emanates from a fundamentally flawed application of Chicago microeconomics to the political marketplace.

CPE draws on the tight prior equilibrium methodology applied by Chicago economists in their analysis of private markets (Reder 1982) in its study of transfer politics. The thrust of this methodology is towards instantaneous and durable equilibrium, with political markets always clearing. In equilibrium no individual can raise his expected utility (wealth) without reducing the expected utility (wealth) of at least one other individual. Political agents (brokers) clear political markets without invading them as principals. They are driven by constraints, not by preferences. There is no role for ideology in the CPE research program.

The auxiliary hypotheses of the CPE program ensure that political market equilibrium is tight and instantaneous. It is assumed that all political actors are price-takers; that there is no discretionary power in political markets; that the prices at which individuals agree to contract are market-clearing prices consistent with optimizing behavior; that such prices reflect all economically relevant information; that individuals engage in optimal search; and that all constraints on political market behavior are efficient, reflecting expected utility maximizing behavior on the part of those who create or modify them.

The auxiliary conditions imposed by CPE do not produce a political market equilibrium based on perfect foresight. Random disturbances cannot be accommodated by the model. Nor will political actors utilize uneconomic information. The system responds with stochastic analogs of determinist general equilibrium. A particular feature of CPE, as of the Chicago School more generally, is the presumption that only propositions derived from the tight prior equilibrium theory are valid. In a sense, CPE demands that the findings of empirical research must be

consistent with the implications of the standard price theory (Reder 1982). This is a dangerous perversion of the methodology of positive economics advanced by Milton Friedman (1953).

Ultimately, the Chicago presumption must give way if confronted with relentlessly adverse evidence. But this can take a very long time, given the malleability of statistical techniques and of political-economic data. When Gary Becker (1976) remains willing to defend in-kind transfers as carrying lower excess burdens than lump sum transfers of income, when George Stigler (1992) argues that all long-lived trade protection tariffs are efficient, when William Landes and Richard Posner (1987) defend U.S. tort law as being economically efficient, and when the *Journal of Political Economy* publishes papers that defend the U.S. federal farm program as an efficient mechanism for transferring income to wealthy farmers, there is justifiable cause to worry whether CPE scholars and their journal editors ever look out from their ivory towers and survey the real world.

Virginia

The Virginia School, with its early roots in the economics of Frank Knight and Henry Simons at the University of Chicago (Mitchell 1988, 2001) and of Lionel Robbins at The London School of Economics, is the most far-reaching program in public choice, provocative because many of its practitioners do not hesitate to step across the divide separating science from moral philosophy. Under the early intellectual leadership of James M. Buchanan, Gordon Tullock and Ronald Coase, the Virginia School established itself in the teeth of active opposition both from orthodox neoclassical economics and from conventional political science. It has challenged successfully, inter alia, Keynesian macroeconomics, Pigovian welfare economics, Musgravian public finance and the veneration of simple-majority democracies.

From the outset, the Virginia School differentiated its research program from the early public choice contributions of Duncan Black (1948) and Anthony Downs (1957) through its focus on the logical foundations of a constitutional democracy. In 1962, Buchanan and Tullock published *The Calculus of Consent*, arguably the single most important text ever written in public choice and constitutional political economy.

This book demonstrates that individuals are capable of long-run expected utility maximization when establishing the rules of the game, even though they will resort to short-run expected utility maximization when playing under those rules. Because constitutional rules are designed to be durable, individuals confront generalized uncertainty with respect to the impact of such rules on their individual lives. This generalized uncertainty makes possible near-universal consent regarding rules even among a heterogeneous electorate without reliance on the artificial assumption of the veil of ignorance later used by John Rawls in his famous book, *A Theory of Justice* (1971).

The Virginia tradition commenced in earnest in 1957, with the founding by James Buchanan and Warren Nutter of *The Thomas Jefferson Center for Studies in Political Economy* at the University of Virginia. For a decade, Buchanan, Tullock

and Ronald Coase pioneered a research program that would fundamentally change the playing field of political economy throughout the Western world by providing an effective scientific counter-balance to the early postwar onslaught by neoclassical economists targeted against the capitalist system.

Throughout the period 1945–1957, Keynesian macroeconomists, Pigovian welfare economists, Arrovian social choice theorists and Musgravian public finance scholars had waged an unrelenting war against free markets, alleging near-universal market failure and exploring the appropriate public sector responses by benevolent and impartial democratic governments. Even such an old-style, free market economist as Milton Friedman (1962) was forced onto the defensive, devising ever more exotic methods of government intervention designed to minimize the discretionary power of government while recognizing that private markets were widely beset by such problems as monopoly, externalities, public goods and bounded rationality. Even Harold Demsetz, whose writing always stressed the importance of a comparative institutions approach to policy formation, had no theory of government from which to launch a scientific counter-attack.

In a tour de force, Buchanan and Tullock (1962) provided the missing theory of government and placed the advocates of market failure on the defensive (Goetz 1991). If problems of monopoly, externalities, public goods, and bounded rationality afflict private markets, they simply ravage political markets that confront individuals with massive indivisibilities and severely limited exit options. The scene was set for a program of scientific endeavor that would expose government failures and for a program of moral philosophy that would support constitutional reforms designed to restrict the scope and size of government.

The Virginia School does not focus primarily on the vote motive as the fulcrum of political markets, in part because of the paradox of voting implicit in rational ignorance and rational abstentions in large numbers in elections (Rowley 1984), in part because of the lengthy period between elections (Mitchell 1988) and in part because of agenda control problems (Romer and Rosenthal 1978). Instead, a great deal of analysis is focused on the behavior of interest groups, the legislature, the executive, the judiciary and the bureaucracy. The results of such scientific inquiry rarely show the political market in a favorable light. Only through constitutional interventions do Virginians see much prospect of utility-enhancing institutional reforms (Buchanan, Rowley and Tollison 1987).

The Virginia research program analyses the government from the perspective of neoclassical price theory as a vehicle used by rational self-seeking individuals to redistribute wealth (Rowley 1992). In this respect, the protected core of the research program closely resembles that of the Chicago program. Yet, its auxiliary hypotheses – suggestive of widespread government failure – could not be more different.

Important differences in the auxiliary statements of the two programs explain this divergence. Virginia, unlike Chicago, does not assume that individuals are always price takers in political markets; significant discretionary power is recognized. Virginia does not assume, as generally as Chicago does, that political markets clear instantaneously and completely. Virginia does not assume that decision-makers in political markets are always fully informed about the present

or that they are capable of forming rational expectations over the future. Virginia does not excise human error or voter stupidity from its theory of political market behavior (Caplan 2007), and does not ignore institutions in favor of the black-box theory (Rowley 1997, Rowley and Vachris 1994).

That its conditional predictions about the quality of government performance differ so sharply from those of a Chicago program that applies unmodified private market theory to political market analysis is only to be expected.

Political Economics

Unlike the other schools outlined above, Political Economics was never geographically concentrated. Rather its contributors distinguish themselves methodologically from their public choice forerunners. Political economics is a strand of literature that emerged in the mid-1980s, a literature that deals with the same topics as public choice, frequently without recognizing the prior contributions of public choice scholarship (Tollison 2007). In a sense, it reinvents the wheel, while re-shaping that wheel in terms of analytical rigor and alleged generality. Its key contributors, Alesina (1987), Persson and Tabellini (2001, 2003), draw, from the rational expectations revolution in macroeconomics, the inference (Kydland and Prescott (1977)) that not all policies are equally plausible; only those that the policymaker has no incentive to abandon become credible. Constitutional rules are necessary to prevent decision-makers from exercising their discretion in a time-inconsistent manner.

The price of (perhaps excessive) rigor and generality often tends to be a weakening of predictive power. This appears to be the case with Political Economics. Their scholars utilize general equilibrium models and non-cooperative game theory to a greater extent than the other schools. However, econometric studies and experimental studies strongly support the partial equilibrium predictions of the Virginia School. In any event, there is much to be gained by increased interaction between Political Economics and public choice. The public choice literature regularly cites Political Economics contributions. This recognition is not reciprocated by scholars of Political Economics. As Robert Tollison (2007) tersely comments, "the work of Torsten Perrson, Guido Tabellini, Alberto Alesina, Daron Acemoglu and their colleagues is a welcome and strong sign of things to come. As we go forward, however, I hope that we shall not forget Alfred Marshall's dictum that we stand on the shoulders of giants (Tollison 2007, p. 4)."

The Vote Motive

The early contributions to public choice (Black 1948, Downs 1957) viewed the vote motive as a key determinant of political market equilibrium. Black (1948) deduced the median voter theorem whereby competing political candidates would be driven

by vote considerations to converge in policy space to a unique and stable equilibrium that reflected the policy preferences of the median voter.

Downs (1957) reinvented Black's wheel albeit without reference to the median voter. He focused on systems of two party representative governments and demonstrated that vote maximizing politicians would formulate policies to win elections rather than seek political victory in order to implement preferred policies. He also noted the tendency for such political competition to converge to the center of the voter distribution, albeit without distinguishing between the mode, the median and the mean, since he deployed normal distributions throughout his analyses.

This equilibrium offered little discretion to political parties unless they had no serious aspiration to govern. As such, it should have been attractive to those wedded to majority-based political outcomes. In reality, it was anathema to conventional political scientists because of its strict adherence to the rational choice approach.

In the event, the median voter theorem, while still attracting attention among public choice scholars, promised more than it could deliver. It rests on a stringent set of assumptions that coincide only rarely in political markets (Rowley 1984):

1. Two political parties must contest the election;
2. The policies at issue must collapse into one dimension of left-right space;
3. Voter preferences must be single-peaked over policy space;
4. Political parties must be able and willing to move across policy space;
5. Political parties must be well informed regarding the preferred policies of the voters;
6. Voters must be well informed regarding the policy positions of the competing parties;
7. Voters must not abstain in significant numbers from voting in elections;
8. Voters must punish governments that deviate from their electoral manifesto.

Once these assumptions are relaxed, individually or severally, to take account of the realities of political markets, the median solution is much less dominant, especially where the distribution of voter preferences is skewed or multi-modal (Rowley 1984). In some circumstances, the mean dominates the median (Romer and Rosenthal 1979). In others, the political equilibrium cycles in single or in multi-dimensional policy space (Black 1948, Arrow 1951). In yet other circumstances, there is no equilibrium as the political parties become immobilized at separate positions in policy space (Rowley 1984). As a consequence the grip of voter majorities over the election manifestos must be viewed as much looser than either Black or Downs was willing to acknowledge.

Enelow and Hinich (1984) challenged the hypothesis, central to both Black and Downs, that competing political parties (or presidential candidates) are mobile over policy space. Their counter-hypothesis is that political parties are immobilized in the short run by the recent history of their political behavior. In such circumstances, political parties (candidates) must advertise to consolidate the voter preference distribution around their respective positions in policy space. Rationally ignorant voters are vulnerable to such persuasive advertising. To the extent that they are correct, and elections are determined by campaign expenditures, the concept of

revealed voter preferences is rendered suspect and, with it, the underlying connection between political equilibrium and majority-rule politics.

The probability that an individual vote will prove to be decisive in a major election is minute (less than 10^{-6} in U.S. presidential elections (Stigler 1971)). This implies that the differential expected benefit to any voter from voting decisively in an election is also trivial, typically far less than the cost of voting. Only some notion of civil duty or some major miscalculation of probabilities will drive the rational instrumental voter to the polls. Only an active consumption interest will motivate the rational individual to become informed about the political market. Otherwise, he will remain rationally ignorant, whether or not he casts his vote, and will rely on opaque ideology indicators to determine his electoral strategy (Downs 1957). Alternatively, knowing that his vote is indecisive, he will vote expressively, following his heart rather than his interest. This serious consequence of the indivisibility of the vote mechanism opens up tempting avenues for interest groups to invade the political process (for a counter –view, however, see Peltzman 1984).

Elections are discrete events in a continuous political process. The vote motive, at its best, is only as influential as elections are in controlling the post-election behavior of the incumbents (Tullock 1976). Such control is limited by the high rate of voter memory decay that protects deviant governments from adverse electoral consequences. It may be weakened by rational irrationality on the part of voters who choose with deliberation to place their votes irrationally (Caplan 2007). It is surely weakened by the ability of political parties to full-line force into the electoral process policy bundles that intermingle popular with less popular policy proposals.

Once again, all these weaknesses open up opportunities for effective interest groups to divert the supply of policies away from the informed preferences of the median voter.

The Special Interests

A special interest issue is one that generates substantial personal benefits for a small number of constituents while imposing a small individual cost on a large number of other potential voters (Gwartney and Wagner 1988). As James Madison recognized in *The Federalist* (Number 51, 1788), a majority-based system of representative government is biased toward the adoption of special interest (or faction-based) policies, even when such policies are generally harmful to society. The Founding Fathers wrote the separation of powers and the bicameral legislature into the United States Constitution to curtail this perceived political bias. The first ten amendments to the Constitution (The Bill of Rights) were added to the Constitution to protect individuals from the perceived excesses of federal and state governments.

Arguably, these constitutional constraints have failed to hold firm against special interest pressures. Facilitated by a weak Supreme Court, that became increasingly deferential toward the legislative branch of government after 1936, parchment has

ceded victory to the guns of the special interests and has allowed factions to roam freely across constitutional constraints.

Special interests emerge to take advantage of rational ignorance within the legislature, through the mechanism of persuasive campaign contributions, to obtain advantages for their members more than commensurate with their relative voting strength. Their success depends on their relative abilities to offer political gains, in the forms of votes and money, to politicians who broker policies beneficial to concentrated interests and detrimental to diffused interests (Ekelund and Tollison 2001). Legislatures infested with such parasites typically manifest weak party allegiance and relatively high incumbent electoral success rates.

The logic of collective action (Olson 1965) demonstrates that interest groups are far from easy to organize. Because many of the benefits to be derived from effective interest group lobbying have public good or public bad characteristics, free riding by members of the group is rational. Such free riding diminishes the pressure that can be exerted on the legislature. The free riding problem is not dispersed equally, however, across potential interest groups. Some groups, notably in the United States, trade unions and professional groups, are able to coerce supply. Other groups, notably producer groups, successfully engage in collective action in the absence of coercion, because they are small and homogeneous. These groups predictably will be differentially successful in the political process.

Large, diffuse groups confront the free riding problem in its most devastating form. In many instances, for example consumers and taxpayers, they simply cannot form an effective coalition. If such interest groups are to be politically effective, they must organize themselves primarily to provide private (or selective) benefits to the membership, bundling their public objectives into the subscription fee as a by-product of their activities. The by-product solution, coupled with the tax privileged status of most such groups, explains the existence and relative success of organizations active on behalf of the elderly, of abortion rights, of the environment etc., each of which is plagued by public good or public bad characteristics.

To the extent that Olson's (1965) theory is correct, and there is a great deal of accumulated evidence in its favor, the implications for the political process are serious. Interest group pressures will divert the political equilibrium away from the median voter even under circumstances outlined by Duncan Black (1948). Moreover, because such diversions are most effectively achieved through redistributions that are opaque and not transparent, interest group politics will impose high excess burdens on society, as regulations and complex in-kind subsidies are favored over lump sum transfers (Olson 1982).

The logic of collective action constitutes a core element of Virginia Political Economy. It has been challenged, inevitably, by the Chicago School, notably in the scholarship of George Stigler (1974), Gary Becker (1983) and more recently of Donald Wittman (1989, 1995). Gary Becker modeled interest groups within a general equilibrium framework, on the assumption that they can be formed and reorganized at a minimal cost, that their policy benefits, for the most part, are private and not public in nature, and that free riding can be limited by low cost

monitoring. It is not surprising that these assumptions result in a much more benign view of such organizations.

Specifically, Becker suggests that interest groups redistribute wealth efficiently, minimizing the deadweight costs to society. Groups that impose high deadweight excess burdens, in this view, are replaced by more efficient alternatives. This *Panglossian* view has its advocates, mostly from the University of Chicago. The public choice evidence almost universally refutes the predictions of the model (Ekelund and Tollison 2001). There are, for example, virtually no instances of lump sum redistribution in any democracy. Sadly, the post-Friedman Chicago is less interested in positive methodology (Friedman 1953, Lakatos 1978) and more interested in the elegance of theory, in this sense, following the standard bias of modern neoclassical economics.

Rent Seeking and Rent Extraction

Rents are here defined as returns in excess of opportunity cost engineered in a market economy through the regulatory intervention of government (Tollison, 1982, 1997, Rowley, Tollison and Tullock, 1988). The availability of such rents gives rise to rent seeking on the part of interest groups, whose members rationally expend resources in a competition to secure the present value of the rents that are potentially available. Whether rent seeking outlays constitute welfare neutral transfers or whether they constitute welfare-reducing wastes of resources depends on the institutional structure, although the general presumption is that some waste occurs even within a well-designed political system. The extent of rent seeking outlays in response to any given aggregate of available rents depends on attitudes towards risk, the nature of returns to scale in rent seeking and the nature of the rent seeking game (Tullock 1980).

As with so many important contributions to public choice, the original insight came from Gordon Tullock, this time in his seminal 1967 paper in *The Western Economic Journal* challenging the then conventional wisdom that the only loss of welfare from monopoly was the deadweight loss characterized as the *Harberger Triangle* (Harberger 1954). Tullock focused attention on the *Tullock Rectangle* of producer's surplus created as a consequence of monopoly and posed the simply but crucially important question: Would not producers when competing for that monopoly rationally expend aggregate resources, in the limit, equal to the present value of that rent? His positive reaction to that question shook the throne of the new welfare economics, and ultimately destroyed the latter's widely endorsed presumption against free markets.

In 1971, Tullock returned to the theme of his 1967 paper, which as yet had made little headway in mainstream economics, shifting attention to the cost of transfers (Tullock 1971). Drawing from his experience in post-Second World War China, where beggars regularly mutilated themselves as a means of making themselves pitiful to potential donors, Tullock argued that many would be recipients of

government transfers in the Western democracies engaged in similar forms of activity. Rejecting the notion that all political redistribution of wealth is voluntary, Tullock focused attention on the resource cost of competitive lobbying of politicians and bureaucrats, both by those who sought transfers and by those who sought to prevent them. He noted that the resources invested in such activities were socially wasteful, irrespective as to whether the transfers were made or not.

By recognizing that government regulatory activities are endogenous, the self-seeking response to resource outlays by influential pressure groups, Tullock explicitly challenged the public interest theory of government. In 1974, Anne Krueger coined the term rent seeking to characterize these activities, a term that would take a central place in the public choice litany. The rent seeking insight plays a central role in Virginia Political Economy, suggesting as it does that there are significant welfare costs to government activity. By the same coin, the concept presents a fundamental challenge to Chicago School notions that democracies are efficient and that the cost of redistribution does not exceed the normal cost of government.

Indeed, in recognizing that successful rent seeking results in a transitional gains trap that obstructs efficient economic reforms (Tullock 1975), the research program explains why clearly inefficient regulations remain on the statute books and offers a clear warning to those who are rationally well-informed to work hard to obstruct the passing of new regulations, however attractive the latter may appear to be.

A recent empirical study by Laband and McClintock (2001) suggests that, for the United States, supposedly a relatively efficient constitutional republic, the annual cost of rent seeking and rent protection amounts at least to $400 billion. Evidently, this is not the normal cost of government, even if rent seeking continues to be downplayed or ignored entirely by Chicago economists (Laband and Sophocleus 1988).

The mainstream rent-seeking literature assumes that politicians are passive brokers of policies that create rents, responding to rent-seeking bids essentially as auctioneers bent on maximizing the size of their brokerage fees. A more recent literature (McChesney 1987, 1997, 2001) presents a yet more dismal picture, by recognizing that politicians may abandon or modify brokerage roles in order to obtain yet more lucrative returns by threatening adverse legislation unless they are paid off in terms of protection money. Rent extraction, as this Mafia-like protection racket is labeled, is very costly in terms of resource misallocation. Yet, like 'the dog that did not bark' in the Sherlock Holmes novel *Silver Glaze*, it does not manifest itself at all in the public accounting system. Even should it be revealed, the politicians who benefit from it, unlike members of *La Cosa Nostra*, are largely immune from investigation and punishment through the criminal law.

The Legislature

Under conditions of democracy, elected politicians serve in legislatures for specified or flexible terms as representatives of the electorate. Legislatures typically are either unicameral or bicameral in nature. They may or may not be constrained by written or conventional constitutional rules.

Organized on the basis of political parties, or coalitions of parties, or committees and sub-committees, politicians essentially are political brokers, pairing demanders and suppliers of legislation, i.e., those willing to pay most for a particular law or transfer with those who are willing to pay the least to prevent such a law or transfer. Typically, such politician-brokers concentrate on legal arrangements that benefit well-organized and concentrated groups at the expense of diffuse interests, each of which latter is taxed a little to fund the transfer or legislation (Tollison 1988).

Although politicians have ideologies of their own, competition among them, together with the contestability of their positions, constrains their ability to pursue such ideologies unless they conform to those of the constituents who secured their election. Of course, that does not imply that some politicians will not risk an election loss by pursuing a goal to which they are especially attracted. Nor does it imply that politicians will misjudge the will of the electorate on ideologically charged issues. Fundamentally, however, in the public choice perspective, politicians are brokers and not purveyors of policy (Rowley 1992).

Politicians expend resources in specific wealth transfer markets in return for brokerage fees that typically take the form of some mixture of campaign contributions, post-political career remuneration and promised votes. The size and continuity of such brokerage fees depend significantly upon the perceived durability of the wealth transfers. Durability, in a political system characterized by cycles, depends upon institutional constraints designed to protect the status quo. Such constraints vary significantly across the world's democratic legislatures. However, both politicians and interest groups share a common interest in promoting institutional arrangements that enhance the durability of laws (Tollison 1988).

In Westminster models of parliamentary democracy, where parliament is supreme and there is no effective separation of powers, durability of laws in a polity characterized by cycles is not easy to achieve. In such systems, the executive branch of the government, the prime minister and the cabinet, are drawn from the elected legislature and are dependent for their continuation in office on the majority support of the legislature. The cabinet possesses the agenda power in preparing legislation, but this is modified by the ongoing threat that alienated members of the majority party may withdraw parliamentary support and force the government to resign. Coalition governments, typical in many democracies in Continental Europe, are yet more vulnerable to cycles. Predictably, campaign contributions will be relatively low and interest group activity relatively less forceful, ceteris paribus, under all such systems of government than under more severely constrained models.

The United States legislature is just such a constrained model, both by constitutional design and by evolved institutional structure. Its bicameral format increases the difficulty both of passing and of repealing laws. The separation of powers allows for bills passed by both chambers to be vetoed by the President, forcing it back onto two-third supra-majority votes to override the veto. The independent federal judiciary patrols the borders of its legislation, in principle at least, to ensure that the Constitution has not been infringed. These constitutional constraints arguably enhance the durability of its determinations (Landes and Posner 1975, Anderson, Shughart and Tollison 1988, Tollison 1988, Mueller 1996).

In an alternative explanation as to why legislative equilibrium tends to be stable and durable, Shepsle (1978) and others focus attention on the role of committees and the nature of committee assignments in both chambers of the United States Congress, but more especially in the House, as coalition-builders. In this analysis, committees substitute for more vulnerable logrolling (Tullock 1959) in overcoming the high transaction costs of contracting in political markets. They do so by providing a division of labor within the legislature, in which representatives specialize in the political issues relevant to their own districts.

The committee structure of Congress is grounded in 'property rights' granted to each committee of jurisdiction, allowing it almost exclusive rights in 'gate keeping' (i.e., in deciding whether or not to allow potential bills onto the floor of the chamber). It is also grounded in the 'deference' accorded to each committee by non-committee members, grounded both in reciprocity, in threat, and in the power of 'ex post settling up' accorded to committees with jurisdiction by the convention that conference committees between the two chambers are manned primarily by members of those original committees (Shepsle and Weingast 1981). In such circumstances, committees can protect the status quo, or their own bills from non-empty win-sets thereby providing protection against cycling in an environment of instability.

Despite the growing literature based on the new institutional economics that focuses attention on the gains-from-trade aspect of 'politics-as-it-is', there is another, darker side of the legislative process that must not be ignored. Politics is not primarily concerned with gains-from-trade, but with obtaining power over the public authority (Moe 1987, 1990). When two poor voters and one rich voter comprise the electorate, the rich voter is in trouble. He is not in trouble because of political cycles and the instability of the political process. He is in trouble because the poor majority may decide to steal his wealth, using the political process as a less costly mechanism than simple theft. To the extent that the legislative process is concerned more with redistribution than with wealth creation, the fear of the rich voter must be increased.

Because there are no long-term property rights in public authority, the governing political party must exercise caution when creating institutions. Political opponents, should they access the power of those institutions, may deploy that power to reward their own constituents. For this reason, the agencies of government are often tightly constrained by legislation, or even designed to fail in their expressed purposes (Moe 1990).

The Presidency

In countries exemplified by the United States, where the separation of powers is enshrined in the Constitution, the president is elected by an independent vote and holds his position for a fixed term quite independently from the wishes of the majority of the legislature. The United States Constitution arms the president with a

veto power over bills emanating from the legislature. To override the presidential veto, each chamber of the Congress must re-pass the vetoed bill with at least a two-third supra-majority vote. The veto threat effectively allows the president to serve as a third chamber of the legislature, logrolling with the other chambers in the shaping of legislation (Carter and Schap 1987). The president also enjoys significant regulatory powers delegated to him by Congress. These powers can be utilized to reward or punish legislators who choose to pursue goals contrary to the preferences of his key constituents in the Electoral College (Moe 1987).

Potential differences in the interest group constituencies of the Congress and the president emanate in part from their different bases of representation. Special interests are much more effective when targeting their rent seeking on the specialized districts of the House than they are state-wide in the Senate. They are least effective in targeting the nation-wide constituency of the president. The special interests are most effective when working in opaque environments (Crew and Rowley 1988). Presidential politics are much more transparent than congressional politics.

One view that has some currency in the public choice literature (Crain and Tollison 1979a,b) is that the President and Congress override the separation of powers and the intent of the Founding Fathers and impose a collusion of powers designed to enhance the durability of legislation and thus to raise the brokerage fees provided by the special interests. While this perspective has some credibility when the presidency and the Congress are both controlled by a single political party, it is difficult to justify when the party of the president does not control the Congress.

When the Congress and the president are at odds with each other, it is by no means clear which branch will dominate. Madison (1788) (*The Federalist*, No. 53) envisaged the legislature as the dominant branch and worried about the power that this would accord to factions. Powerful presidents, (most notably Ronald Reagan) however, have influenced policy even when their parties have been in a legislative minority. Certainly, presidents are able to destabilize a political equilibrium when the policies at issue are of high priority to them and transparent to the voting public (Rowley 1992).

The Judiciary

The United States federal judiciary was established by the Founding Fathers as a co-equal independent branch of government designed to function as an effective counter-weight to the legislative and the executive branches. To limits its powers, the federal judiciary is dependent on the president and the Congress for its appointments, dependent on the executive branch for enforcing its judgments, and on the Congress for appropriating its budget. Within these constraints, however, the judiciary patrols the behavior of the executive and the legislative branches to ensure that the Constitution is not violated.

To secure independence, federal judges are granted lifetime tenure, albeit subject to the sanction of impeachment. Their salaries cannot be reduced in nominal

terms during their tenure. Their judgments, especially those of the Supreme Court, are accorded enormous respect even when they run counter to majority popular opinion. Even so, the federal judiciary has not escaped public choice scrutiny.

Because judges and justices are appointed through a political process, it is extremely unlikely that the 'best and the brightest' will be successful in securing appointment. Typically, such individuals will have delivered judgments deemed too controversial for the politically sensitive ears of the politicians. Potential appointees are scrutinized closely in terms of ideological bias, past political service and campaign contributions.

Where the president's party controls the Senate, and thus the Judiciary Committee, candidates of the alternative political persuasion will not be nominated. Only stealth candidates who provide a false image of their views (notably in recent years, Justice Souter) will wriggle through the selection process. Where the party of the president does not control the Senate, serious candidates will have to display mediocre intellects and enhanced deference to the legislature (in recent years, Justice Kennedy is a good example) or to be willing to play the race card (Justice Thomas).

The interest-group theory of politics (McCormick and Tollison 1981) models legislatures as firms supplying wealth transfers to competing interest groups by means of contracts termed 'laws'. In one interpretation (Anderson 2001), the judiciary confirms such contracts by adjudicating in favor of the short-term interests of pressure groups who successfully bid for political influence. As the balance of pressure groups changes, the courts will shift their judgments, irrespective of the original intent of the legislation.

An alternative interpretation (Landes and Posner 1975) focuses on the long-run effects of judicial independence, arguing indeed that such independence may be an integral component of the interest-group theory of government. They argue that the function of judges is to provide stability to the bargains agreed between the legislature and organized pressure groups, thus increasing the value of the rents that are dispersed. Precisely because the judiciary is independent from the current legislature, the judiciary is able to resolve disputes concerning the interpretation or constitutionality of a law by reference to the intentions of the originally enacting legislative body. Landes and Posner (1975) provide weak empirical support for this proposition. The proposition remains suspect, however, because such long-run stabilization of contracts inevitably reduces the prospects for the forging of new contracts (Benson 2001). Legislators who control the budget appropriations to the judiciary are unlikely to allow strong judicial independence where it threatens their current brokerage fees in the rent-seeking market.

Bureaucracy

The bureaucracy of government, responsible for the implementation of policies that are legislated and signed by the president into law, is located in the executive branch of the government. However, bureaus are dependent on the legislature for

budget appropriations, are subject to its oversight authority, and are vulnerable to new legislation where their activities place them at odds with the current legislative majority.

The traditional political science perspective envisages senior bureaucrats as being impartial and public-spirited in the sense of pursuing either the original intent of the legislation that created their bureaus or the current wishes of their legislative overseers. This perspective has been closely scrutinized by public choice scholars who have focused on the rational choice approach in which senior bureaucrats are viewed as maximizing their personal utilities subject to relevant institutional constraints.

Within the public choice perspective, the senior bureaucrats who exercise authority over the budget are viewed as self-seekers whose goals are defined as some balance between expected wealth, ideology, patronage, discretionary power and ease of management (Tullock 1965, Downs 1967, Niskanen 1971). Budget maximization (Niskanen 1971), or alternatively discretionary budget maximization (Niskanen 1975, 2001) is deployed as a plausible proxy for these various objectives. Senior bureaucrats commit a total output in return for a total budget appropriation.

In negotiating budgets with legislative committees, senior bureaucrats are viewed as possessing information advantages because of the monopoly of their provisions. Their legislative overseers have little access to independent information regarding the bureau's production function. Because of the indivisible nature of the budgetary negotiations, senior bureaucrats are able to operate as discriminating monopolists, extracting the total surplus from the legislature (Niskanen 1971).

The nature of the budgetary outcome under these bargaining conditions depends on a number of factors. First is the nature of the budgetary environment, specifically, whether the bureau is demand-constrained or budget constrained (Niskanen 1971). In circumstances of relaxed oversight, or demand constraint, the budget-maximizing bureau simply maximizes the size of its bureau, unconstrained by output constraints. In circumstances of tightened oversight, or budget constraint, bureaucrats maximize the size of their budget at a lower level than would be possible under conditions of demand constraint.

In both circumstances, the output of the bureau is significantly higher than what the median voter would prescribe. In the former case, the bureau is additionally technically inefficient, supplying its output at costs significantly higher than those minimally available to it. In the latter case, the bureau is technically efficient according to this model (Niskanen 1971).

Second, if discretionary budget maximization replaces maximization of budget size in the utility functions of senior bureaucrats, the outcome of budget negotiations changes. Senior bureaucrats no longer negotiate deals that extend output beyond that optimally demanded by the legislature. Instead they focus their attention on maximizing the budget surplus that can be deployed in pursuit of other goals (Niskanen 1975). A key implication of this outcome is that bureaus are always technically inefficient, securing budgets significantly in excess of the minimal cost of providing output even if the level of their output is not in excess of the optimal requirements of the oversight committee.

Members of the bureaucracy predictably enter the political market place on the demand as well as on the supply side as special interests that are unconstrained by free-rider considerations (Rowley, Shughart and Tollison 1987). They tend to be differentially well-informed concerning the predictable response of legislators to specific initiatives. They are rationally well informed concerning the policies that their bureaus will administer. Predictably, senior bureaucrats favor non-transparent policy initiatives, not only to conceal special interest allocations from electoral scrutiny, but also to maximize their own discretionary power in the provision of commodities subject to their control (Crew and Rowley 1988).

In critical response to Niskanen's seminal works (1971, 1975), the public choice analysis of bureaus has reverted somewhat from his theory of bureau dominance to a view that oversight and appropriation committees exercise significant control and that bureaus respond to a significant degree to the dictates of their political masters (Weingast and Moran 1983). The congressional dominance theory assumes that congressmen on relevant committees possess sufficient incentives and sufficient sanctions to establish effective governance over the agencies that they monitor.

The federal bureaus and agencies established by statute usually, though not universally, are subject to oversight both by the Congress and by the president. Their budgets are appropriated by both chambers of the Congress but are subject to review and potential veto by the president. In such circumstances, it is relevant to analyze bureaucratic behavior from the perspective of a multiple principal–agent relationship (Rowley and Vachris 1993).

The congressional dominance theory (Weingast and Moran 1983) assumes that congressmen on the relevant oversight and appropriations committees possess sufficient incentives and sufficient sanctions to establish governance over the agencies that they monitor. Although the committees are not endowed with sufficient resources to engage in continuous monitoring, special interests keep them well informed about agency performance. By choking off appropriations to recalcitrant agents, by harassing them through the oversight process, and by threatening interventionist legislation, congressional committees are viewed as influential monitors. The threat of ex post sanctions and the promise of ex post settling up create ex ante incentives for agents to reflect the preferences of the majority vote on the relevant congressional committees.

The hub of the efficient governance hypothesis is the assumption that congressional committees exercise a near monopoly jurisdiction over their respective agents, thus benefiting from clearly defined property rights that encourage circumspect monitoring. To the extent that congressmen self-select the committees on which they serve, the near monopoly power that they access provides leverage over precisely those issues relevant to their individual political support and, hence, to their expectations of re-election.

If this hypothesis is correct, there are two testable predictions that should survive empirical testing, namely (1) that specific oversight/appropriations committees should exercise more influence than Congress as a whole over the behavior of particular agents and (2) that if the political complexion of a particular committee should shift, then so should the politically relevant behavior of the associated agent.

Early tests have not refuted either of these hypotheses (Weingast and Moran 1983, Weingast 1984, Grier 1991).

Nevertheless, because of the competition among multiple principals for agency control, agents will not be efficiently monitored. Considerable agency discretion will survive (Rowley and Vachris 1993). The multiplicity of principals arises from at least four sources, namely (1) jurisdictional overlaps among oversight committees in each chamber of Congress, (2) duality of oversight responsibilities in a bicameral legislature, (3) jurisdictional conflicts between oversight and appropriations committees composed of different members and (4) the competing jurisdictions of the Congress and the President, especially when the Congress and the presidency are controlled by different political parties (Rowley and Vachris 1993).

Constitutional Political Economy

According to Buchanan (1990) there is a 'categorical distinction' to be made between constitutional economics and ordinary economics, a distinction in the ultimate behavioral object of analytical attention (*ibid*., p. 2). In ordinary economics, analysis is concentrated on choices made *within* constraints that are imposed exogenously on the person or persons making that choice. Constitutional economics, in contrast, directs analytical attention to the choice *among* constraints, choices that are made ex ante by individuals in seeking to restrict their own and others' subsequent choice sets in the ordinary political sphere.

The seminal contribution to constitutional political economy is *The Calculus of Consent*, co-authored by Buchanan and Tullock (1962). This book outlined for the first time an individualistic theory of the constitution, assigning a central role to a single decision-making rule – that of general consensus or unanimity.

By focusing attention on the nature of expected external costs and expected decision-making costs under decision-rules short of unanimity, and by recognizing that constitutional rules are derived under conditions of generalized uncertainty, Buchanan and Tullock explained why rules of less than unanimity (not necessarily a simple majority rule) would be chosen unanimously by rational individuals at the constitutional stage: 'At best, majority rule should be viewed as one among many practical expedients made necessary by the costs of securing widespread agreement on political issues when individual and group interests diverge' (Buchanan and Tullock 1962).

The hard core of the constitutional political economy research program combines the assumptions of rational choice, methodological individualism and *homo economicus* in a manner that distinguishes it sharply from all mainstream economic research programs designed to evaluate the nature and the role of the state (Brennan and Hamlin 2001). Over the following forty years, the auxiliary assumptions of the model have adjusted to reflect changing circumstances. Those working within the field, however, have not found it necessary to adjust the hard-core assumptions.

As the political environment in the United States deteriorated from the rosy scenario of the second Eisenhower administration through the civil rights crisis, and

the Vietnam fiasco of the Kennedy and Johnson administrations, culminating in the Watergate crisis of the Nixon administration, Buchanan in particular became less enamored of the positive gains-from-trade approach of *The Calculus of Consent*. In *The Limits of Liberty* (1975), he effectively deployed the philosophy of Thomas Hobbes – the threat of beckoning anarchy – to protect the hard core of his research program in a much less favorable external environment. From this insight came some of the best scholarship of the program, most notably *Democracy in Deficit* (Buchanan and Wagner 1977).

There then occurred through the decade of the 1980s a shift of direction from science to moral philosophy as Brennan and Buchanan (1980, 1985) injected propositions from John Rawls (1971) into the protective belt of their theory. With the breakdown of the Soviet Empire in the early 1990s, scholars recognized that Rawls's 'veil of ignorance' played no role in the process of constitution-making that followed in the wasteland left behind by socialism. In 1990, Buchanan returned to science in an important paper introducing his new journal, *Constitutional Political Economy*. Since then the constitutional political economy research program has proceeded successfully along the rational choice lines from whence it originated (Mueller 1996; Brennan and Hamlin 2001), albeit with a shift in the direction of formal modeling and empirical testing (Persson and Tabellini 2003).

Bioeconomics of Non-Human Societies

Innovative public choice scholarship is extending the frontiers of the discipline well beyond the domain of rational economic man to encompass the behavior of other species, notably bees and fishes. Janet Landa, a law-and-economics scholar and a prominent bio-economist, has written two important papers (1986) dealing with these species.

Her paper on the political economy of swarming in honeybees offers a fascinating study of collective action in biological systems. Landa explains the organization of bee swarming as a means whereby honeybees economize on information and decision-making costs when establishing a new nest site. She uses the Buchanan and Tullock (1962) theory of the choice of Pareto-optimal voting rules to explain why scout bees use the unanimity rule when deciding where to establish a new nest.

On one hand, the external costs of using the 'any bee' rule would be very high for the whole bee swarm should the one bee find an unsuitable home. On the other hand, the decision-making costs of using the unanimity rule are low both because scout bees constitute only about 5 per cent of the whole swarm and because they are a homogeneous group, being experienced former foragers. Because of the high external costs relative to decision- making costs, the use of the rule of unanimity by scout bees is efficient.

Just as honeybees 'vote with their wings' (Landa 1986), when swarming out of their nest in search of a new home, so also many species of fish 'vote with their fins'

(Landa 1998), when forming schools in order to migrate to spawn, to search for new foraging areas, to hunt for prey and to organize for defense. In her 1998 paper, Landa applies public choice analysis to the biological literature on schooling fish, using a selfish fish, club-theoretical paradigm.

On this basis she hypothesizes that a selfish fish (a) joins the fish school because it derives hydrodynamic benefits (a club good), (b) has no incentive to completely free-ride because it will be left behind by the school if it attempts to do so, (c) has no incentive to shirk a leadership role because of role reversibility between leaders and followers, (d) derives defense benefits against predators from its membership of the school, and (e) has no incentive to discriminate against odd-looking outsiders since odd-looking fish in a school are attacked more frequently by predators than are look-alikes. On the other hand, outsiders display xenophobia towards insiders because outsiders do not wish to become prime targets for predators. As a consequence, fish schools tend to be homogeneous.

Finally, Landa applies the Buchanan and Tullock (1962) theory of choice of optimal voting rules to explain why the 'any leader' rule for making the collective choice to escape, the main anti-predator defense strategy, is optimal for members of a fish school. In so doing, Landa explains the leaderless, completely decentralized form of organization of fish schools, in contrast to bee swarms. Evidently, the reach of *The Calculus of Consent* extends well beyond Homo sapiens into the bio-economics of non-human societies.

References

Alesina, A. (1987). 'Macroeconomic policy in a two party system as a repeated game'. *Quarterly Journal of Economics*, 102, 651–678.

Anderson, G. (2001). 'The judiciary' In W.F. Shughart and L. Razzolini (eds.) *The Elgar Companion to Public Choice*. Cheltenham, UK: Edward Elgar Publishing, pp. 293–309.

Anderson, G.A., Shughart, W.F. and Tollison, R.D. (1988). 'On the incentive of Judges to enforce legislative wealth transfers'. *Journal of Law and Economics*, 31, 215–228.

Arrow, K.J. (1950). 'A difficulty in the concept of social welfare'. *Journal of Political Economy*, 58, 328–346.

Arrow, K.J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Becker, G.S. (1976). 'Comment on Peltzman'. *Journal of Law and Economics*, 19(2), 245–248.

Becker, G.S. (1983). 'A theory of competition among pressure groups for political influence'. *Quarterly Journal of Economics*. 63, 371–400.

Benson, B. (2001). 'Law and economics' In W.F. Shughart and L. Razzolini (eds.) *The Elgar Companion to Public Choice*. Cheltenham, UK: Edward Elgar Publishing, pp. 547–589.

Black, D. (1948). 'On the rationale of group decision-making'. *Journal of Political Economy*. 56, 23–34.

Brennan, H.G. and Buchanan, J.M. (1980). *The Power to Tax*. Cambridge: Cambridge University Press.

Brennan, H.G. and Buchanan, J.M. (1985). *The Reason of Rules*. Cambridge: Cambridge University Press.

Brennan, H.G. and Hamlin, A. (2001). 'Constitutional choice.' In W.F. Shughart and L. Razzolini (eds.) *The Elgar Companion to Public Choice*. Cheltenham, UK: Edward Elgar Publishing, pp. 117–139.

Buchanan, J.M. (1975). *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: Chicago University Press.

Buchanan, J.M. (1990). 'The domain of constitutional economics'. *Constitutional Political Economy*, 1(1), 1–18.

Buchanan, J.M., Rowley, C.K. and Tollison, R.D. (1987). *Deficits*. Oxford: Basil Blackwell.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Buchanan, J.M. and Wagner, R.E. (1977). *Democracy in Deficit: The Political Legacy of Lord Keynes*. New York: Academic Press.

Carter, J.R. and Schap, D. (1987). 'Executive veto, legislative override and structure-induced equilibrium'. *Public Choice*, 52, 227–244.

Coase, R.H. (1960). 'The problem of social cost'. *Journal of Law and Economics*, 3, 1–44.

Caplan, B. (2007). *The Myth of the Rational Voter*, Princeton, New Jersey: Princeton University Press.

Crain, W.M. and Tollison, R.D. (1979a). 'Constitutional change in an interest group perspective'. *Journal of Legal Studies*, 8, 156–175.

Crain, W.M. and Tollison, R.D. (1979b). 'The executive branch in the interest-group theory of government'. *Journal of Legal Studies*, 8, 555–567.

Crew, M.A. and Rowley, C.K. (1988). 'Toward a public choice theory of regulation'. *Public Choice*, 57, 49–67.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Downs, A. (1967). *Inside Bureaucracy*. Boston: Little Brown.

Ekelund, R.B. and Tollison, R.D. (2001). 'The interest-group theory of government'. In W.F. Shughart and L. Razzolini (eds.) *The Elgar Companion to Public Choice*. Cheltenham, UK: Edward Elgar Publishing, pp. 357–378.

Enelow, J.M. and Hinich, M.J. (1984). *The Spatial Theory of Voting*. Cambridge: Cambridge University Press.

Fiorina, M. (1983). 'Flagellating the Federal Bureaucracy'. *Society*, 66–73.

Friedman, M. (1953). 'The methodology of positive economics.' In M. Friedman, *Essays in Positive Economics*. Chicago: University of Chicago Press, pp. 3–43.

Friedman, M. (1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.

Goetz, C.J. (1991). *Uncommon Common-Sense vs. Conventional wisdom: The Virginia School of Economics*. Fairfax: Center for Study of Public Choice.

Grier, K. (1991). 'Congressional influence on US monetary policy: An empirical test'. *Journal of Monetary Economics*, 28, 202–220.

Gwartney, J. and Wagner, R.E. (1988). *Public Choice and Constitutional Economics*. Greenwich: JAI Press.

Harberger, A.C. (1954). 'Monopoly and resource allocation'. *American Economic Review*, 44, 77–87.

Hotelling, H. (1929). 'Stability in competition'. *American Economic Review*, 39, 41–57.

Kydland, F. and Prescott, E. (1977). 'Rules rather than discretion: The inconsistency of optima plans'. *Journal of Political Economy*, 85, 473–491.

Laband, D.N. and McClintock, G.C. (2001). *The Transfer Society*. Washington, DC: The Cato Institute.

Laband, D.N. and Sophocleus, J.P. (1988). 'The social cost of rent seeking: First estimates'. *Public Choice*, 58, 269–276.

Lakatos, I. (1978). *The Methodology of Scientific Research Programs*. Cambridge: Cambridge University Press.

Landa, J. (1986). 'The political economy of swarming in honeybees: Voting- with-wings, decision-making costs, and the unanimity rule'. *Public Choice*, 51, 25–38.

Landa, J. (1998). 'Bioeconomics of schooling fishes: Selfish fish, quasi-free riders, and other fishy tales'. *Environmental Biology of Fishes*, 53, 353–364.

Landes, W.E. and Posner, R.A. (1975). 'The independent judiciary in an interest group perspective'. *Journal of Law and Economics*, 18, 875–902.

Landes, W.E. and Posner, R.A. (1987). *The Economic Structure of Tort Law*. Cambridge: Cambridge University Press.

Madison, J. (1788). *The Federalist* 51, Publius, February 6.

Madison, J. (1788). *The Federalist* 53, Publius, February 9.

McChesney, F.S. (1987). 'Rent extraction and rent creation in the economic theory of regulation'. *Journal of Legal Studies*, 16, 101–118.

McChesney, F.S. (1997). *Money for Nothing: Politicians, Rent Extraction, and Political Extortion*. Cambridge, MA: Harvard University Press.

McChesney, F.S. (2001). 'Rent seeking and rent extraction.' In W.F. Shughart and L. Razzolini (eds.) *The Elgar Companion to Public Choice*. Cheltenham, UK: Edward Elgar Publishing, pp. 379–395.

McCormick, R.E. and Tollison, R.D. (1981). *Politicians, Legislation and the Economy: An Inquiry into the Interest-Group Theory of Government*. Boston: Martinus Nijhoff.

Mitchell, W.C. (1988). 'Virginia, Rochester and Bloomington: Twenty-five Years of public choice and political science'. *Public Choice*, 56, 101–120.

Mitchell, W.C. (1989). 'Chicago political economy: A public choice perspective'. *Public Choice*, 63, 282–292.

Mitchell, W.C. (2001). 'The old and the new public choice: Chicago vs. Virginia.' In W.F. Shughart and L. Rozzalini (eds.) *The Elgar Companion to Public Choice*. Cheltenham, UK: Edward Elgar Publishing, pp. 3–30.

Moe, T.N. (1987). 'An assessment of the positive theory of congressional dominance'. *Legislative Studies Quarterly*, 12, 475–520.

Moe, T.N. (1990). 'Political institutions: The neglected side of the story'. *Journal of Law, Economics and Organization*, 6, 213–253.

Mueller, D.C. (1996). *Constitutional Democracy*. Oxford: Oxford University Press.

Mueller, D.C. (ed.) (1997). *Perspectives on Public Choice*. Cambridge: Cambridge University Press.

Niskanen, W.A. (1971). *Bureaucracy and Representative Government*. New York: Aldine-Atherton.

Niskanen, W.A. (1975). 'Bureaucrats and Politicians'. *Journal of Law and Economics*, 18, 617–643.

Niskanen, W.A. (2001). 'Bureaucracy', In W.F. Shughart and L. Razzolini (eds.) *The Elgar Companion to Public Choice*. Cheltenham, UK: Edward Elgar Publishing, pp. 258–270.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Olson, M. (1982). *The Rise and Decline of Nations: Economic Growth, Stagflation and Social Rigidities*. New Haven: Yale University Press.

Peltzman, S. (1984). 'Constituent interest and congressional voting'. *Journal of Law and Economics*, 27, 181–210.

Persson, T. and Tabellini, G. (2001). *Political Economics*. Cambridge: MIT Press.

Persson, T. and Tabellini, G. (2003). *The Economic Effects of Constitutions*. Cambridge: MIT Press.

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Belknap Press.

Reder, M.W. (1982). 'Chicago economics: Permanence and change'. *Journal of Economic Literature*, 20, 1–38.

Riker, M.W. (1962). *The Theory of Political Coalitions*. New Haven: Yale University Press.

Riker, M.W. (1964). *Federalism: Origins, Operation Significance*. Boston: Little Brown.

Riker, W.H. and Ordeshook, P.C. (1973). *An Introduction to Positive Political Theory*. Englewood Cliffs: Prentice Hall.

Romer, T. and Rosenthal, H. (1978). 'Political resource allocation: Controlled agendas and the status quo'. *Public Choice*, 33, 27–43.

Romer, T. and Rosenthal, H. (1979). 'The elusive median voter'. *Journal of Public Economics*, 12, 143–170.

Rowley, C.K. (1984). 'The relevance of the median voter theorem'. *Journal of Institutional and Theoretical Economics*, 140, 104–135.

Rowley, C.K. (1992). *The Right to Justice*. Aldershot, UK: Edward Elgar Publishing.

Rowley, C.K. (1993). 'Introduction' In C.K. Rowley (ed.) *Public Choice Theory*, Volume 1, Aldershot, UK: Edward Elgar Publishing, pp. ix–xxix.

Rowley, C.K. (1997). 'Donald Wittman's the myth of democratic failure', *Public Choice*, 92, 15–26.

Rowley, C.K. (2001). 'The international economy in public choice perspective' In W.F. Shughart and L. Razzolini (eds.) *The Elgar Companion to Public Choice*. Cheltenham UK: Edward Elgar Publishing, pp. 645–672.

Rowley, C.K. and Schneider, F. (eds.) (2004). *The Encyclopedia of Public Choice*. Dordrecht: Kluwer Academic Publishers.

Rowley, C.K., Shughart, W.F. and Tollison, R.D. (1987). 'Interest groups and deficits.' In J.M. Buchanan, C.K. Rowley and R.D. Tollison (eds.) *Deficits*. Oxford: Basil Blackwell, pp. 263–280.

Rowley, C.K., Tollison, R.D. and Tullock, G. (eds.) (1988). *The Political Economy of Rent Seeking*. Dordrecht: Kluwer Academic Publishers.

Rowley, C.K. and Vachris, M.A. (1993). 'Snake oil economics versus public choice' In C.K. Rowley (ed.) *Public Choice Theory*, Volume 3, Aldershot, UK.: Edward Elgar Publishing, pp. 573–84.

Rowley, C.K. and Vachris, M.A. (1994). 'Why democracy does not necessarily produce efficient results', *Journal of Public Finance and Public Choice*, 12, 95–111.

Shepsle, K.A. (1978). *The Giant Jigsaw Puzzle*. Chicago: Chicago University Press.

Shepsle, K.A. (1982). 'The budget: Will a constitutional amendment help?' *Challenge*, 25, 53–56.

Shepsle, K.A. and Weingast, B.R. (1981). 'Structure-induced equilibrium and legislative choice'. *Public Choice*, 37, 503–520.

Shughart, W.F. and Razzolini, L. (eds.) (2001). *The Elgar Companion to Public Choice*. Cheltenham, UK.: Edward Elgar Publishing.

Stigler, G.J. (1971). 'The theory of economic regulation'. *Bell Journal of Economics and Management Science*, 2, 137–146.

Stigler, G.J. (1988). *Chicago Studies in Political Economy*. Chicago: Chicago University Press.

Stigler, G.J. (1992). 'Law or economics', *Journal of Law and Economics*, 35, 455–468.

Tollison, R.D. (1982). 'Rent seeking: A survey'. *Kyklos*, 35, 575–602.

Tollison, R.D. (1988). Public choice and legislation'. *Virginia Law Review*, 74, 339–371.

Tollison, R.D. (1989). 'Chicago political economy'. *Public Choice*, 63 (3), 293–298.

Tollison, R.D. (1997). 'Rent seeking.' In D.C. Mueller (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 506–525.

Tollison, R.D. (2007). 'Old wine, new wine'. *Public Choice*, 132, 3–5.

Tullock, G. (1959). 'Problems in majority Voting'. *Journal of Political Economy*, 67, 571–579.

Tullock, G. (1965). *The Politics of Bureaucracy*. Washington, DC: Public Affairs Press.

Tullock, G. (1967). 'The welfare costs of tariffs, monopolies and theft'. *Western Economic Journal*, 5, 224–232.

Tullock, G. (1971). 'The cost of transfers'. *Kyklos*, 4, 629–643.

Tullock, G. (1975). 'The transitional gains trap'. *The Bell Journal of Economics and Management Science*, 6, 671–678.

Tullock, G. (1976). *The Vote Motive*. London: Institute of Economic Affairs.

Tullock, G. (1980). 'Efficient rent-seeking' In J.M. Buchanan, R.D. Tollison and G. Tullock (eds.) *Toward the Theory of the Rent-Seeking Society*. College Station: Texas A & M University Press, pp. 97–112.

Tullock, G. (1981). 'Why so much stability?' *Public Choice*, 37, 189–202.

Tullock, G. (1991). 'Duncan Black: The founding father'. *Public Choice*, 71, 125–128.

Weingast, B.R. (1984). 'The Congressional Bureaucratic System: A principal-agent perspective'. *Public Choice*, 44, 147–191.

Weingast, B.R. and Moran, N.J. (1983). 'Bureaucratic discretion or congressional control?' *Journal of Political Economy*, 91, 765–800.

Wittman, D. (1989). 'Why democracies produce efficient results'. *Journal of Political Economy*, 97, 1395–1424.

Wittman, D. (1995). *The Myth of Democratic Failure: Why Political Institutions are Efficient*. Chicago: University of Chicago Press.

Chapter 2
Public Choice: An Introduction[1]

Dennis C. Mueller

Origins

Public Choice has been defined as the application of the methodology of economics to the study of politics. This definition suggests that public choice is an inherently interdisciplinary field, and so it is. Depending upon which person one selects as making *the* pioneering contribution to public choice, it came into existence either in the late 18th century as an offshoot of mathematics, or in the late 1940s as an offshoot of economics. The case for the earlier date rests on the existence of publications by two French mathematicians, J.C. de Borda (1781) and M. de Condorcet (1785). Condorcet was the first person, as far as we know, to discover the problem of *cycling*, the possibility when using the simple majority rule that an alternative x can lose to y in a vote between the two, y can lose to another alternative z, but z will also lose to x. The existence of such a possibility obviously raises the issue of how a community can decide among these three alternatives, when a cycle exists, and what the normative justification for any choice made will be. No cycle exists, of course, if some alternative, say y, can defeat both x and z. The literature has commemorated Condorcet's contribution by naming such an issue like y a *Condorcet winner*. A vast number of papers and books have analyzed both the normative and positive implications of the existence of cycles.

Condorcet gave his name to one other important part of the public choice literature, when he proved what he called a theorem about juries, and what we now call the *Condorcet jury theorem*. This remarkable theorem provides both a justification for making collective decisions with the simple majority rule, and for the institution of democracy itself. It rests on three assumptions: (1) The community faces a binary choice between x and y, with only one of the two choices being the "right" choice for the community. (2) Everyone in the community wants to make the right choice. (3) The probability p that a citizen votes for the right choice is greater than 0.5. The theorem states that the probability that the community makes

---

the right choice when it uses the simple majority rule increases as the number of voters increases approaching one in the limit.

That the theorem provides a normative case for the simple majority rule is obvious, if one accepts its premises. Condorcet described the collective decision as one regarding the determination of whether a person had committed a particular crime or not—hence the theorem's name. For this type of collective decision the definition of "the right decision" is fairly controversial—the person is declared innocent only if she is in fact innocent. The assumption that everyone wants to make the right choice in this situation also seems uncontroversial.

The argument that the theorem also provides a justification for democracy is more subtle, and under it the assumptions underpinning the theorem become more controversial. Imagine, however, that everyone in the community agrees that they would like a "good government" that would be honest and provide goods and services and levy taxes so as to maximize the welfare of the community. Two parties compete for the honor of becoming the government, and each citizen votes for the party that he believes will form the best government. If each citizen has a greater than 0.5 probability of picking the party that will form the best government (two-party) democracy chooses the best government in a large electorate with near certainty.

The second and third assumptions take on extreme importance, when the theorem is used as a defense of democracy. Citizens share a common goal—good government. Each citizen has a greater than 0.5 probability of picking the party that will provide the best government. Citizens do not merely flip coins to decide how to vote, they study the parties and make an *informed* choice.

The assumption that everyone agrees on what good government is, becomes more controversial when we are thinking of the whole panoply of things governments do. If citizens disagree about what the government should do, there will be no "right choice" for all citizens. This being the case, parties will compete not only on the basis of how good they will be at advancing the community's welfare, but how that welfare should be defined. Finally, when one is thinking of a large electorate, even the assumption that voters are well-informed becomes controversial.

Many studies in public choice employ some of the assumptions needed to apply the Condorcet jury theorem to the study of politics; many others do not. All the work on party competition that uses "spatial modeling" assumes, for example, that voters are well-informed, that they know the positions of the parties in the issue space. At the same time, however, this literature does not assume that voters agree on where the parties should be located in the issue space. Conflicts of interest or preferences are assumed, and thus voters do not agree on which party is best even when they are certain about what the parties will do in office—assuming, that is, that the parties will do different things. There is another branch of the public choice literature, however, that *does* assume common interests among citizens, and thus does accord with the second assumption underlying the jury theorem. This work often focuses on decisions made at the constitutional stage of the political process and today often goes by the name of constitutional political economy.

Thus, directly or indirectly, Condorcet's pioneering work raised many of the questions with which the modern public choice literature has been concerned.

Do individuals share common interests? Is democracy stable or not (produces cycles)? Are voters sufficiently well-informed that one gains information by aggregating their preferences? What voting rule should be used to aggregate these preferences?[2]

Borda was critical of the use of the simple majority rule to aggregate preferences, and proposed instead a rule which today carries his name. If there are n possible outcomes to a collective decision, each voter assigns a one to his most preferred choice, a two to his second most preferred choice, and so on. The scores awarded are then added across all voters, and the Borda-count rule selects as the winner the alternative receiving the lowest score. With only two alternatives from which to choose, the Borda-count is equivalent to the simple majority rule. When $\bar{n}$ 2, it avoids cycling and has additional desirable properties that make it attractive.[3]

Three more names deserve brief mention before we end this discussion of the forerunners to public choice. Another mathematician, the Reverend Charles L. Dodgson, better known today as Lewis Carroll, wrote a series of pamphlets analyzing the properties of voting procedures roughly a century after the work of Borda and Condorcet.[4] J.S. Mill's *Considerations on Representative Government* (1861) must also be mentioned, since he was one of the great economists of the 19th century, although the work is arguably an early contribution to political science rather than to public choice, as it makes no noticeable use of economic reasoning. Nevertheless, the great thinker's logical mind is quite evident, and it is one of the few works in political science from the 19th century that still warrants reading by students of public choice.

The same can be said of K. Wicksell's (1896) classic essay on *Just Taxation* written as the 19th century was coming to a close. As the title suggests, it is as much or more a contribution to public finance than to the study of politics, but it contains an early normative economic justification for the state, and a spirited defense of the unanimity rule for aggregating individual preferences.

Early Classics

The modern literature on public choice came into being with the publication of articles by D. Black (1948a, b), J.M. Buchanan (1949) and K.J. Arrow (1950) in the late 1940s and 1950. Retrospectively, one can identify three important contributions between Wicksell and Black, namely Hotelling (1929), Schumpeter (1942) and Bowen (1943), but it was Black, Buchanan and Arrow who got the public choice ball rolling.

Duncan Black's two articles, first published in 1948 and then republished with extensions and an interesting account of the history of ideas lying behind his work, take up the problem of cycling under the simple majority rule and provide a proof of

---

[2] For additional discussion on Condorcet and the jury theorem, see Young (1997).

[3] See in particular, Saari (1994).

[4] See discussion in Black (1958, Ch. 20).

the famous median voter theorem. This theorem has been frequently invoked to describe equilibrium in theoretical studies and has been the analytical foundation for much of the empirical work in public choice.

Arrow proved that no procedure for aggregating individual preferences could be guaranteed to produce a complete social ordering over all possible choices and at the same time to satisfy five, seemingly reasonable, axioms. Indirectly Arrow's theorem invoked the problem of cycling again, since one of his axioms was intended to ensure that cycling did not occur. Arrow's 1950 article and 1951 book spawned much controversy and a huge literature.

Although Buchanan published several important articles prior to 1962, it was the book *The Calculus of Consent*, published in that year and coauthored with Gordon Tullock that established Buchanan and Tullock as leading scholars in the field. Although the book contains many interesting discussions of the properties of the simple majority rule, logrolling and the like, its most lasting contribution to the literature has been to introduce the distinction between the constitutional stage of collective decision making in which the voting rules and other institutions of democracy are selected, and the applications of these rules to the actual work of making collective choices.

In *Capitalism, Socialism and Democracy*, Schumpeter put forward "another theory of democracy" in which the social function of democracy is fulfilled incidentally by the competitive struggle for power between parties, just as the social function of markets is fulfilled incidentally by the competitive struggle for profits among firms (Schumpeter, 1950, Ch. 22). Anthony Downs did not cite this argument of Schumpeter directly, but he did state that: "Schumpeter's profound analysis of democracy forms the inspiration and foundation for our whole thesis" (Downs 1957, p. 27, n. 11). Downs was a student of Kenneth Arrow, and it appeared that with his dissertation he wished to develop Schumpeter's insight and demonstrate how political competition between parties could produce a welfare maximum and thus avoid the dire implications of Arrow's impossibility theorem. Downs ultimately failed in this endeavor, but succeeded in introducing a mode of analysis of competition using spatial modeling that was to have a profound impact on the development of the field, particularly among practitioners trained in political science. Building again on insights from Schumpeter (1950, pp. 256–264), Downs also developed a model of the rational voter who, among other things, rationally chooses to remain ignorant of most of the issues in an election (Chaps. 11–14).

Another doctoral dissertation that was to have a profound impact on both the public choice field and political science in general was that of Mancur Olson, published in book form in 1965.[5] Just as Downs had shown that the logic of rational decision making led individuals to invest little time in collecting information to help them decide how to vote, "the logic of collective action" would prevent individuals from voluntarily devoting time and money to the provision of public

---

[5] Alt (1999) describes the impact of Olson's work on political science literature.

goods. Mancur Olson did not invent the "free-rider problem," but no one has put it to better use than he did in this and his subsequent contributions to the literature.

All the "early classics" discussed so far were written by economists. One contribution by a political scientist that certainly falls into this category is William Riker's *The Theory of Political Coalitions* (1962). In this book Riker developed the logic of coalition formation into a theory that could explain among other things why "grand coalitions" were short lived. Riker's book foreshadowed a large literature that would apply tools of the game theory to political analysis.

Deciding when the early classics end and the "late" ones begin is a somewhat subjective judgment. Perhaps from the vantage point of 2002, however, the definition of 'early' can be extended through the early 1970s to include three more sets of works. First of these in chronological order would be an article published by Gordon Tullock in 1967. This article might be dubbed a "hidden classic," since its seminal nature did not become apparent to the profession at large until its main idea was rediscovered and developed by Anne Krueger (1974) and Richard Posner (1975) some time later. It was Krueger who gave the idea the name of rent seeking. Until Tullock's 1967 article appeared, standard discussions of "the social costs of monopoly" measured these costs solely in terms of the "deadweight triangle" of lost consumers' surplus resulting from the monopolist's restriction of output. The rectangle of monopoly rents was treated as a pure transfer from consumers to the monopolist and as such devoid of any welfare significance. Tullock pointed out, however, that the right to supply the monopolized product or service was a valuable right, and that individuals could be expected to invest time and money to obtain or retain this right. These investments constitute a pure social waste as they only serve to determine the identity of the monopoly rent recipient. They have no positive impact on the allocation of resources.

The social costs of rent seeking are potentially very large. Numerous articles have appeared since the pioneering contributions of Tullock and Krueger. One branch has analyzed theoretically the conditions under which the total resources invested in rent seeking fall short of, equal, or exceed the size of the rents pursued. A second branch has sought answers to the same questions empirically.[6] One of the curiosities of this literature has been that it has by and large analyzed rent seeking as if it were exclusively a problem of the public sector, even though the logic of rent seeking applies with equal validity to the private sector.[7]

While Tullock's rent-seeking article has proved to be a hidden classic, Sen's (1970) article about the Paretian liberal might be dubbed as an "unassuming classic." Sen put forward another sort of paradox, in the spirit of the Arrow paradox, but neither the author nor any of the readers of this six page note is likely

---

[6] For recent surveys of this literature, see Magee (1997), Tollison (1997) and Mueller (2003, Ch. 15).

[7] The same might be said of the implications of Arrow's impossibility theorem. The theorem establishes that no method for aggregating preferences is consistent with the five "Arrow axioms." The theorem thus casts a shadow over both market and non-market methods for aggregating individual preferences, and yet most discussions of the theorem's import focus only on democratic procedures.

to have appreciated, at the time it appeared, the impact it was to have on the literature.[8] Where Arrow proved that it was impossible *not to have a dictator* and satisfy four other axioms, Sen proved that it was impossible *to allow someone to be a dictator* over even one simple choice—as for example whether he sleeps on his back or his stomach—and satisfy three other axioms.

The last early contribution that qualifies as a classic is William Niskanen's (1971) book on bureaucracy. Niskanen posited that bureaucrats seek to maximize the size of their budgets and then proceeded to derive the implications of this assumption. A, by now, huge literature has been built on the analytical foundation that he laid.[9]

The Second Generation

More Impossibilities

During the 1970s several papers appeared, which extended the dire implications of Arrow's impossibility theorem and the literature it spawned. Satterthwaite (1975) and Gibbard (1977) demonstrated the incompatibility of having a preference aggregation procedure that was both nondictatorial and *strategyproof*, where strategyproof meant that everyone's best strategy was to faithfully reveal their true preferences. These theorems illustrated the close relationship between Arrow's independence-of-irrelevant-alternatives axiom and the goal of having a preference aggregation procedure in which individuals did not have an incentive to behave strategically.

McKelvey (1976) and Schofield (1978) drew out a further implication of a procedure's failure to satisfy the transitivity axiom. When a procedure leads to voting cycles it is possible to move anywhere in the issue space. An agenda setter can take advantage of this feature of cycling to lead a committee to the agenda setter's most preferred outcome.

The Veil of Tears Rises

The theorems of McKelvey and Schofield might be regarded as the capstones—or should we say tombstones—for the literature initiated by Arrow. It paints a very negative picture of the capacity for democratic procedures to aggregate information on voter preferences in a normatively appealing matter. Collective decisions were likely to be arbitrary or dictatorial. Free riding and the strategic concealment of individual preferences undermined democracy's legitimacy. Rent seekers and bureaucrats contributed to the "waste of democracy." William Riker's (1982)

---

[8] See, for example, Sen (1996).

[9] For recent surveys of this literature, see Moe (1997), Wintrobe (1997) and Mueller (2003, Chs. 16 and 17).

attack against "populist democracy"—the idea that democratic procedures could aggregate individual preferences reasonably—accurately conveys the flavor of this literature. Even before Riker's book appeared, however, several developments in the public choice literature were taking place that painted a far more cheery picture of democracy's potential. The first of these concerned the potential for direct revelation of preferences.

Voting Rules

In his classic article deriving the conditions for the Pareto optimal allocation of private and public goods, Paul Samuelson (1954) matter-of-factly proclaimed that it would be impossible to get people to honestly reveal their preferences, because no person could be excluded from consuming a pure public good. So things stood for nearly 20 years, when Clarke (1971) and Groves (1973) showed that individuals could be induced to reveal their preferences for public good honestly by charging them a special "incentive tax" equal to the costs that their participation in the collective choice process imposed on the other voters. This class of procedures was first discovered in another context by William Vickrey (1961), and has come to be known in the public choice literature as "demand revelation" processes.

Mueller (1978, 1984) showed that the preference revelation problem could be solved using a three-step procedure in which each individual first makes a proposal—say a quantity of public good and a tax formula to pay for it; and then following a random determination of an order of veto voting removes (vetoes) one element from the set of all proposals.

Hylland and Zeckhauser (1970) added to the list of preference-revelation procedures by showing that individuals will allocate a stock of "vote points" across a set of issues to reveal the intensities of their preferences on these issues, if the quantities of public good provided are determined by adding the square roots of the points each individual assigns to an issue. During the decade of the 1970s, one new method appeared after another to solve the heretofore seemingly insoluble problem of inducing people to reveal their preferences for public good honestly.

Two-Party Competition

During the decade of the 1980s, several papers appeared that suggested that two-party representative governments were far better at aggregating individual preferences than had previously been demonstrated. One set of these articles simply replaced the assumption of the Downsian voter model that each individual votes with probability one for the candidate promising her a higher utility, with the assumption that the probability of an individual's voting for a candidate increases when the candidate promises her a higher utility. Substituting this "probabilistic voting" assumption for the standard Downsian deterministic voting assumption allowed Coughlin and Nitzan (1981a, b) and Ledyard (1984) to prove that the

competition for votes between two candidates led them to select an equilibrium pair of platforms that maximized some form of social welfare function. Schumpeter's assertion that the competition for votes between parties resulted in a form of "invisible hand theorem" for the public domain was, after forty years, finally proved.

In a multidimensional issue space, every platform choice by one party can be defeated by an appropriate choice of platform by the other, and the two candidates might cycle endlessly, under the Downsian assumption of deterministic voting. Such cycling could in theory take the candidates far away from the set of most preferred points of the electorate. A platform x, lying far from the set of most preferred points of the electorate would, however, be dominated by some other point y, lying between x and the set of most preferred points of the electorate, in the sense that y could defeat every platform that x could defeat, and y could also defeat x. By restricting one's attention to points in the issue space that are not dominated in this way, the set of attractive platforms for the two candidates shrinks considerably. The cycling problem does not disappear entirely, but it is reduced to a small area near the center of the set of most-preferred points for the population.[10]

These results clearly sound a more optimistic note about the potential for preference aggregation than many of the early classics and the works discussed in the earlier sections. The reader can see how dramatic the difference in perspectives is by comparing the books by Wittman (1995) and Breton (1996) to that of Riker (1982).

Political Business Cycles

Almost all Nobel prizes in economics have been awarded for contributions to economic theory. All of the early classics in public choice have been theoretical contributions, as have the subsequent contributions reviewed so far.[11] As the public choice field has matured, however, an increasing number of studies have appeared testing every and all of its theoretical propositions. Space precludes a full review of the many empirical contributions that have been made to the field. We have therefore selected only three areas, where a lot of empirical work has been done, beginning with the area of "political business cycles."

One of the most frequently quoted propositions of Anthony Downs (1957, p. 28) is that "parties formulate policies in order to win elections, rather than win elections in order to formulate policies." Among the policies of great concern to voters few stand higher than the state of the economy. If the quoted proposition of Downs is correct, then parties should compete for votes on the basis of their promised

---

[10] Gordon Tullock's (1967) claim that this was the case was rigorously proved Miller (1980, 1983) and McKelvey (1986) among others.

[11] Riker (1962) demonstrated the explanatory power of his theory of coalitions with historical examples, but the main contribution of the book was to propose a theory.

macroeconomic policies, and both parties in a two-party system should offer the same set of policies. Kramer (1971) was the first to test for a relationship between the state of the economy and votes for members of the House and the president. Nordhaus (1975) and MacRae (1977) were among the first to develop a Downsian model of the political business cycle in which both parties are predicted to follow the same strategy of reducing unemployment going into an election to induce short-sighted voters to vote for the incumbent party/candidates.

Numerous observers of politics in both the United States and the United Kingdom have questioned the prediction of the one-dimensional Downsian model that both parties adopt identical positions at the most-preferred outcome for the median voter. This prediction appears to be blatantly at odds with the evidence concerning macroeconomic policies, where right-of-center parties clearly seem to be more concerned about inflation, while left-of-center parties are more concerned about unemployment. Early contributions by Hibbs (1977, 1987) and Frey and Schneider (1978a, b) incorporated these "partisan effects" into a political model of macroeconomic policy and provided empirical support for them.

In some areas of public choice, data for testing a particular proposition are difficult to obtain and empirical work is accordingly sparse. Such is not the case with respect to hypotheses linking policy choices to macroeconomic outcomes. Data on variables like unemployment and inflation rates are readily available for every developed country, as are data on electoral outcomes. Each passing year produces more observations for retesting and refining previously proposed hypotheses. The empirical literature on political business cycles is, by now, vast. The main findings grossly condensed are that partisan differences across parties are significant and persistent, but that both the parties of the left and parties of the right do tend to become more "Downsian" as an election approaches and adapt their policies to sway the uncommitted, middle-of-the-road voters.[12]

Public Choice Goes Multinational

All the early classics discussed earlier were written by either American or British authors. It is thus not surprising that the literature on representative government, as for example in the political business cycle area, has almost always assumed the existence of a two-party system—even when testing the model using data from countries with multiparty systems. In the last couple of decades, however, considerably more attention has been devoted to analyzing properties peculiar to multiparty systems. This literature has been heavily populated by persons trained in public choice, and is one in which the lines between political science and public choice are particularly blurred.

---

[12] For recent surveys of this literature, see Paldam (1997), Drazen (2000) and Mueller (2003, Ch. 19).

A salient feature of multiparty systems is that no single party typically wins a majority of seats in the parliament, and thus no single party is able to *form the government*. Consequently, a coalition of parties must come together if the cabinet is to reflect the wishes of a majority of the parliament, or a minority government forms. Two important questions arise: (1) Which parties will build the coalition that forms the government, and (2) How long will it last?

Game theory provides the ideal analytical tool for answering the first question, and it has been used to make a variety of predictions of the coalition that will form after an election. Riker's (1962) prediction, that a *minimum winning coalition* forms, receives as much support as any theory, although it accounts for less than half of the governments formed in European countries since World War II.[13] In particular, it failed to predict the many minority governments that have existed.

A theory that can account for the existence of minority governments has been put forward by van Roozendaal (1990, 1992, 1993). His theory emphasizes the pivotal position of a party that includes the median member of the parliament (a *central party*), under the assumption that the parties can be arrayed along a single, ideological dimension. Under the assumption that each party favors proposals coming close to their position along the ideological dimension over proposals lying far away, a central party will be a member of every coalition that forms. A large central party is likely to be able to successfully lead a minority government by relying on votes from the left to pass some legislation and votes from the right for some other legislation.

When the issue space cannot be reasonably assumed to be one-dimensional, cycling is likely to arise, which in the context of cabinet formation implies unstable party coalitions. Here concepts of game theory such as the covered set and the heart have proved useful for identifying the likely members of the coalitions that eventually form the government.[14]

A long literature beginning with Taylor and Herman (1971) has measured the length of a government's life and related this length to various characteristics of the government. One of the regularities observed is that minority governments tend to be relatively short lived and governments formed by a single, majority party long lived.[15]

Persson and Tabellini (2000, 2003) have developed and tested hypotheses about the effects of electoral rules on political outcomes such as the size and composition of state budgets, rent seeking and corruption. Their work uncovers significant differences between two- and multi-party systems, and between presidential and parliamentary systems.

---

[13] See Laver and Schofield (1990). A minimum winning coalition is one which constitutes a majority of the seats in the parliament, but falls to a minority coalition through the defection of any member party.

[14] For discussions of these concepts and surveys of this literature, see Laver and Schofield (1990), Schofield (1997), and Mueller (2003, Ch. 13).

[15] For a recent survey of this literature see Müller and Kaare (2000).

Other recent work with a multinational focus includes the work of Alesina and Spolaore (2003) on the size of nations, and of Filippov, Ordeshook and Shvetsova (2004) on federalism.

Experimental Economics

Experimental economics can be rightfully thought of as a separate field of economics and not just a "topic" in public choice. Two of its pioneering scholars—Vernon Smith and Charles Plott—have also been major contributors to the public choice field, however, and an important stream of the experimental economics literature has dealt with public choice issues. It thus constitutes an important body of empirical evidence corroborating, or in some cases undermining, certain hypotheses in public choice.

The first experimental study of the new voting mechanisms was by Vernon Smith (1979). He ran experiments on the Groves and Ledyard (1977) iterative version of the demand revelation process, and on a somewhat simpler auction mechanism that he had developed. In most experiments the subject chose a public good quantity and set of contributions that was Pareto optimal. The experiments also served to demonstrate the feasibility of using the unanimity rule, as the participants had to vote unanimously for the final set of contributions and public good quantity for it to be implemented.

Hoffman and Spitzer (1982) devised an experiment with an externality to test the Coase theorem and found that in virtually every run of the experiment the subjects were able to reach a bargain that was Pareto optimal.

A third set of experiments might in some way be thought of as rejecting a prediction of an important theory, but it rejects the theory in favor of alternatives that support the behavioral premises underlying the public choice methodology. Frohlich et al. (1987) presented students with four possible redistribution rules—Rawls's (1971) rule of maximizing the floor, maximizing the average, maximizing the average subject to a floor constraint, and maximizing the average subject to a range constraint. The students were made familiar with the distributional impacts of the four rules and were given time to discuss the merits and demerits of each rule. In 44 experiments in which students were uncertain of their future positions in the income distribution, five students in each experiment reached unanimous agreement on which redistributive rule to use to determine their final incomes in every case. Not once did they choose Rawls's rule of maximizing the floor. The most popular rule, chosen 35 out of 44 times, was to maximize the average subject to a floor constraint. Similar experiments conducted in Canada, Poland and the United States all found (1) that individuals can unanimously agree on a redistributive rule, and (2) that this rule is almost never Rawls's maximin rule, but rather some more utilitarian rule like maximizing the mean subject to a floor (Frohlich and Oppenheimer, 1992). While these results may constitute bad news for Rawlsians, they

lend support to the assumptions that underlie economic and public choice modeling. They suggest further that individuals are not concerned merely with their own welfare, but are also motivated by considerations of fairness and justice, although apparently not in the extreme form posited by Rawls.

The last set of experiments is less comforting for students of public choice. At least since the publication of Mancur Olson's *Logic of Collective Action* in 1965, a basic tenet in the public choice literature is that individuals will *free ride* in situations where contributions to the provision of a public good are voluntary. Countless experiments have demonstrated that they do free ride, but to a far smaller degree than one might have expected. If 100 is the contribution to the public good that produces the optimum quantity of the good for the collective, and 1 is the contribution that is individually optimal, then the typical finding in an experiment testing for free rider behavior is that the *mean* contribution of the participants is around 50. Some people do free ride, but many make contributions that are far larger than is individually optimal. In aggregate the total contributions fall far short of what would be optimal for the group, but far above what pure free riding behavior would produce.[16]

Many additional types of experiments have been run that have important implications for both public choice and other branches of economics, and many more will be run in the future. Experimental economics seems destined to remain an important source of empirical evidence for testing various theories and propositions from the field.[17]

Constitutional Political Economy

Over the last 25 years or so, considerable work has appeared that falls under the heading, Constitutional Political Economy, and a journal with this title has appeared. Much of this research builds on the two-stage modeling of the political process in *The Calculus of Consent* and is theoretical and often normative.[18] More recent contributions have sought to test hypotheses about the effects of constitutional institutions on political outcomes. Persson and Tabellini (2003) falls into this category. Voigt (1997) surveys this literature.

---

[16] The pioneering contributions to this strand of the literature were by Marwell and Ames (1979, 1980).

[17] For recent surveys of this literature, see Ledyard (1995) and Hoffman (1997). Ostrom and Walker (1997) also survey large parts of the literature.

[18] I tried to use developments in public choice to discuss possible contents of constitutions that would improve on present constitutions in Mueller (1996). Cooter (2000) is a nice application of game theoretic reasoning to constitutional issues. Ferejohn, Rakove and Riley (2001) contains several interesting essays. Blankart and Mueller (2004) contains essays focusing on the new EU constitution, as originally drafted.

The Next Generation

The field of public choice is now some sixty years old and befitting its age has begun to resemble other mature fields in economics. Important theoretical break-throughs are fewer and farther between than during the field's first 25 years. Much current research consists of extending existing theories in different directions, and of filling in the remaining empty interstices in the body of theory. Much current research also consists of empirically testing the many theoretical propositions and claims that have been made until now. The future development of the field will most certainly parallel that of other mature fields in economics—with continually in-creasing use of sophisticated mathematics in theoretical modeling and continual use of more and more sophisticated econometrics applied to larger and larger data sets when estimating these models.

Two other trends are apparent. Although public choice is destined to remain just one of many fields in economics, it is possible—I would dare to say likely—that it eventually takes over the entire discipline of political science, takes over in the sense that all political scientists will eventually employ rational actor models when analyzing various questions in political science and all will test their hypotheses using the same sort of statistical procedures that economists employ. Political institutions are sufficiently different from market institutions to require important modifications in the assumptions one makes about the objectives of rational actors in politics and about the constraints under which they pursue these objectives. Nevertheless, the assumption that individuals rationally pursue specific objectives has proven to be so powerful when developing testable hypotheses about their behavior, that this methodology—the methodology of public choice—must even-tually triumph in some form throughout the field of political science.

The second discernable trend is the full internationalization of the discipline. In addition to the US-based Public Choice Society, there is now a Japanese as also a European Public Choice Society. In the spring of 2007, the American and European branches jointly held the first *World* Public Choice Society meeting in Amsterdam. Public Choice can now be said to be both an interdisciplinary and an international field of research.

References

Lesina, A. and Spolaore, E. (2003). *The Size of Nations*. Cambridge, MA: MIT Press.

Alt, J. E. (1999). "Obituary: Thoughts on Mancur Olson's contribution to political science." *Public Choice*, 98: 1–4.

Arrow, K. J. (1950). "A difficulty in the concept of social welfare." *Journal of Political Economy*, 58: 328–346.

Arrow, K. J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Black, D. (1948a). "On the rationale of group decision making." *Journal of Political Economy*, 56: 23–34; reprinted in K.J. Arrow and T. Scitovsky (eds.) (1969) 133–146.

Black, D. (1948b). "The decisions of a committee using a special majority." *Econometrica*, 16: 245–261.

Black, D. (1958). *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Blankart, C. B. and D. C. Mueller. (eds.) (2004). *A constitution for the European Union*. Cambridge, MA: The MIT Press (CESifo Seminar Series).

Bowen, H. R. (1943). "The interpretation of voting in the allocation of economic resources." *Quarterly Journal of Economics*, 58: 27–48.

Breton, A. (1996). *Competitive Governments*. Cambridge: Cambridge University Press.

Buchanan, J. M. (1949). "The pure theory of government finance: A suggested approach." *Journal of Political Economy*, 57: 496–506.

Buchanan, J. M. and Tullock, G. (1962). *The Calculus of Consent. Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Clarke, E. H. (1971). "Multipart pricing of public goods." *Public Choice* 11: 17–33.

Cooter, R. D. (2000). *The strategic constitution*. Princeton: Princeton University Press.

Coughlin, P. and Nitzan, S. (1981a). "Electoral outcomes with probabilistic voting and nash social welfare maxima." *Journal of Public Economics*, 15: 113–122.

Coughlin, P. and Nitzan, S. (1981b). "Directional and local electoral equilibria with probabilistic voting." *Journal of Economic Theory*, 24: 226–239.

de Borda, J. C. (1781). *Memoire sur les Elections au Scrutin*. Paris: Histoire de l'Academie Royale des Sciences.

de Condorcet, M. (1785). Essai sur l'Application de L'Analyse à la Probabilité des Décisions Rendues à la Pluraliste des Voix. Paris.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Drazen, A. (2000). *Political Economy in Macroeconomics*. Princeton: Princeton University Press.

Ferejohn, J., Rakove, J. N., and Riley, J. (eds.). (2001). *Constitutional Culture and Democratic Rule*. Cambridge: Cambridge University Press.

Filippov, M., Ordeshook, P. C., and Shvetsova, O. (2004). *Designing Federalism*. Cambridge: Cambridge University Press.

Frey, B. S. and Schneider, F. (1978a). "An empirical study of politico- economic interaction in the U.S." *Review of Economics and Statistics*, 60: 174–183.

Frey, B. S. and Schneider, F. (1978b). "A politico-economic model of the United Kingdom." *Economic Journal*, 88: 243–253.

Frohlich, N. and Oppenheimer, J. A. (1992). *Choosing Justice: An Experimental Approach to Ethical Theory*. Berkeley: University of California Press.

Frohlich, N., Oppenheimer, J. A., and Eavey, C. L. (1987). "Laboratory results on Rawls's distributive justice." *British Journal of Political Science*, 17: 1–21.

Gibbard, A. (1977). "Manipulation of schemes that combine voting with chance." *Econometrica*, 45: 665–668.

Groves, T. (1973). "Incentives in teams." *Econometrica*, 41: 617–631.

Groves, T. and Ledyard, J. J. (1977). "Optimal allocation of public goods: a solution to the 'free rider' problem." *Econometrica*, 45: 783–809.

Hibbs, D. A., Jr. (1977). "Political parties and macroeconomic policy." *American Political Science Review*, 71: 1467–1487.

Hibbs, D. A., Jr. (1987). *The Political Economy of Industrial Democracies*. Cambridge: Harvard University Press.

Hoffman, E. (1997). "*Public choice experiments*," in Mueller, D.C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 415–426.

Hoffman, E. and Spitzer, M. L. (1982). "The coase theorem: Some experimental tests." *Journal of Law and Economics*, 25: 73–98.

Hotelling, H. (1929). "Stability in competition." *Economic Journal*, 39: 41–57.

Hylland, A. and Zeckhauser, R. (1970). "A mechanism for selecting public goods when preferences must be elicited." KSG Discussion Paper 70D, Harvard University.

Kramer, G. H. (1971). "Short run fluctuations in U.S. voting behavior, 1896–1964." *American Political Science Review*, 65: 131–143.

Krueger, A. O. (1974). "The political economy of the rent-seeking society." *American Economic Economic Review*, 64(3): 291–303; reprinted in Buchanan, J. M., Tollison, R. D., and Tullock, G. (eds.) (1980) 51–70.

Laver, M. and Schofield, N. (1990). *Multiparty Government*. Oxford: Oxford University Press.

Ledyard, J. O. (1984). "The pure theory of large two-candidate elections." *Public Choice*, 44(1): 7–41.

Ledyard, J. O. (1995). "Public goods: a survey of experimental research," in Kagel, J. H. and Roth, A. E. (eds.) *The Handbook of Experimental Economics*. Princeton: Princeton University Press, pp. 111–251.

MacRae, C. D. (1977). "A political model of the business cycle." *Journal of Political Economy*, 85: 239–263.

Magee, S. P. (1997). "Endogenous protection: the empirical evidence," in Mueller, D. C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 526–561.

Marwell, G. and Ames, R. E. (1979). "Experiments on the provision of public goods I: resources, interest, group size, and the free rider problem." *American Journal of Sociology*, 84: 1335–1360.

Marwell, G. and Ames, R. E. (1980). "Experiments on the provision of public goods II: provision points, stakes, experience and the free rider problem." *American Journal of Sociology*, 85: 926–937.

McKelvey, R. D. (1976). "Intransitivities in multidimensional voting models and some implications for agenda control." *Journal of Economic Theory*, 12: 472–482.

McKelvey, R. D. (1986). "Covering, dominance, and institution-free properties of social choice." *American Journal of Political Science*, 30: 283–314.

Mill, J. S. (1861). *Considerations on Representative Government*. New York: Bobbs-Merrill, 1958.

Miller, N. R. (1980). "A new solution set for tournaments and majority voting: further graph-theoretical approaches to the theory of voting." *American Journal of Political Science*, 24: 68–96.

Miller, N. R. (1983). "The covering relation in tournaments: Two corrections." *American Journal of Political Science*, 27: 382–385.

Moe, T. M. (1997). "The positive theory of public bureaucracy," in Mueller, D.C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 455–480.

Mueller, D. C. (1978). "Voting by veto." *Journal of Public Economics*, 10: 57–75.

Mueller, D. C. (1984). "Voting by veto and majority rule." in Horst, H. (ed.) *Public Finance and the Quest for Efficiency*. Detroit: Wayne State University Press, pp. 69–85.

Mueller, D. C. (1996). *Constitutional Democracy*. Oxford: Oxford University Press.

Mueller, D. C. (2003). *Public Choice III*. Cambridge: Cambridge University Press.

Müller, W. C. and Kaare, S. (2000). "Coalition governance in Western Europe," in Müller, W.C. and Strøm, K. (eds.) *Coalition Governments in Western Europe*. Oxford: Oxford University Press, pp. 559–592.

Niskanen, W. A., Jr. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton.

Nordhaus, W. D. (1975). "The political business cycle." *Review of Economic Studies*, 42: 169–190.

Olson, M., Jr. (1965). *The Logic of Collective Action*. Cambridge: Harvard University Press.

Ostrom, E. and Walker, J. (1997). "Neither markets nor states: Linking transformation processes in collective action areas," in Mueller, D. C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 35–72.

Paldam, M. (1997). "Political business cycles," in Mueller, D. C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 342–370.

Persson, T. and Tabellini, G. (2000). *Political Economics*. Cambridge, MA: MIT Press.

Persson, T and Tabellini, G. (2003). *The Economic Effect of Constitions*. Cambridge, MA: MIT Press.

Posner, R. A. (1975). "The social costs of monopoly and regulation." *Journal of Political Economy*, 83: 807–827; reprinted in Buchanan, J. M., Tollison, R. D., and Tullock, G. (eds.) (1980) 71–94.

Rawls, J. A. (1971). *A Theory of Justice*. Cambridge, MA: Belknap Press.

Riker, W. H. (1962). *The Theory of Political Coalitions*. New Haven: Yale University Press.

Riker, W. H. (1982). *Liberalism Against Populism*. San Francisco: Freeman.

Samuelson, P. A. (1954). "The pure theory of public expenditure." *Review of Economics and Statistics*, 36: 386–389.

Saari, D. G. (1994). *Geometry of Voting*. Berlin Heidelberg New York: Springer.

Satterthwaite, M. A. (1975). "Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions." *Journal of Economic Theory*, 10: 187–217.

Schofield, N. (1978). "Instability of simple dynamic games." *Review of Economic Studies*, 45: 575–594.

Schofield, N. (1997). "Multiparty electoral politics," in Mueller, D. C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 271–295.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*, 3rd edn, (1950) New York: Harper & Row.

Sen, A. (1970). "The impossibility of a paretian liberal." *Journal of Political Economy*, 78: 152–157.

Sen, A. (1996). "Rights: Formulation and consequences." *Analyse & Kritik*, 18: 153–170.

Smith, V. L. (1979). "An experimental comparison of three public good decision mechanisms." *Scandinavian Journal of Economics*, 81(2): 198–215.

Taylor, M. J. and Herman, V. M. (1971). "Party systems and government stability." *American Political Science Review*, 65: 28–37.

Tollison, R. D. (1997). "Rent seeking," in Mueller, D.C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 506–525.

Tullock, G. (1967). "The welfare costs of tariffs, monopolies and theft." *Western Economic Journal*, 5: 224–232; reprinted in Buchanan, Tollison and Tullock (eds.) (1980) 39–50.

van Roozendaal, P. (1990). "Centre parties and coalition formations: A game theoretic approach." *European Journal of Political Research*, 18: 325–348.

van Roozendaal, P. (1992). "The effect of dominant and central parties on cabinet composition and durability." *Legal Studies Quarterly*, 17: 5–36.

van Roozendaal, P. (1993). "Cabinets in the Netherlands (1918–1990): The importance of 'dominant' and 'central' parties." *European Journal of Political Research*, 23: 35–54.

Vickrey, W. (1961). "Counterspeculation, auctions, and competitive sealed tenders." *Journal of Finance*, 16: 8–37.

Voigt, S. (1997). "Positive constitutional economics – A survey." *Public Choice*, 90: 11–53.

Young, H. P. (1997). "Group choice and individual judgements," in Mueller, D.C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 201–225.

Wicksell, K. (1896). "Ein neues Prinzip der gerechten Besteuerung." *Finanztheoretische Untersuchungen*, Jena; translated as "A New Principle of Just Taxation," 1958; reprinted in (1967) Richard, A. M., and Peacock, A. T. (eds.) *Classics in the Theory of Public Finance*. London: Macmillan, pp. 72–118.

Wintrobe, R. (1997). "Modern bureaucratic theory," in Mueller, D.C. (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, pp. 429–454.

Wittman, D. (1995). *The Myth of Democratic Failure: Why Political Institutions are Efficient*. Chicago: University of Chicago Press.

Chapter 3
Public Choice Trailblazers versus the Tyranny of the Intellectual Establishment[1]

Charles K. Rowley[2]

Introduction

Public choice—or the economics of politics—is a relatively new science located at the interface between economics and politics. It was founded during the late 1940s in a sequence of papers by Duncan Black, primarily focused on voting within committees and elections. Black died in 1991 without achieving full recognition as a Founding Father of the discipline.

Public choice was extended into the arena of representative government first by Anthony Downs (1957), who outlined a theory that emphasized the centripetal tendencies of two parties competing for electoral votes, and then by Buchanan and Tullock (1962), who provided the long-term foundations for the subject within the framework often referred to as politics without romance. The practitioners of this latter brand of public choice seek to understand and to predict the behavior of political processes by utilizing the analytical techniques of private market economics—most notably methodological individualist and the rational choice postulates—albeit identifying the crucial differences between the behavior of the private market and non-market decision-making processes.

In this paper, I suggest that leading United States neoclassical economists, both from saltwater and freshwater universities, have systematically ignored, downplayed, or distorted the scholarship of the public choice trailblazers, thereby slowing down the impact of public choice ideas on the intellectual mainstream. Because of the radical nature of public choice thinking, its trailblazers ran head on into the tyranny of the intellectual establishment, as reflected by four highly prestigious Nobel Prize winning economists, namely Paul Samuelson, Kenneth Arrow, George Stigler and Gary Becker. I explore the conflict in the context of four episodes in the

---

development of Public Choice ideas: Cycling under the rule of simple majority voting, the burden of the national debt, interest-group theory, and rent-seeking and the efficiency of government legislation.

The Saltwater 'Social Welfare Function' Economists

Paul Samuelson and Kenneth Arrow are pre-eminent figures of modern neoclassical economics. They and their innumerable disciples would offer the first line of defense against public choice ideas, relying on a complete separation between, on one hand, private market analysis with behavior evaluated against a carefully specified social welfare function, and, on the other hand, on a public-interested state as responding impartially to the considered advice of philosopher-king economists. Inevitably, the approach established an uneven playing field that unfairly favored government over private markets across a significant range of policy margins.

Paul A. Samuelson

Paul Samuelson received his bachelor's degree from the University of Chicago (where he studied under Frank Knight and Jacob Viner) in 1935, his master's degree from Harvard University in 1936 and his doctorate from Harvard University (where he studied under Joseph Schumpeter and Wassily Leonief) in 1941. He joined MIT as an assistant professor in 1940, was promoted to associate professor in 1944 and to full professor in 1947. In 1966, he became Institute Professor, a rank of high honor reserved for only a few members of the MIT faculty.

In 1947, Samuelson earned international renown for his magnum opus, *Foundations of Economic Analysis*, based on his doctoral dissertation. In that book, Samuelson confronted the perceived contradictions, overlaps and fallacies in classical economics and attempted to unify economics through the language of mathematics. The American Economic Association recognized this book in 1947 by awarding Samuelson the John Bates Clark Medal, the first of what would be many illustrious awards.

In 1948, Samuelson confirmed his international reputation with the publication of his famous textbook, *Economics*, which, under single-authorship until 1986, and thereafter under co-authorship, has survived 60 years in multiple editions, translated into 41 different languages, and selling well in excess of four million copies. This book introduced generations of students to Keynesian economics and pursued a consistently social-democratic approach to economic policy throughout the second half of the twentieth century.

By 1970, Samuelson was widely regarded as the pre-eminent economist of the West, with several hundreds of articles to his name spanning almost all fields of neoclassical economic theory. In that year, only the second year of the Prize, he became the first American economist to be awarded the Bank of Sweden Prize in

Economic Sciences in Memory of Alfred Nobel. The following excerpt from the Nobel citation summarizes the reasoning behind this award:

> More than any other contemporary economist, Samuelson has helped to raise the general analytical and methodological level in economic science. He has simply rewritten considerable parts of economic theory. He has also shown the fundamental unity of both the problems and analytical techniques in economics, partly by a systematic application of the methodology of maximization for a broad set of problems. This means that Samuelson's contributions range over a large number of different fields.

Samuelson was never wary of intermingling policy advice with scientific contributions. For many years, he wrote a column on economics for *Newsweek Magazine*. He was an advisor to Presidents Kennedy, Johnson and Carter while adding to President Nixon's list of enemies for his consistently harsh criticism of the Nixon administration's economic policies. In 1996, he was awarded the National Medal of Science by President Clinton. The award recognized "his fundamental contributions to economic science, education and policy for nearly 60 years, establishing both the agenda of modern economics and scientific standards for economic analysis of a wide range of problems including social security and the public debt, welfare and international trade."

So the public choice trailblazers ran into a very serious road block when they challenged Samuelson's new welfare economics, and set in motion a research program that would eventually level the playing field in all political economic debates concerning the appropriate limits of the state, for Samuelson's interest in welfare economics dates back certainly to his *Foundations of Economic Analysis* (Samuelson 1947).

The chapter on welfare economics in that book provides a brief but fairly complete survey of the entire field. It also exposits and develops what became commonly known as the Bergson-Samuelson social welfare function, a device that shows how to rank different feasible economic policy alternatives as better than, worse than, or indifferent to each other from the perspective of welfare economics.

In subsequent papers, Samuelson emerged as one of the most aggressive purveyors of the mid-twentieth century economics of market failure. As such, he became an inevitable target for, and a forthright opponent of, the public choice trailblazers.

Kenneth J. Arrow

Kenneth Arrow was born in August 1921 in New York City. He attended the City College in New York, graduating in 1940 with a bachelor's degree in social science with a major in mathematics. He continued his studies at Columbia University, first in mathematics and then in economics, though his primary interest at that time was mathematical statistics. He received his master's degree in 1941. He served as a weather officer in the United States Army Air Corps from 1942 to 1946, utilizing mathematical models to analyze the optimal use of winds for flight planning.

The years 1946 to 1949 were spent studying for a doctorate in economics at Columbia University under the mentorship of Harold Hotelling, while simultaneously engaging in research in mathematical economics at the Cowles Commission for Research in Economics at the University of Chicago. Arrow was appointed Assistant Professor of Economics in the University of Chicago for the year 1948/49. He also consulted at the RAND Corporation, where he first formulated the theory of social choice that proved his famous 'impossibility theorem' of group decision-making.

In 1949, Arrow was appointed Acting Assistant Professor of Economics and Statistics at Stanford University, two years before completing his doctorate at Columbia University in 1951. He remained at Stanford University until 1968, eventually becoming Professor of Economics, Statistics and Operations Research. In 1968, he left Stanford University to become Professor of Economics at Harvard University, where he remained until 1979, becoming James Bryant Conant University sity Professor in 1974. In 1979, he returned to Stanford to take up the position of Joan Kenney Professor of Economics and Professor of Operations Research. Arrow retired in 1991 at 70 years of age.

During the course of a stellar career, the scientific community has recognized Arrow's work with some of its most prestigious awards. In 1957, he received the John Bates Clark Medal from the American Economic Association. In 1972, he shared the Royal Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel with Sir John Hicks for the development of theories underlying the assessment of business risk and government economic and welfare policies. In 2004, he received the National Medal of Science for his ground-breaking contributions to economic theory and for his broad understanding of the social science arena in which theories are confronted and practical lessons worked out.

Although Kenneth Arrow, in one of his earliest papers (Arrow 1950) demonstrated the impossibility of formulating any social welfare function capable of satisfying simultaneously a number of so-called 'reasonable assumptions', nevertheless, he was a prolific user of social welfare functions (notably utilitarian in nature) throughout the remainder of his career.

With the possible exception of his 1954 paper co-authored with Gerard Debreu (Arrow and Debreu 1954) all of Arrow's policy papers have focused on the prevalence of market failure, and all his policy recommendations are grounded on the assumption of a benevolent welfare-maximizing government. Inevitably, therefore, he would prove to be a formidable establishment figure obstructing the real public choice trailblazers, even if some second-generation public choice scholars choose to put the fox in the henhouse by claiming him as one of their own.

The Freshwater 'Democracy is Efficient' Economists

George Stigler and Gary Becker are widely regarded as the joint founders of the Chicago political economy research program (CPE). The CPE research program was a relatively late starter, launched by George Stigler's 1971 article on economic

regulation. Stigler retained the intellectual leadership of the program until his death in 1991, when the mantle passed to Gary Becker. CPE is overtly positivist in its methodology, asserting for the most part that economists can observe, explain and predict, but cannot influence the course of political economy, and that attempts to change the world, by and large, are futile and wasteful of scarce resources (Rowley 1993, pp. 38–41).

CPE is a body of literature that analyzes political markets from the perspective of price theory and tight equilibrium theory (Mitchell 1989, Tollison 1989, Reder 1982). It views government primarily as a mechanism utilized by rational self-seeking individuals to redistribute wealth within a society. *Homo politicus* is modeled as a solipsist who exclusively pursues expected wealth maximization. Thus, CPE denies all credibility to the public interest theory of government. Yet, its analytics indicate that political markets efficiently allocate scarce resources among competing ends (Rowley 1993).

The thrust of their theory is towards instantaneous and durable equilibrium, with political markets always clearing. In equilibrium no individual can raise his expected wealth without reducing the expected wealth of at least one other individual. Political agents (brokers) clear political markets without invading them as principals. In this sense, political markets are not driven by broker preferences, aside from the desire for personal wealth. Political ideology plays no role, at least among the brokers of the political process.

The assumptions of CPE do not produce equilibrium based on perfect information or perfect mobility across political market space. High transaction costs will impede exchanges within the political process much as they impede private market exchanges. Uneconomic information will not be utilized by individuals within the political process. Random shocks may well create temporary disequilibrium, as actors respond to false signals. Nevertheless, CPE places a strong presumption in favor of stochastic analogues of deterministic general equilibrium in its analytic modeling. Such modeling differs sharply from that developed by the public choice trailblazers. Leading exponents of CPE, therefore, have attempted to sideline key contributions by public choice pioneers, seeking to uphold the rational expectations perspective of the late twentieth-century Chicago intellectual establishment.

George J. Stigler

George Stigler was born in 1911 in Renton, Washington. He received a Bachelor of Arts degree in Economics from the University of Washington in 1931 and an MBA from Northwestern University in 1932. Stigler returned to the University of Washington for one more year of graduate study, before receiving a tuition scholarship that enabled him to enroll at the University of Chicago in 1933. There, working under the guidance of scholars such as Frank Knight, Jacob Viner, John Nef, Henry Simons, and others, Stigler bathed in an intense intellectual atmosphere

that would captivate him for the remainder of his life. Notable among his Chicago peers were Milton Friedman, Paul Samuelson, Kenneth Boulding and Allen Wallis, each of whom would become a stellar academic economist following the Second World War.

In 1936, while still working on his doctoral dissertation under the chairmanship of Frank Knight, Stigler accepted an appointment as Assistant Professor of Economics at Iowa State College. He completed his doctoral dissertation in Spring 1938 and immediately moved to the University of Minnesota. In 1942, following the outbreak of war, Stigler took a sustained leave of absence, first to the National Bureau of Economic Research, where he became acquainted with Arthur Burns, and then to the Statistical Research Group at Columbia University, where he engaged in statistical analysis related to military problems. The Director was Allen Wallis, and other senior figures included Milton Friedman, Harold Hotelling, L.J. Savage and Abraham Wald.

When the war ended in 1945, Stigler returned to the University of Minnesota, where he was joined briefly by Milton Friedman. Within a year, however, Stigler moved to Brown University to take up a professorship, having been turned down for a position at the University of Chicago in favor of his erstwhile colleague, Milton Friedman. After a year at Brown, Stigler moved to Columbia University in New York, where he remained until 1958, despite several attempts by Theodore Schultz to lure him to Chicago.

In 1958, Allen Wallis, Dean of the University of Chicago Business School, persuaded Stigler to accept the Charles R. Walgreen Professorship of American Institutions. Stigler would remain at Chicago for the remainder of his life, editing the *Journal of Political Economy*, directing the Industrial Organization Workshop and founding, in 1977, the Center for the Study of the Economy and the State, serving as its director until his death. From 1971 until his death Stigler also held a fellowship at the Hoover Institution at Stanford where he spent part of almost every year.

Stigler was elected to the National Academy of Sciences in 1975. He received the Royal Bank of Sweden Prize in Economic Sciences in the Memory of Alfred Nobel in 1982 'for his seminal studies in industrial structures, functioning of markets and causes and effects of public regulation'. In 1987, he received the National Medal of Science.

Until late in his career, Stigler was a consistently pro-market economist who made stellar contributions to the history of economic thought (Stigler 1947), price theory, the economics of information, economic regulation (Stigler 1971) and the organization of industry. For these contributions, Stigler justly earned his Nobel Prize.

Unfortunately, Stigler's late contributions to Chicago Political Economy (Stigler 1988 a, b, 1992) proved to be less insightful, as the supreme market analyst failed to recognize significant institutional differences between private markets and political processes. This failure would lead Stigler to challenge head-on most of the major findings of the public choice trailblazers, with his final (posthumous) contribution on 'law or economics?' (1992) providing an unfortunate monument to his fundamental lack of understanding of the nature of the state.

Gary S. Becker

Gary Becker was born in December 1930 in Pottsville, Pennsylvania, but grew up in Brooklyn, New York. At school, he specialized in mathematics and the natural sciences, sharing team prizes in math and science competitions against specialized high schools. He won a math scholarship to Princeton University at age seventeen, and enrolled in an economics elective, where he studied microeconomics by reading the works of Paul Samuelson, George Stigler and John Hicks before taking a range of more advanced courses.

Graduating in three years with a bachelor's degree, Becker published (with William Baumol) his junior-year thesis on classical monetary theory and his senior-year thesis on multi-country trade, both in 1952 in the *American Economic Review*. With two well-received publications under his belt, he would have his pick among the top economics graduate schools, including Princeton. In the most important decision of his professional career, Becker chose the University of Chicago, despite the fact that it was not as highly ranked in economics as Princeton, Harvard, and Yale.

At Chicago, Becker would learn price theory from Milton Friedman, human capital theory from Ted Schulz, labor economics from Gregg Lewis (who would chair his doctoral dissertation committee), and statistics and probability theory from L.J. Savage. He obtained a master's degree in 1953 and a doctorate in 1955. His doctoral dissertation on the economics of discrimination was published in 1957 by the University of Chicago Press. The publisher received a subsidy from the Department of Economics (no market brief here for every tub standing on its own bottom). The book, though ultimately very successful, was initially viewed askance by an economics profession largely skeptical of its economic relevance.

Becker, without acceptable outside offers, stayed on at Chicago as Assistant Professor of Economics over the period 1954–1957 before accepting an Assistant Professorship at Columbia University. In 1958, he was promoted to Associate Professor and in 1960 to Professor of Economics. In 1968, he was appointed to the Arthur Lehman Professorship of Economics at Columbia University. Throughout his stay at Columbia University, Becker also researched at the National Bureau of Economic Research.

In 1969, upset by the weak-kneed responses both by the Columbia University administrators and his economist colleagues to student riots, Becker returned to Chicago as Ford Foundation Visiting Professor of Economics. In 1970, Chicago appointed him University Professor of Economics. In 1983, in recognition of his work in sociology, Becker became University Professor of Economics and Sociology at Chicago, a position that he still holds. Since 1990, he has also held the position of Senior Research Fellow at the California-based Hoover Institution on War, Revolution and Peace.

Becker has produced a prodigious volume of high quality scholarship, deploying the rational choice approach of neoclassical economics effectively across diverse areas of human behavior, many of which were previously assumed to be beyond the reach of

economic analysis (Rowley 1999). His work has brought him great recognition within the economics profession, starting as early as in 1967, when he was awarded the John Bates Clark Medal by the American Economics Association.

In 1992, Becker was awarded the Royal Bank of Sweden Prize in Economic Sciences in honor of Alfred Nobel for his pioneering work on the behavior of the family, including distribution of work and allocation of time, crime and punishment, and discrimination. In the judgment of the Swedish academy, Becker's contribution consisted "primarily in having extended the domain of economic theory to aspects of human behavior which had previously been dealt with—if at all—by other social science disciplines such as sociology, demography and criminology."

In December 2000, Becker received the National Medal of Science. He was cited for his pioneering the economic analysis of racial discrimination, inventing the economics of human resources, producing the major modern innovations in economic demography, and leading recent developments in how social forces shape individual economic behavior. In November 2007, Becker was awarded the Presidential Medal of Freedom.

Once Milton Friedman retired from Chicago in 1976, Becker increasingly came under the intellectual influence of George Stigler. Following the lead of his new mentor, Becker began to model the political process as conveying many of the efficiency characteristics of private markets. Becker's ideas about the working of pressure groups under conditions of democracy, as we shall see, place him in the front line among the rational expectations establishment at Chicago and in direct conflict with the writings of the early public choice trailblazers in this important area of public choice analysis.

Cycling under the Rule of Simple Majority Voting

Duncan Black (1908–91) was a Scot, born in Motherwell, an industrial town located some 12 miles from Glasgow. He obtained an M.A. degree in mathematics and physics at the University of Glasgow in 1929, and an M.A. in political economy and political philosophy, also at Glasgow, in 1932. He was appointed in 1932 to an assistantship in economics at the University of Dundee, joining Ronald Coase in a two-year exposure to rigorous analytical economics (Coase 1981). For a short biography of Duncan Black, see Chapter 4 of this book.

From the outset of his scientific career, Black's vision had been grand yet simple, namely to develop a pure science of politics as a generalized theory of committees (Grofman 1981). To this end, Black modeled political phenomena 'in terms of the preferences of a given set of individuals in relation to a given set of motions, the same motions appearing on the preference schedule of each individual' (Black 1948a). This body of scholarship, on what came to be called the theory of committees and elections, constitutes one of the major pillars of public choice.

Some 230 years ago, the Marquis de Condorcet demonstrated that majority rule need not yield a stable outcome when there are more than two alternatives to be

considered. Although periodically rediscovered or reinvented by succeeding generations of scholars, "the paradox of cyclical majorities" for all practical purposes was unknown to twentieth century scholars of democratic theory until Duncan Black called it to their attention in two innovative papers (Black 1948a and b). Although Black was not the first to discover the phenomenon, his work is the foundation of all subsequent research on the problem. Yet, ironically, Kenneth Arrow, not Duncan Black, is incorrectly credited with this rediscovery.

In his seminal paper (Black 1948a), Duncan Black first formulated the median voter theorem—a contribution which is widely acknowledged as providing the intellectual foundation for the spatial theory of voting under majority vote rules—which earned him recognition as a founding father of public choice. The median voter theorem provides an optimistic insight that the majority-vote rules, under favorable conditions where all voters' preferences are single-peaked over policy space, and provides unique and stable election outcomes, much as perfect competition provides such outcomes in private markets.

Less well-known, however, is the fact that, in the same article, Black 'discovered' that once the assumption of single-peaked voter preferences is relaxed, no motion "need exist which is able to get at least a simple majority over every other" (Black 1948a, p. 32). Black illustrates this 'paradox of voting' by reference to the following set of preferences by three voters, A, B and C over three alternative motions, a_1, a_2 and a_3 (Fig. 3.1).

Black notes that motion a_1 wins a pair-wise majority against a_2; that a_2 wins a pair-wise majority against a_3; and that a_3 wins a pair-wise majority against a_1. In such circumstances, none of the three motions is able to get a simple majority over the other two. He then demonstrates that, "the greater the number of motions put forward in a committee of any given size, the greater will be the percentage of the total number of possible cases in which there exists no motion which is able to get a simple majority over each of the others" (Black 1948a, p. 33).

Black further notes that, in such a state of affairs, the particular motion that is adopted by the committee, using the majority vote rule, will depend on chance—the chance of particular motions coming earlier or later into the voting process. He does not mention, in this paper, the opportunity thus provided for agenda-manipulation, but his analysis takes the astute reader precisely to this extension.

Duncan Black's 1948a article was published by the *Journal of Political Economy*, one of the world's leading economic journals. In 1950, just two years later, Arrow published his famous article on the concept of social welfare in the same journal (Arrow 1950). Arrow's article makes no reference to Black's contribution

			Voters		
			A	B	C
Preference Order		1	a_1	a_2	a_3
		2	a_2	a_3	a_1
		3	a_3	a_1	a_2

Fig. 3.1 Duncan Black's paradox of voting

despite the fact that Arrow utilizes this paradox as the fulcrum of his analysis of democracy as an unacceptable mechanism for making collective choices.

How come, one may well ask, Arrow was able to evade the attention of the very same editors at the JPE who, just two years earlier, had approved the publication of Black's article? The answer is that he claimed in one short sentence that the paradox of voting was "well-known" (Arrow 1950, p. 328), before proceeding to illustrate its nature with an example almost identical to that outlined by Black. On the basis of this illustration Arrow proceeded to make his now-famous claim that "the method just outlined for passing from individual to collective tastes fails to satisfy the condition of rationality as we understand it ordinarily" (Arrow 1950, p. 329). In his paper, Arrow does not cite any source for the paradox of voting.

In his follow-up book, *Social Choice and Individual Values* (Arrow 1951), which comprised his doctoral dissertation at Columbia University, Arrow recognized the importance of citing a source for the paradox of voting: "the 'paradox of voting' seems to have been first pointed out by E.J. Nanson (*Transactions and Proceedings of the Royal Society of Victoria*, Vol 19, 1882, pp. 197–240)" (Arrow, 1951/1963, p. 3). Arrow acknowledges, however, that he had not read that reference. In 1963, now somewhat more defensively, Arrow expands on this cryptic reference:

> When I first studied the problem and developed the contradictions in the majority rule system, I was sure that this was no original discovery, although I had no explicit reference, and sought to express this knowledge by referring to 'the well-known paradox of voting'. When the basic ideas of the book were first read as a paper to the Econometric Society in December 1948, Professor C.P. Wright of the University of New Brunswick called my attention to the work of E.J. Nanson. Nanson, in discussing the proposal of his for a method of election, refers without great emphasis to the possibility of intransitivity arising from majority choice . . . for which he gives no previous reference (Arrow 1963, p. 93).

Bernard Grofman (Grofman 1981), one of the world's leading scholars of public choice, is skeptical (to say the least) about Arrow's 1950 and 1951 claim that the paradox of cyclical majorities was well known prior to Black's 1948 publication. Let me quote Grofman in full on this issue:

> The paradox was rediscovered by Huntington (1938), but this work had no discernible impact on subsequent research and also did not connect the problem to issues in democratic theory. Certainly, in political science, the then standard texts on democratic theory and political philosophy make no mention of the paradox. We are aware of only two 20th century, pre-WWII references to it other than Huntington (1938): the 1907 reprint of Nanson (1882) and Hoag and Hallett (1926). Furthermore, Riker (1965, p. 43) has asserted that as far as he knew 'there was no handbook of parliamentary law that mentions the cyclical majority.' The present author is a professional parliamentarian familiar with well over a dozen parliamentary manuals and has no evidence to contradict Riker's assertion (Grofman 1981, p. 14).

Black's paper was published by the JPE in February 1948, and must have circulated among economics scholars at Chicago well before that date. Between 1946 and 1948, Arrow worked for the Cowles Commission at the University of Chicago. During 1948/49, he was a junior member of the Chicago economics faculty. It is inconceivable that he would not have read Black's 1948 article at some stage during

that period, especially given his direct interest in the subject-matter of Black's article.

In 1949, Duncan Black (this time with R.A. Newing as co-author) extended his work on the paradox of voting by focusing attention on a situation in which the preferences of voters over any set of alternatives depend on what decisions have previously been made. The order in which decisions are made now plays a crucial role in the ultimate outcome. The ensuing paper was submitted to *Econometrica* in November 1949.

Econometrica did not provide the authors with a publication decision for 18 months. When the decision came from the Managing Editor in Chicago, in a letter dated May 24, 1951, it 'had a very peculiar character' (Coase 1981, p. 8). The Editor stated that he was willing to recommend the paper for publication "if the interrelationships with Arrow's recent monograph could be brought out clearly throughout the paper" (Coase 1981, p. 8). As Coase put it:

> Kenneth Arrow's monograph Social Choice and Individual Values had been published in 1951 shortly before the managing editor's letter was written. The suggestion that Black and Newing should revise a paper written and submitted for publication in 1949 so as to relate to a book which had appeared in 1951 (and which they had not even seen) was obviously completely unacceptable. (Coase, 194–195 of the reprint in *Essays on Economics and Economists* 1994).

Black and Newing withdrew the manuscript from *Econometrica*, and published it elsewhere as a little read short monograph in 1951 (Black and Newing 1951).

Those who concern themselves only with the advancement of ideas, and not with whom those ideas should be associated, may conclude that the Black-Arrow story is much ado about nothing. However, they would be incorrect, for the purpose and ultimate thrust of the two scholars differed sharply, with significant consequences for subsequent scholarship.

Duncan Black was concerned with understanding the strengths and weaknesses of simple majority voting in committees and elections and to discover institutional and procedural reforms that would remedy perceived defects. To this end, he wrote a sequence of papers following his 1948 masterpiece, designed to find avenues to ameliorate the problem of cyclical voting (Black 1948a, 1948b, 1949, 1950; Black and Newing 1951). With his work overshadowed by Arrow, however, these contributions largely fell upon deaf ears, at least until Anthony Downs re-awakened interest in the median voter theorem in 1957.

In contrast, Arrow was much less interested in understanding, navigating, and improving real-world voting mechanisms. The thrust of his research was much more abstract, and it tended to accentuate the philosopher-king approach to collective decision-making, indeed to accentuate the discretionary role of official social-decision-makers. Let me justify this interpretation by reference to Kenneth Arrow's own words when commenting on the following 1954 statement by Abram Bergson:

> I have been assuming that the concern of welfare economics is to counsel individual citizens generally. If a public official is counseled, it is on the same basis as any other citizen. In every instance reference is made to some ethical values which are appropriate for

the counseling of the individual in question. ... But some may be inclined nevertheless to a different conception, which allows still another interpretation of Arrow's theorem. *According to this view the problem is to counsel not citizens generally but public officials.* Furthermore, the values to be taken as data are not those which might guide the official if he were a private citizen. The official is envisaged instead as more or less neutral ethically. His one aim in life is to implement the values of other citizens as given by some rule of collective decision-making. (Bergson 1954).

Arrow's commentary on this statement is as follows:

> I need only add that my interpretation of the social choice problem agrees fully with that given by Bergson beginning with the italicized statement (Arrow 1963, p. 107).

The Burden of the National Debt

The perspective held by the classical economists on the burden of debt is best outlined by brief references to the writings of Adam Smith and David Ricardo.

Smith (1776) directly confronted the view prevalent among mercantilist thinkers of the late eighteenth century, that the public debt bears no burden: 'In the payment of interest of public debt, it has been said, it is the right hand which pays the left. This apology is founded, altogether, in the sophistry of the mercantile system' (Smith 1776/1976, pp. 926–927). Smith rejected this view on three grounds.

First, as Smith chronicled in the case of Britain, part of the British national debt was held by the Dutch, as well as by other foreign nations, in which case clearly 'we do not owe it to ourselves'. Second, even with respect to the internal debt, public debt does more than induce the annual transfer of interest from taxpayers to bondholders. It involves a reallocation of resources from productive to unproductive agents. Third, in a comment full of insight, Smith noted that the purchaser of a government bond undergoes no sacrifice at the time of purchase, when the debt is created, but rather makes money by lending to the government. In essence, therefore, the entire burden of the debt falls on future generations when the debt is redeemed:

> By lending money to government, they do not even for a moment diminish their ability to carry on their trade and manufactures. On the contrary, they commonly augment it . . . . The merchant or monied man makes money by lending money to the government, and instead of diminishing, increases his trading capital' (Smith 1776/1976, pp. 910–12).

Writing at the end of the Napoleonic Wars (1820), Ricardo, like Smith before him, emphasized that the primary burden of government war expenditures lies in the loss of the original capital as represented by the resources withdrawn from productive activity. Ricardo, like Smith, recognized that debt financing places no current burden on society.

Ricardo was more sanguine than Smith, however, concerning the debt burden actually placed upon future generations, arguing that future tax payments may be fully capitalized by rational citizens: 'The argument of charging posterity with the interest of our debt, or of relieving them from a portion of such interest, is often

used by otherwise well-informed people, but we confess to see no weight in it (1820/1951, p. 187). This should not be taken to imply that Ricardo accepted the actual equivalence between debt and taxes, as the following passage clearly proves:

> But the people who pay taxes never so estimate them, and therefore do not manage their private affairs accordingly. We are too apt to think that the war is burdensome only in proportion to what we are at the moment called to pay for it in taxes, without reflecting on the probable duration of such taxes. It would be difficult to convince a man possessed on 20,000L, or any other sum, that a perpetual payment of 50L per annum is equally burdensome with a single tax of 1000L. (Ricardo, 1820/1951, p. 186).

The classical economists all agreed that debt imposed some kind of debt on the future generations, even when all the debt was held internally. They differed somewhat concerning the precise nature of the burden. This consensus disappeared, however, in the wake of the Keynesian revolution.

The most categorical denouncement of the classical debt doctrine emanated from Abba Lerner (1948), possibly Keynes' most fervent disciple. Lerner acknowledged that external borrowing imposed a burden on future generations. He categorically denied that any such burden was imposed by the internal debt:

> Very few economists need to be reminded that if our children or grandchildren repay some of the national debt these payments will be made to our children or grandchildren and to nobody else. Taking them together they will be no more impoverished by making the repayments than they will be enriched by receiving them (Lerner 1948, p. 256).

Lerner displayed no overt understanding that he was reverting to the mercantilist doctrine of the late eighteenth century in arguing that *'we owe it to ourselves'*. He did recognize that the creation of debt was likely to involve income redistribution, but argued that such redistribution 'can be ignored because we have no more reason for supposing that the new distribution is worse than the old one as for assuming the opposite (Lerner 1948, p. 261). As a committed Keynesian, he rejected out of hand any suggestion that deficit-financed government expenditure might crowd out private investment. He concluded his essay by urging that 'the kinds of evil most popularly ascribed to national debt are wholly imaginary' (Lerner 1948, p. 275).

Lerner's views on national debt, even in 1948, were not fully shared by Paul Samuelson, then well on his way to becoming the leading Keynesian within the US academy. Certainly, Samuelson endorsed Lerner's view that internal debt imposes no direct burden on future generations:

> 'Can it be truthfully said that "internal borrowing shifts the war burden to future generations while taxing places it on the present generation?" A thousand times no!' (Samuelson 1948, p. 427).

Nevertheless, even in 1948, Samuelson was aware of the possibility that private investment might be deterred by the taxes required to finance interest on the debt. Even if each individual's taxes exactly met the debt interest that he received, a deadweight loss would occur as taxes distorted the relationship between work and leisure and adversely affected willingness to venture capital on risky projects. In

1948, Samuelson followed Lerner in making no reference to the possibility that increased government expenditure might crowd out private investment.

In the meantime, James Buchanan (1958) had mounted a major counter-revolution in favor of the position of Adam Smith. For a biography of James Buchanan, see Chapter 7 of this book. Starting with the classical notion of full-employment and with the assumption that the debt is created for real purposes, thus drawing resources entirely from private capital formation, Buchanan convincingly argued that the issuing of new debt imposes no burden on the current generation:

> The mere shifting of resources from private to public employment does not carry with it any implication of sacrifice or payment. If the shift takes place through the voluntary actions of private people, it is meaningless to speak of any sacrifice having taken place. An elemental recognition of the mutuality of advantage from trade is sufficient to show this. If an individual freely chooses to purchase a government bond, he is presumably moving to a preferred position on his utility surface by so doing. He has improved, not worsened, his lot by the transaction. This must be true for each bond purchaser, the individual who only actually gives up a current command over economic resources (Buchanan, 1958/1999, p. 28).

Buchanan drew the following inference from this analysis:

> The primary burden of the debt, in the only sense in which this concept is meaningful, must rest with the future generations, at least in large part. These are the only individuals who suffer the consequences of wasteful government expenditure and who reap the benefits of useful government expenditure. All other parties to the debt transactions are acting in accordance with ordinary economic motivations. (Buchanan, 1958/99, p. 37).
>
> Buchanan (1958, 1964) analyzes the public debt burden from the perspective of individual actors who collectively choose, through a democratic polity, both the level of government expenditure and the method of financing such expenditure. To understand this process better, Buchanan (1969) focused attention on the subjective nature of the opportunity cost considerations relevant to making such individual choices.

From this perspective, choice-influencing cost, the true opportunity cost, consists of each individual's own evaluation, in utility space, at the moment of his decision, of the sacrifice he is making in selecting his course of action. Subjective cost is imbued with the following important characteristics:

1. It is borne exclusively by those who choose; it cannot be shifted onto others who do not make the choice.
2. It is subjective, existing in the mind of each individual chooser, and nowhere else.
3. It is an *ex ante* concept, based on anticipations and not on retrospective calculations.
4. It is dated at the moment of choice and at no other point in time.

Quite distinct from the above-mentioned choice-influencing cost are the consequences that eventually flow from the choices that are made. Buchanan (1969) designates such consequences as choice-influenced cost. Choice-influenced cost does not reflect an evaluation of sacrificed alternatives, since such alternatives are in the past. Thus it does not represent opportunity cost. It is a future burden that emanates from current decisions.

Choice-influenced cost may be experienced both by individuals who were party to the original choice and by individuals who were not. The burden of the debt imposed on the future generations is of this latter type. At the moment of choice, rational individuals would not endorse a fiscal outcome unless anticipated benefits outweigh their respective choice-influencing cost.

Financing government expenditures by immediate taxation would raise choice-influencing cost. A vote to finance the same expenditures by long-term bonds designed to mature after the decision-makers are dead, does not do so, at least in the absence of some far-fetched Ricardian-equivalence (Buchanan 1964). Thus, not only are the future generations burdened with choice-influenced costs, but, because they cannot vote at the time of the decision, they play no role in the current collective choice. The opportunity to avoid such choice-influencing cost, at the moment of decision, thus biases democracies in favor of excessive levels of public expenditure.

By 1970, in the eighth edition of *Economics*, Samuelson was much more cautious concerning the burden placed by internal debt upon future generations. Specifically, he recognized the crowding out hypothesis, acknowledging that 'the main way that one generation can put a burden on a later generation is by using up currently the nation's stock of capital goods, or by failing to add the usual invest-ment increment to the stock of capital' (Samuelson 1970, p. 341). Yet, even as Paul Samuelson and his Keynesian colleagues shifted perceptibly back towards the classical viewpoint, nowhere within their camp was there any recognition of the insight of Adam Smith concerning the mutuality of exchange and its implications for the burden of the debt. The Keynesian intellectual establishment remained implacably hostile to any notion that debt financing placed a direct burden only on future generations.

Indeed, Samuelson (1970) reacted caustically to the notion that deficit-financing imposed any primary burden on future generations:

> We have seen that the public debt, prorated over the population, is sometimes regarded to be a load on each man's back. According to this same image, when Congress adds a dollar to the debt by running a current deficit of a dollar, that is like just one more rock added to the load our children or grandchildren will already have to carry on their backs. This image is misleading . . . .Suppose that all debt came from World War II. This war is over. Suppose that all America's families (1) share equally in ideal non-distorting taxes, (2) hold equal shares of public-debt bonds, (3) all live forever (as individual or as a cohesive family). With no debt abroad, (4) "we owe it all to ourselves" (Samuelson 1970, p. 351).

The devil, of course, lies in the assumptions. Samuelson hides behind the above-mentioned set of unrealistic assumptions designed to rule out any future genera-tions' problem. Just as Samuelson (1963) characterized Milton Friedman's alleged misuse of assumptions designed to justify his methodology of positive economics as the F-Twist, so I shall designate Samuelson's use of the above-mentioned burden of the debt assumptions as the S-Twist.

The motivation behind the S-Twist may well have been ideological rather than scientific. Samuelson would be well aware that a shift from deficit-financing of public expenditures to taxation under a balanced-budget rule would significantly

constrain the fiscal role of government under conditions of democracy. As a liberal democrat, this would be a very hard pill to swallow.

In any event, James Buchanan's 1964 predictions, as further elaborated in his 1978 monograph co-authored with Richard Wagner (Buchanan and Wagner, 1977) have come to pass with a vengeance. Throughout the period 1961 to 2007, U.S. governments have delivered budget deficits in all but three years, irrespective as to whether the economy has been in boom or in recession. Thus has the intellectual establishment blocked out a public choice trailblazing insight, arguably with serious long-term consequences for future generations. James Buchanan (1987) sums up the fiscal masquerade as follows:

> While confusion and ambiguity described economists' discussion of public debt, the politicians had learned the Keynesian policy lessons with roughly a two-decade lag. By the early 1960s, the 'old-time fiscal religion', based on adherence to the normative precepts of the classical analysis, had lost its constraining influence. The political leaders of the 1960s and beyond had learned that demand-enhancing deficits may be justified in some economic settings. Their natural proclivities to spend without the levy of taxes on consti-tuents caused them to look on economic settings in a biased or one-sided fashion. The idealized Keynesian policy set—deficits in depression, surpluses in booms—proved to be unworkable in democratic politics. (Buchanan 1987, p. 1046).

The Nature and Political Impact of Interest Groups

The notion of 'interest group' goes back into antiquity, starting perhaps with the claim by Thrasymachus, in Plato's *Republic*, that 'justice is nothing else than the interest of the stronger'. In the dialogue that follows, Socrates and Thrasymachus debate the idea of interest, finding the notion hard to specify because each individual appears to have both a self and a collective interest and that these interests may conflict with each other

Research on interest groups goes back certainly to Cicero who distinguishes parties from factions. Parties, in the view of Cicero, seek to discover the common good, whereas factions work only in their respective interests. This distinction surely influenced the American Founding Fathers in their efforts to create a Republic that might withstand the adverse impulses of faction. James Madison and Alexander Hamilton succinctly identify the tension between party and faction in *Federalist* 10 and *Federalist* 51:

> The latent causes of faction are thus sown in the nature of man and we see them everywhere brought into different degrees of activity, according to the different circumstances of a civil society. The regulation of these various and interfering interests forms the principal task of modern legislation, and involves the spirit of party and faction in the necessary and ordinary operations of the government. (Madison, *Federalist* #10).

and

> In framing a government which is to be administered by men over men, the great difficulty lies in this: you must first enable the government to control the governed: and in the next

place oblige it to control itself. This policy of supplying, by opposite and rival interests, the defect of better motives, might be traced through the whole system of human affairs, private as well as public. (Hamilton and Madison, *Federalist* #51).

The American Founders clearly recognized that organized interest groups, or factions, aim to manipulate politics to their own advantage and rarely display concern for the well-being of the other members of society. They believed, perhaps erroneously, that the adverse effects of faction could be limited by a territorially large, diverse republic, invested with separation of powers among its governing institutions. In general, however, both the Founders and the enlightenment scholars who influenced their thinking, held a very dim view of factions.

By the turn of the twentieth century, however, such concerns of political economy had been eroded and a more optimistic view about interest groups had begun to emerge. In his influential 1901 book, *Representative Democracy*, the institutional economist, John R. Commons, shifted attention away from the geographical in favor of the occupational interest groups. Commons evidenced strong support for occupational interest groups, arguing that they were the most representative and beneficial forces affecting American economic policy.

Indeed, Commons went so far as to advocate direct election of representatives for each interest group as the basis for the effective legislature of the country. In his judgment, economic interest groups were more representative of the people than were legislatures based on territorial representation. In this respect, Commons' thinking anticipated the emergence of Benito Mussolini's Italian occupational state (Olson 1965, p. 116). It also anticipated John Kenneth Galbraith's 1952 concept of 'countervailing' power.

John Commons' institutional economic views failed to attract much interest within an economics profession still dominated by the classical economic models. However, they provided an intellectual platform seized upon by a number of influential 'analytical-purist' political scientists anxious to distance themselves from methodological individualism and to develop alternative theories of group behavior. The most important of these was Arthur F. Bentley's 1908 book *The Process of Government* that advanced the hypothesis that pressure groups play a dominant role in economic and political life.

According to Bentley, '[t]he great task in the study of any form of social life is the analysis of . . . groups'; and 'when the groups are adequately stated, everything is stated'. It is group *interests*, moreover, that are basic. 'There is no group without its interest'. Whereas group interests are everything, individual interests are nothing. What matters are the common interests of groups of people and not the gains and losses of individuals.

In such circumstances, conflicting group pressures completely explain government policies, with groups exercising pressure more or less in proportion to their numbers. The logrolling of special interests is an excellent device for adjusting group interests. For the most part, group pressures improve on the performance of democracy by allowing preference intensities to register in the policy-making process. Bentley was an enthusiast and an apologist for group participation in the political process.

Bentley paid little attention to the factors that caused groups to organize and to act effectively through the political process. His political science disciples attempted to fill this lacuna. In particular, David B. Truman's influential 1951 book *The Governmental Process* explores the issue in some detail, developing a sociological theory of voluntary associations to argue that effective group pressures emerge as societies become more complex. In this perspective, suffering, dislocation and disturbance almost inevitably result in organized political pressure as disadvantaged individuals form themselves into effective groups.

Truman mirrors Bentley in his belief that group pressures alone determine equilibrium through the political process. He suggests, more strongly even than Bentley, that group equilibrium is just and desirable, arguing that groups would become weak and divided if they asked too much and would thereby expose themselves to potential competition from new, better adjusted groups. From this perspective, Truman belittles all attempts to improve the existing system of legislation and lobbying (Olson, 1965, p. 125).

In 1965, in his book *The Logic of Collective Action*, a young trailblazing economist, Mancur Olson, challenged head-on the Panglossian thinking of conventional political scientists concerning the nature and role of interest groups under conditions of democracy. In so doing, Olson justly earned his place among the founding fathers of public choice (Rowley 1999). For a short biography of Mancur Olson see Chapter 9 of this book.

In essence, Olson (1965) demonstrates that rational individuals, at least when solicited for support by *large* interest groups, would tend to free ride in the pursuit of their common interests. This free-riding is a direct consequence of the 'publicness' characteristics of most interest groups' objectives and would result in an under-provision of interest group pressure on government:

> If the members of a large group rationally seek to maximize their personal welfare, they will *not* act to advance their common or group objectives unless there is coercion to force them to do so, or unless some separate incentive, distinct from the achievement of the common or group interest, is offered to the members of the group individually on the condition that they bear the costs or burdens involved in the achievement of the group objectives. Nor will such large groups form organizations to further their common goals in the absence of the coercion or the separate incentives just mentioned. These points hold true even when there is unanimous agreement in a group about the common good and the methods of enforcing it. (Olson 1965, p. 2).

Olson further demonstrates that, even in the smallest interest group, the public good (bad) will not ordinarily be provided on an optimal scale. That is to say, the members will not provide as much of the commodity as it would be in their common interest to provide. This tendency toward sub-optimal provision lies in the fact that other members of the group cannot be excluded from consuming the commodity once an individual has provided it for himself. The only exception is the case where a group is composed of members of greatly different size and/or interest in the public good. In such circumstances, the largest member may exert pressure on its own equal to the optimal pressure for the group as a whole, a situation that Olson labels as 'exploitation of the great by the small'. A possible example of this

was the role played by the United States in pressing successfully for the creation of the North Atlantic Treaty Alliance during the early phase of the Cold War.

Nevertheless, small groups typically are more effective in supplying pressure than are large groups:

> In this, small groups differ from larger ones. The larger the group is, the farther it will fall short of obtaining an optimal supply of any collective good, and the less likely that it will obtain even a minimal amount of such a good. In short, the larger the group, the less it will further its common interests. (Olson 1965, p. 36).

Olson thus establishes three exceptions to the general expectation that interest groups will under-provide political pressure in pursuit of their members' respective goals. The first exception is where the leaders of an interest group are able to coerce members to provide pressure. Examples of this are labor unions and trial lawyers' associations. The second exception is where groups are sufficiently small in number, and sufficiently homogeneous in nature to overcome free-riding by a mixture of economic and social incentives. Examples of this are producer groups, like the petroleum lobby, the tobacco lobby, the farming lobby and the automobile lobby. The third exception is where the leaders of large, heterogeneous groups are able to lock in members' pressure by offering selective benefits of a private nature in return for overall membership fees. An example of this is the American Association of Retired Persons which provides access to medical and life insurance benefits to members, on the basis of a large membership, on superior terms to those available (if available at all) to individuals bargaining on their own.

The implications of Olson's analysis suggest that political systems that react to lobbying pressures (i.e. all political systems) will concentrate political benefits on the more influential, typically concentrated, interest groups while laying off the political costs onto large, dispersed groups and onto individuals who cannot organize themselves into groups (for example taxpayers and consumers). In so doing, the political system may diverge significantly in its implemented policies from the expressed electoral preferences of the median voter.

Although *The Logic of Collective Action* was primarily a theoretical treatise, Olson was an empiricist of the first order, deeply concerned to make sense of the world around him. Acutely aware of the differential economic success of the victors and losers in World War II, with West Germany and Japan out-performing the United States and the United Kingdom, Olson turned the searchlight of the *Logic* onto the conundrum. The result was his second public choice masterpiece, *The Rise and Decline of Nations* (Olson 1982).

In *The Rise and Decline of Nations*, Olson argues that, because the free rider problem is pervasive, even the best advantaged interest groups find great difficulty in organizing and effectively invading the political system in pursuit of favors. Once established, however, they have significant sunk cost advantages over potential new entrants and are able to exercise muscle to protect the interests of their members from economic change. Countries that experience long periods of political stability (the United States and the United Kingdom) tend to suffer from internal economic sclerosis, as powerful interest groups lobby as successful Luddites in defending their

members from economic innovations. In contrast, in countries that have recently experienced political upheaval, as a consequence of internal revolution or unconditional defeat in war, pre-existing interest groups typically are severely weakened (as in Japan) or completely wiped out (as in West Germany). Because of the free rider problem, new interest groups are slow to merge. In the interim, the channels through which economic change flows are much less obstructed.

Taken together, *The Logic (1965)* and *Rise and Decline* (1982) identify interest groups as a significant source both of economic inefficiency and of serious potential political market failure. These trailblazing results, though widely accepted by political scientists and sociologists, as well as by public choice scholars, would not long be allowed to stand without a major challenge from the Chicago political economy program, this time in the form of two articles by Gary Becker (Becker 1983, 1985).

Prior to Becker's contributions, the Chicago political economy had taken no position on the role of interest groups except for Stigler's (1974) critique of Olson's (1965) selective benefits model of interest group formation. Becker followed this lead by his colleague and mentor (Becker 1983) with a theory of competition among pressure groups for political influence predicated on the assumption of widespread rational ignorance among the voting public. In direct contradiction to Olson's emphasis on the harmful political consequences of interest group activity, Becker steadfastly followed the developing Chicago political economy tradition with an optimistic political market theory in which wealth redistribution through pressure group competition occurs at least economic cost to society:

> Political equilibrium depends on the efficiency of each group in producing pressure, the effect of additional pressure on their influence, the number of persons in different groups, and the deadweight cost of taxes and subsidies. An increase in deadweight costs discourages pressure by subsidized groups and encourages pressure by taxpayers. This analysis unifies the view that governments correct market failures with the view that they favor the politically powerful: both are produced by the competition for political favors. (Becker 1983, p. 371).

In making his case, Becker (1983) relies on a number of unrealistic Chicago-type assumptions, without, for the most part, acknowledging any shift from those deployed by Olson (1965 and 1982). First, rather than dealing with obviously high excess burden transfers in kind, Becker assumes in his model that all transfers take the form of lower excess burden taxes and subsidies (even he does not imply that such transfers are always lump-sum in nature). Second, rather than deal with the possibility that interest group pressures might create or exacerbate government fiscal deficits, Becker (1983) blithely assumes that a balanced budget constraint always holds. Third, although Becker (1983) recognizes that the cost of controlling free-riding will affect the size of pressure groups (incidentally without any explicit recognition of Olson on this point), he does not acknowledge that free-riding and other transaction costs may prevent the formation of many potential pressure groups, even where effective pressure might conceivably reduce the excess burden of wealth transfers. Fourth, and crucially, Becker (1983) assumes that individuals

who may exert pressure through interest groups are motivated by private returns. Nowhere, other than in a brief reference to a potential free-riding problem in large interest groups, does Becker acknowledge that benefits from interest group lobbying primarily take the form of public good or public bad. Becker (1983) cites Olson's 1965 paper in the references, but nowhere in the text, of his paper.

In 1985, Becker (1985/88) extended his pressure group model to encompass altruism and envy, in addition to solipsism, among members of pressure groups. He also introduced explicitly the principle of the potential compensation test as a means of evaluating the social welfare implications of pressure group activities. It becomes quickly evident that the purpose of these adjustments, for the most part, is to consolidate his earlier view that most pressure group activity is socially beneficial:

> Some policies might raise social output because of altruism by taxpayers or envy by recipients. Redistributions are Pareto-improving when altruistic taxpayers also benefit. Although altruists would be harmed by redistributions beyond the Pareto-efficient point, social output would be increased as long as the monetary value of the gains to beneficiaries exceeds the monetary value of the loss to altruists. (Becker 1985/1988, p. 94)

> and

> If the *intent* of public policies were fully known, I am confident that the public sector would be revealed to be a far more efficient producer and distributor than is popularly believed... Redistribution should be included among the 'measured outputs' of public and regulated enterprises before one can conclude that they are less efficient than private enterprises. (Becker 1985/88, p. 95)

Becker (1985/88) critically addresses Olson's (1982) assertion that established interest groups are responsible for sluggish economic growth and, eventually, for the economic decline of nations. In Becker's (1985/88) judgment, this view is excessively pessimistic:

> the condemnation of special interest groups is excessive because competition among these groups contributes to the survival of policies that raise output: favorably affected groups tend to lobby more for those policies than unfavorably affected groups lobby against. Indeed, no policy that lowered social output would survive if all groups were equally large and skillful at producing political influence, for the opposition would always exert more influence than proponents. (Becker 1985/88, p. 102).

Becker's Panglossian model of interest group behavior under conditions of democracy runs starkly counter to Olson's logic of collective action. Without accounting for any of the empirical work that supports Olson's theory, Becker airily replaces it with a model founded on extremely suspect assumptions. Of course, the opportunity afforded by Becker, a leading member of the Freshwater academy, for those well-disposed of by government to mark down the trailblazing theoretical and empirical work of Olson, was seized upon and an influential literature has now emerged dedicated to the glorification of democratic government (Wittman, 1989, 1995). That literature is not supported by any significant body of empirical evidence. It is long on idolatry and singularly short on facts.

Rent-Seeking and the Efficiency of Government Legislation

Prior to 1967, when Gordon Tullock (for a biography see chapter 8 of this book) published his seminal paper on the '*welfare costs of tariffs, monopolies and theft*' (Tullock 1967) economists regarded legislation largely as the public-spirited response by government to the expressed preferences either of a majority of its electorate or of its well-informed interest groups. Even those who were acquainted with Olson's (1965) book viewed the costs associated with such legislation as minimal, or as simply the ordinary transaction costs of the political process. For example, if government legislates in favor of a tariff, or a monopoly, this legislation redistributes producers' and consumers' surplus among those involved while reducing overall economic welfare only by the deadweight loss, excess burdens involved (the so-called Marshallian or Harberger triangles of welfare loss). Such welfare losses, even in a highly monopolized society, tend to be small, relative to the overall benefits generated by the activities involved. Even theft can be viewed, according to this perspective, primarily as a socially costless transfer of resources between the thief and his victim.

Tullock was far from convinced by such Panglossian analysis. In his 1967 paper (Tullock 1967) he centered attention on the rectangle of producers' surplus transferred from consumers as a consequence of the creation of a monopoly. Surely, he argued, the potential for such a transfer gain would induce wasteful expenditures lobbying government in order to secure the monopoly (or tariff) under consideration:

> Surely we should expect that with a prize of this size dangling before our eyes, potential monopolists would be willing to invest large resources in the activity of monopolizing. In fact the investment that could be profitably made in forming a monopoly would be larger than this rectangle, since it represents merely the income transfer. The capital value, properly discounted for risk, would be worth much more. Entrepreneurs should be willing to invest resources in attempts to form a monopoly until the marginal cost equals the properly discounted return. The potential customers would also be interested in preventing the transfer and should be willing to make large investments to that end. (Tullock 1967, p. 231)

Tullock noted that the welfare cost of securing a monopoly is not limited to the outlays made by the winning firm. Rather it is measured by the total outlays of all the would-be monopolists, losers as well as winners. These additional welfare costs, he argued, might be ameliorated either by free trade or by an active antitrust policy. In 1967, Tullock had not yet completely blended 'rent-seeking', as it eventually was named by Anne Krueger in 1974, with public choice theory.

In his 1971 paper on *the cost of transfers*, Tullock extended his rent-seeking theory to income redistribution. Even with respect to entirely private charitable giving, he argued that competitive rent-seeking on the part of would-be beneficiaries of such transfers wasted resources, as those seeking charity render themselves appealing objects for it. With respect to income transfers mediated through the welfare state, he suggested that rent-seeking is endemic, with perhaps as much as 95 per cent off all such transfers involuntary from better-off members of society

coerced by middle-class lobbies. In two follow-up papers (Tullock 1974 and 1975a), he outlined the working of exactly the same principle when lower-level governments wastefully compete for aid from higher-level governments in a federal system by rendering their own programs ineffective.

In his well-known 1975 paper on *the transitional gains trap* (Tullock 1975b), Tullock explains why many government programs do not appear to benefit their targeted recipients for very long. Basically, Tullock suggests that initial benefits are quickly capitalized and often sold to others at rent-inflated capital values. Subsequent holders of such assets thus earn only a normal return on their investments. If the privilege should ever be withdrawn, the later entrants would incur a capital loss. To avoid such an outcome, the holders of rents will lobby aggressively against any such removal of privilege.

In his 1980 paper on *efficient rent-seeking*, Tullock deploys game theory to define the circumstances under which competing rent seekers exactly dissipate the rent available (he labels this 'efficient rent-seeking'), the circumstances under which they under-dissipate rents and the circumstances under which they over-dissipate rents. Typically, he suggests, rents will be under-dissipated through political rent-seeking not least because of the logic of collective action. Typically, rents will be transferred at considerable social cost and rarely without any social cost. Available evidence supports both hypotheses.

Mancur Olson's 1965 book, *The Logic of Collective Action*, based on his 1963 doctoral dissertation, was completed well before Tullock's 1967 paper, and, therefore could not incorporate the rent-seeking insight. His 1982 book, *The Rise and Decline of Nations*, however, recognized and applied Tullock's insights with powerful effect. By 1982, the trailblazing rent-seeking contribution, with its implications for the failure of political markets, was fully embedded in the literature of economics.

Such knowledge was not embedded even by 1982 in the thinking of members of the Chicago School of Political Economy, as first became evident in Becker's 1983 and 1985 papers on interest groups. In his 1983 paper, Becker fails to cite any of Gordon Tullock's contributions to the rent-seeking literature. Throughout his 1983 paper, he focuses attention exclusively on the Marshallian (or Harberger) dead-weight loss triangles and never on the Tullock rectangles. In his 1985 paper, Becker once again sidesteps any reference to rent-seeking behavior, whether by Tullock or by any of a large number of contributing scholars. It is almost as though a rapidly growing literature remained entirely outside his comprehension. The only other explanation is that Becker's methodological emphasis on efficiently functioning political markets might have been grounded had he taken account of 'Tullock rectangles' as well as 'Harberger triangles'.

If Becker's contributions provide primarily a passive resistance to the public choice theory of political market failure, largely by sidestepping the issues that public choice trailblazers like Tullock have developed, George Stigler's contributions take on a much more aggressive tone. As we noted earlier, George Stigler was the founding father of Chicago Political Economy (Stigler 1971) and, until his death in 1992, the acknowledged leader of the research program (Stigler 1988) which combines regular

price theory with positive economics in its analysis of political market behavior (Mitchell 1989, Tollison 1989). Always skeptical about the ability of economists to influence economic policy, (Stigler 1982) as he aged, Stigler became ever more committed to the notion that political markets were efficient (though he was ambiguous as to whether such efficiency was political or economic). His posthumous 1992 publication (*law or economics*?) effectively summarizes his willingness to apply the Coase theorem (Coase 1960) well beyond private exchange, and the common law, to the law of legislation, essentially turning public choice on its head.

Stigler (1966) first introduced the concept of the Coase theorem to economics albeit focusing attention at that time exclusively on the zero transaction cost model. In his 1992 paper, he was more catholic in his definition:

> Coase reminded economists and taught lawyers that, in a world of exchange by agreement rather than by coercion, the costs and benefits of agreement determine its scope. Because agreements can be costly, many will not be struck, and these unachieved agreements will have been inhibited by the smallness of the benefits or the largeness of the costs of agreement. (Stigler 1992, p. 456).

Stigler defines maximum efficiency to infer 'that a given goal is achieved as best as one knows how' (ibid., 458). He recognizes that the economist's conventional concept of efficiency 'turns on the maximization of the output of an economic process or of an economy (ibid., 458). However, in the case of political markets, he suggests that a broader meaning can be inferred:

> In policy analysis, one may legitimately imply an alternative definition of efficiency that rests on the goals adopted by the society through its government. When a society wishes, for example, to give more income to a group, than the market provides, we may surely analyze the efficiency with which this is done. (Stigler 1992, pp. 458–459).

He deploys this new definition, some may think outrageously, to suggest that all durable social institutions, including common and statute laws must be efficient:

> In this latter view, every durable social institution or practice is efficient, or it would not persist over time. New and experimental institutions or practices will rise to challenge the existing systems. Often the new challenges will prove to be inefficient or counterproductive, but occasionally they will succeed in replacing the older system. Tested institutions and practices found wanting will not survive in a world of rational people. To believe the opposite is to assume that the goals are not desirable: who would defend a costly practice that produces nothing? (Stigler, 1992, p. 459).

Stigler's deconstruction of the notion of efficiency with respect to political markets, without any reference to the institutions and agents that form that market-place, without any reference to the logic of collective action or to the unnecessary costs of rent-seeking behavior, boils down to a remarkable judgment that 'whatever exists is efficient' (Rowley 1997). If one accepts such a judgment, then there is little or no place for public choice analysis or even for constitutional political economy.

All that is necessary for efficiency is democracy—democracy of any kind, anywhere, irrespective of the institutions that emerge, just as long as democratically-established institutions persist over time. Any old tariff is a good tariff. Any old quota is a good quota. Any old public enterprise is a good public enterprise. Any

old income transfer system is a good system. Hail to the United States Post Office, hail to state-run Departments of Motor Vehicles! Hail to sugar-beet subsidy programs! Get lost, public choice!

Conclusions

It is somewhat ironic that key insights of four important public choice trailblazers ran head on into the adverse winds of the intellectual economics establishment, on two occasions (Black and Buchanan) from the left and on two occasions (Olson and Tullock) from the right. In part, no doubt, these adverse winds stemmed from differences of methodology and of science. In part, no doubt, they stemmed from ideology and reputation-protection.

In any event, ultimately, these adverse winds helped rather than hindered the progress of public choice and constitutional political economy. Battle-hardened, the trailblazers pressed forward with their insights against opposing forces with a vigor and zest that might have been less evident had the establishment initially proved more welcoming to their then-alien ideas. Now that the ideas of the trailblazers have become part of the intellectual economics establishment, their public choice successors perhaps should think long and hard before abruptly rejecting or side-stepping equally alien ideas that may emanate in the future from an even younger generation of thinkers.

References

Arrow, K.J. (1950). 'A Difficulty in the Concept of Social Welfare', *Journal of Political Economy* 58, 328–346.

Arrow, K.J. (1951/1963). *Social Choice and Individual Values*. New York: Wiley, 2nd edn.

Arrow, K.J. and Debreu, G. (1954). 'Existence of an equilibrium for a competitive economy', *Econometrica* 22, 265–290.

Becker, G.S. (1983). 'A theory of competition among pressure groups for political influence', *Quarterly Journal of Economics* 98, 371–400.

Becker, G.S. (1985). 'Public policies, pressure groups, and dead weight costs', *Journal of Public Economics* 28, 329–347.

Bentley, A.F. (1908). *The Process of Government*. Chicago: University of Chicago Press.

Bergson, A. (1954). 'On the concept of social welfare', *Quarterly Journal of Economics* 68, 233–252.

Black, D. (1948a). 'On the rationality of group decision-making', *Journal of Political Economy* 56 (1), 23–34.

Black, D. (1948b). 'The decisions of a committee using a special majority', *Econometrica* 16, 245–261.

Black, D. (1949). 'The theory of election in single-member constituencies', *Canadian Journal of Economics and Political Science* 14(2), 158–175.

Black, D. (1950). 'The unity of political and economic sciences', *Economic Journal* 60, 506–514.

Black, D. and Newing, R.A. (1951). *Committee Decisions with Complementary Valuation*. Glasgow: William Hodge.

Buchanan, J.M. (1958). *Public Principles of Public Debt: A Defense and Restatement*. New York: Irwin.

Buchanan, J.M. (1964). 'Public debt, cost theory, and the fiscal illusion', In James M. Ferguson (ed.), *Public Debt and Future Generations*. Chapel Hill: University of North Carolina Press, pp. 150–163.

Buchanan, J.M. (1969). *Cost and Choice*. New York: Markham.

Buchanan, J.M. (1987). 'Keynesian Follies', In David A. Reese (ed.), *The Legacy of Keynes*. San Francisco: Harper Collins Publishers, pp. 130–145.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Buchanan, J.M. and Wagner, R.E. (1977). *Democracy in Deficit: The Political Legacy of Lord Keynes*. New York: Academic Press.

Coase, R.H. (1960). 'The problem of social cost', *Journal of Law & Economics* 3, 1–44.

Coase, R.H. (1981). 'Duncan Black: A biographical sketch', in G. Tullock (ed.), *Toward a Science of Politics: Essays in Honor of Duncan Black*. Blacksburg, Virginia: Center for Study of Public Choice, pp. 1–10.

Commons, J.R. (1901). *Representative Government*. New York: National Bureau for Economic Research.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Galbraith, J.K. (1952/1993). *American Capitalism: The Concept of Countervailing Power*. New Jersey: Transaction Publishers.

Grofman, B. (1981). 'The theory of committees and elections: The legacy of Duncan Black', In G. Tullock (ed.), *Toward a Science of Politics: Papers in Honor of Duncan Black*. Blacksburg, Virginia: Center for Study of Public Choice, pp. 11–58.

Krueger, A. (1974). 'The political economy of the rent-seeking society', *American Economic Review* 64, 291–303.

Lerner, A.P. (1948). 'The burden of the national debt', in *Income, Employment, and Public Policy: Essays in Honor of Alvin H. Hansen*. New York: Norton.

Madison, J. (1787). *Federalist* No. 51.

Mitchell, W.C. (1989). 'Chicago political economy: A public choice perspective', *Public Choice* 63(3), 282–292.

Olson, M. (1965). *The Logic of Collective Action*. Cambridge: Harvard University Press.

Olson, M. (1982). *The Rise and Decline of Nations: Economic Growth, Stagflation and Social Rigidities*. New Haven: Yale University Press.

Reder, M.W. (1982). 'Chicago economics: Permanence and change', *Journal of Economic Literature* 20, 1–38.

Ricardo, D. (1820/1951). 'Funding system', In P. Sraffa and M.H. Dobb (eds.), *Pamphlets and Papers*, 1815–1823, *The Works and Correspondence of David Ricardo*. Volume 4, Cambridge: Cambridge University Press, pp. 143–200.

Rowley, C.K. (1993). *The Right To Justice*. Aldershot, UK: Edward Elgar Publishing.

Rowley, C.K. (1999). 'Five market friendly Nobelists: Friedman, Stigler, Buchanan, Coase and Becker', *The Independent Review* 3, 413–432.

Samuelson, P.A. (1947). *Foundations of Economic Analysis*. Cambridge: Harvard University Press.

Samuelson, P.A. (1948). *Economics: An Introductory Analysis*. New York: McGraw-Hill, 1st edn.

Samuelson, P.A. (1963). 'Comment on Ernest Nagel's "Assumptions in Economic Theory."' *American Economic Review, Papers and Proceedings of the American Economic Association*. (May), pp. 211–219.

Samuelson, P.A. (1970). *Economics: An Introductory Analysis*. New York: McGraw-Hill, 8th edn.

Smith, A. (1776/1976). *An Inquiry into the Nature and Causes of the Wealth of Nations: Glasgow Edition of the Works and Correspondence of Adam Smith*. Volume 2, Book 5, Chapter 3, Oxford: Clarendon Press, pp. 907–947.

Stigler, G.J. (1947). *Production and Distribution Theories: 1870–1895*. New York: Macmillan.

Stigler, G.J. (1966). *The Theory of Price*. New York: Macmillan, 3rd edn.

Stigler, G.J. (1971). 'The theory of economic regulation', *Bell Journal of Economics and Management* 2, 3–21.

Stigler, G.J. (1982). *The Economist as Preacher*. Oxford: Basil Blackwell.

Stigler, G.J. (1988a). *Memoirs of an Unregulated Economist*. New York: Basic Books.

Stigler, G.J. (ed.) (1988b). *Chicago Studies in Political Economy*. Chicago: University of Chicago Press.

Stigler, G.J. (1992). 'Law or Economics?' *Journal of Law and Economics* 35(2), 455–468.

Tollison, R.D. (1989). 'Chicago political economy', *Public Choice* 63(3), 293–298.

Truman, D.B. (1951). *The Governmental Process: Political Interests and Public Opinion*. New York: Alfred A. Knopf.

Tullock, G. (1967). 'The welfare costs of tariffs, monopolies and theft', *Western Economic Journal* 5, 224–232.

Tullock, G. (1971). 'The cost of transfers', *Kyklos* 24(fasc. 4), 629–643.

Tullock, G. (1974). 'More on the welfare costs of transfers', *Kyklos* 27(fasc. 2), 378–381.

Tullock, G. (1975a). 'Competing for aid', *Public Choice* 21, 41–51.

Tullock, G. (1975b). 'The transitional gains trap', *Bell Journal of Economics and Management Science* 6, 671–678.

Tullock, G. (1980). 'Efficient rent seeking', in J.M. Buchanan, R.D. Tollison and G. Tullock (eds.), *Toward a Theory of the Rent-Seeking Society*. College Station: Texas A & M University, 97–112.

Part II
Biographies of the Founding Fathers

Chapter 4
Duncan Black (1908–1991)[1]

Charles K. Rowley[2]

Biography

Duncan Black was born on May 23, 1908, in Motherwell, part of the industrial belt of central Scotland, located some 12 miles from Glasgow. He was raised in the quaint village of Tayvallich, Argyle, with its spectacular views of Loch Sween, in the same general area of Scotland. His father, a boiler-maker, was a Highlander, born in the Western Isle of Mull. His mother, a Motherwell Lowlander, owned a drapery and millinery shop until 1931, when the expansion of public transportation (buses) opened up the Glasgow stores to eager Motherwell shoppers and bankrupted her store (Rowley, 1991, p. 83).

According to Black's former colleague Ronald Coase (1981), Black's highly intelligent mother was the dominant influence in his life, followed by his Scottish Presbyterian upbringing. Both influences instilled in him a discipline and a work ethic absent which real scholarly achievement is all but unattainable. Black benefited from a rigorous and formal education then available for clever children of working-class parents (McLean, McMillan, and Monroe, 1996, p. xvi). He graduated from Dalziel High School in 1926 and enrolled at the University of Glasgow to study mathematics and physics, graduating in 1929 with a mediocre second class honors Master of Arts degree (the M.A. is the first degree in Scotland). Black's interest in logic and pure mathematics had been disappointed by a degree program in which the approach was much more practical, geared towards engineering students (Coase, 1981).

Disappointed with this experience, Black contemplated a future career in the civil service. To this end, he enrolled in an M.A. course in economics and politics, also at the University of Glasgow, graduating in 1932, this time with a first-class honors degree. During this degree program, Black was influenced by the teachings of W.R. Scott, a distinguished economist whose interest was in the relationship

---

between economics and philosophy, and most especially by A.K. White, a political scientist who fired Black with enthusiasm for constructing a pure science of politics, comparable to the general equilibrium analysis of economic theory.

In 1932, Black was appointed Assistant Lecturer in the School of Economics and Commerce at Dundee University, a School that had been established under the intellectual guidance of leading members of the London School of Economics (LSE) (including Sir William Beveridge). Not surprisingly, his three economist colleagues at Dundee, J.K. Eastham, J.C. Gilbert, and a very young Ronald H. Coase, had all been educated in economic theory at the LSE. They would provide the fulcrum for Black's two-year exposure to a cutting-edge economic theory that simply was unavailable at the University of Glasgow.

In 1934, Black accepted a lectureship in economics at the University College of North Wales, Bangor, where he remained (except for a period of wartime duty as a member of the British civil service) until 1945. In 1945, he accepted a senior lectureship in economics at Queen's College, Belfast, but one year later, in 1946, he returned to his alma mater, the University of Glasgow, also as senior lecturer in economics. In 1952, Black returned to the University of Wales, Bangor, as Professor of Economics where he remained until his retirement in 1968 at the comparatively early age of 60 years.

Black's career pattern of movement between Scotland, Wales, and Northern Ireland, but never to England, was not uncommon among academic Scots of Black's generation, partly because the structure of academic education differed greatly between the Anglo-Saxon heartland and the Celtic fringes of the United Kingdom (McLean, McMillan, and Monroe, 1996, p. xvi). An unfortunate consequence of this segmentation, as I shall explain in Section 2, was Black's marginalization within the British academy throughout his academic career, a factor, no doubt, that led to his early retirement decision.

As Coase (1981) has noted, however, there was a much greater interest in Black's scholarship in the more open, entrepreneurial, and energetic economics profession in the United States than in the United Kingdom during the middle years of the twentieth century: "once more, the New Word was called in to redress the balance of the Old" (Coase, 1981, p. 9). Black's (1958) work on committees and elections quickly attracted the attention of James Buchanan and Gordon Tullock at the Thomas Jefferson Center for Political Economy in the University of Virginia. Black first visited the University of Virginia as a Visiting Professor in 1962. In 1963, he was Visiting Professor in William Riker's Department of Political Science at the University of Rochester.

Following his retirement from Bangor in 1968, Black visited the United States on many occasions. He was a Research Fellow in Law and Economics at the University of Chicago (1968–1969), and a National Science Foundation Fellow at Virginia Polytechnic Institute (1970–1971). He was a Visiting Professor in the Department of Political Science at the University of Chicago in 1969, 1972, 1973, and 1976. He was a Visiting Professor in the Department of Political Science at Michigan State University in 1971, 1972, 1973, 1975, and 1976. These visits were a source of great pleasure for this neglected scholar, most especially his visits to

Michigan State University, which almost became his second home during the 1970s. At long last, the lonely prophet had found an intellectual home.

Following the death of his beloved and highly supportive wife of 31 years, Almut, in 1977, and without any children, Duncan Black continued his scholarship (see Section 2) quietly from his home in Cambridge, England, before moving, finally, to the lovely coastal Devonshire town of Paignton in southwest England, where he died (alone) on January 14, 1991, at 83 years of age.

The Father of All of Us

"I'd like to close (my obituary) by repeating something I said in my salute to Black in the (1981) Festschrift for him. The Chinese at a dinner will drink a toast to the host calling him "the father of all of us." Black indeed was the father of all of us." (Tullock 1991, p. 128).

The debt that we owe Black for his historical scholarship is immense, but it is outweighed in importance by the magnitude of Black's own research contributions to the pure theory of politics in calling attention to the importance of "procedures, agendas, and the search for principles which govern the behavior of voting processes" (Plott, personal communication, February 5, 1980). It can, with considerable justice, be claimed that Black was the first "public choice" economist. Of him (as of Condorcet and Carroll) it can be said, "And here be giants." (Grofman, 1981, p. 46).

Black's scholarship is so central to public choice that it permeates much of this book (see especially Chapters 1, 2, 3, 12, and 18). Therefore, my treatment will be brief on details and will focus on placing his contributions into perspective from the standpoint of the development of the discipline. As I shall explain, Black played two roles, namely (1) that of pioneering his own important contributions to the literature and (2) (subsequently) that of rediscovering long-lost past treasures. Taken together, these roles justly earn for Duncan Black recognition as the founding father of public choice (Tullock, 1981; Grofman, 1981; Tullock, 1991; Rowley, 1991).

The Median Voter Theorem

Duncan Black honed his analytical economic tools at Dundee over the period 1932 to 1934, benefiting greatly from discussions with Ronald Coase, who at that time was framing his paper on the nature of the firm (Coase, 1937). In 1934, after his move to Bangor, Black began to apply economic concepts to the analysis of political systems. Unfortunately progress was not sufficiently swift and, reluctantly, Black felt obliged to set the project aside in late 1935 in order to secure his position at Bangor by publishing in the more conventional field of public finance (Coase, 1981, p. 5).

Black focused particular attention on Italian writings in public finance, which, at that time, were much more closely related to political theory than was the case in the United Kingdom and the United States. By September 1939, he had published several articles and a book (Black, 1939), thereby securing his academic position. At that moment in time, however, the end of the 'phony war' with Germany and its Axis supporters led Black to take a leave of absence from Bangor in order to serve the British wartime government as a member of its civil service.

Black did not abandon his pure politics project during his time in government. He proved the median voter theorem in February 1942 in extremely unusual circumstances, as his own typically modest testimony explains:

> I was "fire-watching" in case of air raids, around midnight in the green drawing room of Warwick Castle, one of the most stately rooms in the whole of England, though now there was a strange contrast between the coats of medieval armour and the walls and the long narrow tables strewn about the room and cluttered with civil service paraphernalia. Acting apparently at random, I took a sheet of civil-service notepaper and wrote down a diagram bearing three curves, and I saw in a shock of recognition that if I interpreted points on the horizontal axis as motions before the committee, and took the preferences of the members in relation to these motions to be represented by the three single-peaked curves, the decision by a committee using a majority-based procedure must correspond to the median option (Black, 1991, p. 262).

Duncan Black (Black, 1948a) proved that the median voter theorem holds for committee decision making by majority vote, in single-dimensional space, whenever the preferences of committee members are single-peaked and whenever the number of committee members is odd. In such circumstances, there exists a unique and stable alternative capable of receiving a majority vote in pairwise competition against all other alternatives. In such circumstances, the preference of the median voter is decisive (see Chapter 12 for further details). This result has exercised enormous influence on public choice scholarship and forms the basis for a great deal of highly-regarded theoretical and empirical analysis (McLean, MacMillan, and Monroe, 1996).

Although the median voter theorem is frequently assigned to Anthony Downs (Downs, 1957) by public choice scholars, this is completely incorrect (see Chapter 12). Downs makes no reference to the concept of the median anywhere in his 1957 book and, unfortunately, appears at that time to have been totally unaware of Black's earlier, path-breaking contributions.

Majority Vote Cycles in Single Dimensional Space

A short while after his identification of the median voter theorem, Black returned to Bangor for a brief sojourn from the civil service, in order to teach the classes of a colleague who was ill. By now, he was completely enthused with his discovery, and confident that he had discovered that unique and stable political equilibrium that would establish a pure theory of politics comparable to the general equilibrium theory of economics. To his utter dismay, in August 1946, Black stumbled upon

the conditions that underpin vote cycles and quickly recognized that, with vote cycles, there is no majority winner and no median voter theorem (for details see Chapter 12):

> I took it for granted that with a simple majority in use, the answer, irrespective of the shapes of the preference curves, would be determinate. Later, working out an arithmetical example in which an intransitivity arose, it seemed to me that this must be due to a mistake in the arithmetic. On finding that the arithmetic was correct and the intransitivity persisted, my stomach revolted in something akin to physical sickness. Not only was the problem to which I had addressed myself more complicated than I had supposed, it was of a different kind. The result would be determinate for only certain shapes of preference curves (Black 1991, p. 262).

Following the Allied victory in World War II, Black dedicated his immediate scholarship to researching the nature of spatial equilibrium and disequilibrium under the majority vote rule (see Chapter 12). The bulk of this scholarship, focused on determining the frequency and consequences of majority vote cycles and other majority voting problems, was published between 1948 and 1950 (Black, 1948a, 1948b, 1948c, 1949a, 1949b, 1950). These early contributions had a major impact on the development of public choice. As Grofman notes, "[i]n the first 32 issues of *Public Choice*, Black's work is cited in 20 issues, a total of 31 times. Only Downs (1957) and Buchanan and Tullock (1962) are more frequently cited. Indeed, the *Theory of Committees and Elections* (Black, 1958) is one of the most cited works in modern political science, being cited dozens of times each year in the published economics and political science literature" (Grofman, 1981, p. 13).

Majority-Vote Cycles in Multidimensional Space

The notion of viewing choices in spatial terms was not original to Duncan Black. However, his antecedents in the economics literature, Harold Hotelling (1929) and Arthur Smithies (1941) focused attention on oligopolistic competition between firms in economic markets (Grofman, 1981, p. 36). The first full-blown political application of the spatial model, in which alternative social states are viewed as points in a convex policy space (such as E^n) was by Black and Newing (1951) (see Chapter 3 for details).

Black and Newing's monograph, complex as it is, is a pioneering contribution to understanding the necessary and sufficient conditions for a stable voting equilibrium in multidimensional issue space. The authors demonstrate that majority voting equilibrium under such circumstances requires extremely restrictive assumptions concerning the nature of the intersection of voters' indifference contours. In this respect, Black and Newing (1951), whose monograph was submitted to *Econometrica* two years prior to the publication of Arrow's (1951) book, clearly anticipated Arrow's contribution, at least on the issue of intransitivity problems in multidimensional issue space (for details see Chapter 3). Moreover, Black and Newing (1951) pioneered the concept of *win-set analysis*, a concept that is now absolutely central

to a great deal of public choice scholarship related to perceived voting stability within legislatures and their committees in environments where underlying cycles typically are abundant (Tullock, 1981).

The Rediscovery of Long Lost Treasures

Much of Duncan Black's career, working without support at an interface between economics and political science that neither profession cared (at least openly) to acknowledge, was lonely and introspective. Perhaps these experiences of indifference and rejection led Black to search back into the history of thought for any signs of earlier recognition of his apparently novel and disturbing insights. In the event, whatever the cause, Black's voyage into the past proved to be fruitful, well beyond his original expectations.

During the late 1940s, Black became aware that two eighteenth century French noblemen—Jean-Charles, *Compte de Borda* and Marie Jean Antoine Nicolas Caritat, *Marquis de Condorcet*—had written papers on the theory of elections, papers that had long ago disappeared from the public domain. Intrigued by a few slim leads, Black visited the British Museum in August 1948 where he discovered the link between Borda, Condorcet, and the Academy of Sciences in Paris. In 1949, he visited Paris and became acquainted with the writings of both scholars as well as those by a third important academician, Pierre-Simon, *Marquis de Laplace*. In 1951, he also located important nineteenth century contributions to election theory by Charles Dodgson (Lewis Carroll) while serving as a member of the Governing Body of Christ Church College, Oxford.

Black worked on these lost treasures throughout the 1950s, writing up his discoveries in his 1958 masterpiece, *The Theory of Committees and Elections*. Suffice it here to note that from Black's archival work the world rediscovered the Borda Count (the Method of Marks), the nature of vote cycling, the concepts of the *Condorcet winner* and the *Condorcet loser*, issues surrounding the nature of the so-called *jury problem*, and many other insights that continue to stimulate the public choice research program during the early years of the twenty-first century (for details see Chapter 12).

Conclusions

Duncan Black was a modest, shy, and retiring man, a lone pioneer, perhaps too easily brushed aside by more aggressive, more ambitious, and more fortunately located scholars, themselves surrounded by equally ambitious and equally aggressive acolytes. Fortunately, Black's contributions, unlike those of Condorcet, Borda, Laplace, and Dodgson, were not lost in the swirling maelstrom of historical events. They survive and thrive, as do the once-lost treasures, but only because of the ever-diligent and scrupulously honest scholarship of the *Founding Father of Public Choice*.

References

Arrow, K.J. (1951/1963). *Social Choice and Individual Values*. New York: Wiley.

Black, D. (1948a). 'On the Rationale of Group Decision-Making', *Journal of Political Economy* 56, 23–34.

Black, D. (1948b). 'The Decisions of a Committee Using a Specific Majority', *Econometrica* 16, 245–261.

Black, D. (1948c). 'The Elasticity of Committee Decisions with an Altering Size of Majority', *Econometrica* 16, 261–272.

Black, D. (1949a). 'The Theory of Elections in Single-Member Constituencies', *Canadian Journal of Economics and Political Science* 15 (2), 158–159.

Black, D. (1949b). 'Some Theoretical Schemes of Proportional Representation', *Canadian Journal of Economics and Political Science* 15 (3), 334–335.

Black, D. (1950). 'The Unity of Political and Economic Science', *Economic Journal* 60, 506–514.

Black, D. (1958). *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Black, D. (1991). 'Arrow's Work and the Normative Theory of Committees', *Journal of Theoretical Politics* 3, 259–276.

Black, D. and Newing, R.A. (1951). *Committee Decisions with Complementary Valuation*. Glasgow: William Hodge.

Coase, R.H. (1937). 'The Nature of the Firm', *Economica* 4, 386–405.

Coase, R.H. (1981). 'Duncan Black: A Biographical Sketch', in G. Tullock (ed.) *Toward a Science of Politics: Essays in Honor of Duncan Black*. Blacksburg: Center for Study of Public Choice, pp. 1–10.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Grofman, B. (1981). 'The Theory of Committees and Elections: The Legacy of Duncan Black', in G. Tullcok (ed.) *Toward a Science of Politics: Essays in Honor of Duncan Black*. Blacksburg: Center for Study of Public Choice.

Grofman, B. (1988). 'Black Duncan' in J. Eatwell, M. Milgate and P. Newman (eds.) *The New Palgrave: A Dictionary of Economics*, Vol I, London: Macmillan, pp. 250–251.

Hotelling, H. (1929). 'Stability in Competition', *Economic Journal* 39, 41–57.

McLean, I., McMillan, A. and Monroe, B.L. (1996). *A Mathematical Approach to Proportional Representation: Duncan Black on Lewis Carroll*. Boston: Kluwer Academic Publishers.

Rowley, C.K. (1991). 'Duncan Black: Pioneer and Discoverer of Public Choice', *Journal of Public Finance and Public Choice* 2, 83–88.

Smithies, A. (1941). 'Optimum Location in Spatial Competition', *Journal of Political Economy* 49, 423–439.

Tullock, G. (1981). Foreword' in G. Tullock (ed.) *Toward a Science of Politics*. Blacksburg: Center for Study of Public Choice.

Tullock, G. (1991). 'Duncan Black: The Founding Father', *Public Choice* 71 (3), 125–128.

Chapter 5
Kenneth J. Arrow (1921-)[1]

Bernard Grofman[2]

Introduction

Kenneth Arrow, a 1972 Nobel Laureate in Economic Sciences who has held faculty positions in economics at the Universities of Chicago, Harvard, and Stanford (with most of his career spent at the last institution), is one of the most important social scientists of the post-WWII era. While social choice theory is only one of his many research interests, and Arrow would still be a towering figure, famous for his work on general equilibrium models, models of risk and uncertainty, information economics, health economics, etc., even if he had never written *Social Choice and Individual Values* (1951, revised edition 1963), it is this seminal book on which this essay will focus.

Social Choice and Individual Values was Arrow's 1951 doctoral dissertation in economics at Columbia, and published that same year by the Cowles Foundation, then located at the University of Chicago. Reprinted and continuously in print for well over 50 years, and continuing to be among the most highly cited publications in economics, it has spawned a vast and still growing literature. Any scholar in constitutional political economy must come to grips with its key results and the mathematical structure of its arguments.

The Problem of Social Choice

The best known aspect of the book, probably the most famous result in mathematical social sciences, is what has come to be called "Arrow's Impossibility Theorem." Inspired by the then little known *paradox of cyclical majorities* (a phenomenon subject to periodic rediscovery, e.g., by Duncan Black (1958) and, prior to Black, by E.J. Nanson), Arrow offered a way to generalize this paradox,

---

showing in the process that, rather than merely being a kind of mathematical curiosity, intransitivity in social choice was instead, apparently, a fundamental problem at the heart of democratic theory. The impossibility theorem shows that no aggregation process satisfying certain desirable and seemingly noncontroversial requirements can yield transitivity of collective preferences analogous to transitivity of individual preferences, i.e., such that if A is preferred to B and B preferred to C, then A ought to be preferred to C as well.

While there are different ways to state the theorem's requirements for reasonable aggregation processes, one set includes the assumptions that individual preference rankings are themselves transitive and that they are ordinal in character, the requirement that collective rankings respond to the preferences of more than a single individual (*non-dictatorship*), the requirement that alternatives that become more highly ranked by individuals do not ever thereby become lowered in the social ranking (*positive responsiveness*), the requirement that changes in the collective relative rankings of any given pair of alternatives be affected only by changes that affect the relative rankings of that same pair of alternatives among one or more individual rankings (*independence of irrelevant alternatives*), and the stipulation that there are no restrictions on the feasible individual rankings that may be introduced into the aggregation process (*absence of domain restriction*).

Arrow's impossibility result has very wide ramifications. By showing that no "fully satisfactory under all circumstances" aggregation process can exist, Arrow tossed a monkey-wrench both into welfare economics (where Arrow's theorem is taken to be about social welfare orderings) and into democratic theory (where Arrow's theorem is taken to be about voting rules and constitutional choice). In political science, Arrow changed the way political scientists think of the process of voting by problematizing the choice of voting rule. Indeed, in political science, Arrow's impossibility theorem (usually abbreviated as "Arrow's Theorem" as if it were the only theorem he ever proved) is often taken to be an attack on the possibility of meaningful democratic decision-making.

The search for ways to escape Arrow's impossibility result has been a major intellectual enterprise for more than 50 years. The Theorem's highly counterintuitive and apparently incontrovertible negative conclusion about the possibility for the aggregation of individual preferences by any reasonable decision rule that avoids dictatorship (including majority rule) has led to a variety of responses.

Alternative Escape Routes

There are six basic avenues of "escape" that have been pursued:[3]

---

[3] In the discussion that follows, of necessity, various mathematical complexities will be sloughed over.

Reject the Relevance of the Arrowian Framework

James Buchanan's view has long been to insist that we are making a category mistake if we seek to impose rationality conditions based on individuals on a collectivity. Groups are not individuals and we are wrong to be bothered if they do not behave as such. Another attack on the basic Arrowian framework is based on the claims that judgments are fundamentally different from preferences (Grofman and Feld, 1986); that the processes of deliberative judgments do not give rise to Arrowian problems since the search for truth allows only one best (available) answer; and that deliberation operates to move collectivities toward that answer (Aaken, List, and Luetge, 2007). Another attempt to escape from the dilemma involves shifting from an ordinal to a cardinal framework, but here the cost is attempting to provide a plausible solution to the problem of interpersonal comparisons of utility.

Modify the Fundamental Framework of the Theorem vis-à-vis the Transitivity Requirement

Here, the most important approaches include replacing transitivity with *quasi-transitivity* or with the requirement that the social welfare ordering only be a *semi-order*. This line of escape has proved to be largely a dead end in that, at best, we get weaker results, such as results that avoid dictatorship, but only at the cost of escaping from impossibility by requiring rule by something like an oligarchy; or we end up permitting dictators as long as they are *representative dictators* (Tanguiane, 1991). Relatedly, allowing for some intransitivity as long as a unique social choice is made does not help much in that we often get a choice set for the "best" alternative that is huge.

Eliminate or Modify Other Axiomatic Requirements of the Theorem

Here the most important proposed modification is with respect to domain restrictions. Arrow's "Possibility Theorem for Single-Peaked Preferences," the lesser-known accompaniment to his impossibility theorem, is built on Black's classic work on single-peakedness (Black, 1948), and inspired a search for plausible ways to limit preference orderings in such a fashion that the impossibility result does not hold. The definitive results for majority rule processes over a finite set of alternatives are due to Sen (1970a, b) in terms of a condition called *value restriction*, which encompasses the *NW* (*single-peakedness*) *condition*, the *NB* (*single-troughedness condition*), and the *NM* (*not middle condition*), the only one of the three conditions not to have a clear geometric representation. Sen shows that these three specify all the possible ways to guarantee transitivity on triples of alternatives in terms of limitations on the types of preferences that are admissible. Unfortunately, value restriction is a very strong

requirement, violated by virtually all empirically observed preference distributions. However, the Sen value restriction is in terms of conditions on profiles; and we can readily get transitivity even when value restriction is not satisfied (Feld and Grofman, 1986; Regenwetter et al., 2006). Moreover, if we restrict preferences to those that are generated by alternatives (and voter ideal points) that can be viewed as points in multidimensional issue space, a variety of solution concepts exist that would allow us to establish that the most socially desired alternatives are located within some particular relatively delimited area of the space (Feld et al., 1987).

After the "no domain restriction" requirement, the Arrowian requirement that is regarded as the most vulnerable to challenge is the criterion of "independence of irrelevant alternatives". Donald Saari (2001) has argued that this requirement is fundamentally misguided in that it throws away information and renders us unable to distinguish between transitive and intransitive individual preferences. Once we eliminate this requirement, aggregation, even if we allow all conceivable alternatives, methods exist that satisfy the remaining axioms, such as scoring rules like the *Borda rule*, of which Saari (2004, 2005) is an advocate.

Accept the Theorem, but Reject the Implications for the Infeasibility of Meaningful Collective Choice

This approach begins by emphasizing that, while for any aggregation process satisfying the theorem's assumptions there are always circumstances in which intransitivities will result, the theorem tells us nothing about the probability to be attached to the occurrence of such intransitivities. Those who take this approach customarily present evidence that the theorem is generally irrelevant in that, in most real-world settings involving a finite set of alternatives, for the set of observed individual preference orderings—typically highly asymmetric—there are social welfare functions that can be expected to yield transitive choice or voting rules that can be expected to yield a definitive outcome (Regenwetter et al. 2006). Alternatively, they assert that, even though cycles may occur, indeterminacies will largely or entirely be among alternatives that are virtually indistinguishable from one another (Feld and Grofman, 1992); or that the cycles that are present can be viewed as of little moment because collectivities do not even notice when they are there; or that we can cheerfully go on about our business even in the presence of cycles as long as outcomes are generated by rules that "legitimate" social choices (Grofman and Uhlaner, 1985).

Accept the Theorem, but Argue that Any Problematic Features of Social Choice It Identifies are to be Found in an Analogous Form When We Look at Individual Choices

This approach recognizes that individual preference orderings must aggregate/ reconcile multiple criteria in a fashion that parallels the way in which collective

preference orderings must aggregate/reconcile the preferences of multiple actors. The approach implies that there are no realistic institutional alternatives to the Arrow problem.

Go about Your Business Just as if the Theorem did not Exist

This "ostrich with its head in the sand" mode of response has been most common among political scientists who write about democratic theory and representation.

Other Implications of Arrow's Contribution

Even beyond the specific results of the "Impossibility Theorem" and the "Possibility Theorem for Single-Peaked Preferences", *Social Choice and Individual Values* has other important implications. Its proposed replacement of a cardinal utility framework for social welfare with an ordinal framework based in binary relations was a key part of an ordinalist revolution in welfare economics; and its demonstration of the power of an axiomatic approach led to an explosion in the application of such methods not just to voting/social welfare functions but to many other topics in economics and political science (for social choice applications see Kelley, 1988). Furthermore, later work would show hidden and deep mathematical connections between Arrow's work on potential intransitivity of social welfare functions and three other topics highly relevant to constitutional political economy: the manipulability of social choice (in particular, the Gibbard–Sattherthwaite Theorem: (Gibbard, 1973; Sattherthwaite, 1974, Craven 1992)); the infeasibility of a rights-based approach to liberal democracy (Sen, 1970a, b); and the feasibility of decentralized institutional decision-making (Saari and Sieberg, 2001).

References

Aaken, A., Christian, L., and Christoph, L. (eds.) (2007). *Deliberation and Decision: Economics, Constitutional Theory and Deliberative Democracy*. London: Ashgate.

Black, D. (1958). *The Theory of Committees and Elections*. Cambridge University Press: Cambridge.

Craven, J. (1992). *Social Choice: A Framework for Collective Decisions and Individual Judgements*. Cambridge: Cambridge University Press.

Feld, S.L. and Grofman, B. (1986). "Partial single-peakedness: An extension and clarification". *Public Choice*, 51, 71–80.

Feld, S.L. and Grofman, B. (1992). "Who's afraid of the big bad cycle?". *Journal of Theoretical Politics*, 4(2), 231–237.

Feld, S.L., Grofman, B., Hartley, M., Kilgour, M.O. and Miller, N. (1987). "The uncovered set in spatial voting games". *Theory and Decision*, 23, 129–156.

Gibbard, A. (1973). "Manipulation of voting schemes: A general result". *Econometrica*, 41(4), 587–601.

Grofman, B. and Feld, S. L. (1986). "Rousseau's general will: A condorcetian perspective". *American Political Science Review*, 82(2), 567–576.

Grofman, B. and Uhlaner, C. (1985). "Metapreferences and reasons for stability in social choice: Thoughts on broadening and clarifying the debate". *Theory and Decision*, 19, 31–50.

Regenwetter, M., Grofman, B., Marley, A.A.J., and Tsetlin, I. (2006). *Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications*. New York: Cambridge University Press.

Saari, D.G. (2001). *Decisions and Elections; Explaining the Unexpected*. Cambridge: Cambridge University Press.

Saari, D. G. and Sieberg, K. (2001). "The sum of the parts can violate the whole". *The American Political Science Review*, 95(2), 415–433.

Sen, A. (1970a). *Collective Choice and Social Welfare*. San Francisco: Holden-Day.

Sen, A. (1970b). "The impossibility of a Paretian liberal". *Journal of Political Economy*, 78(1), 152–157.

Tanguiane, A.S. (1991). *Aggregation and Representation of Preferences: Introduction to the Mathematical Theory of Democracy*. Berlin Heidelberg New York: Springer.

Chapter 6
Anthony Downs (1930-)[1]

Bernard Grofman[2]

Introduction

Anthony Downs' *An Economic Theory of Democracy* (Downs, 1957) is one of the founding books of the Public Choice movement, and one of the most influential social science books of the twentieth century. It has been reprinted and continuously in print (with the exception of a handful of months) for 50 years. Citations to it now surpass that of any other book on American electoral politics, including such classics as Campbell et al.'s *The American Voter* (1960), which laid the foundations for the behavioral approach to the study of voter behavior that has come to be associated with the University of Michigan (Gray and Grofman, 2005). Downs' book introduced seminal ideas, such as a cost-benefit calculation of political participation, a spatial model of party competition, knowledge about public affairs as a by-product of other more directly instrumental activities, and concepts such as rational ignorance and cue-taking behavior. While Downs has written more than a dozen other books since 1957, the sum of all this work has not come close to the influence of his first book, and our discussion will focus on this work.

An Economic Theory of Democracy was Downs' doctoral dissertation in economics at Stanford, and was published without any changes. Several political scientists who read it, including scholars such as Robert Dahl, recognized its innovation and importance, and after being shown the work by Dahl, Charles Lindblom, whom Downs had never met, arranged for three different publishers to offer Downs contracts for its publication (Downs, 1993). Yet, at the time of its publication no one could have truly anticipated its long-run impact on the political science discipline. Remarkably, too, it is largely an accident that Downs wrote on the topic he did, rather than on the political economy of expressways (Downs, 1993). Downs was encouraged to switch dissertation topics by scholars such as

---

[1] This is an original paper written specifically for this book.

[2] Professor of Political Science and Social Psychology and Adjunct Professor of Economics at the University of California at Irvine, 3151 Social Science Plaza, Irvine, California 92697. e-mail: bgrofman@uci.edu.

Julius Margolis, who introduced Downs to the writings of Joseph Schumpeter, and by his thesis supervisor, Kenneth Arrow (Downs, 1993).

The Paradox of Voting

Like most classic works in the social sciences, *An Economic Theory of Democracy* tends to be more cited than read (or reread), and its main ideas have entered the social sciences (especially political science) in only bare bones form. In particular, among many political scientists, the book has come to stand primarily for the simple (and simplistic) idea that rational voters should not bother to vote since the expected direct benefits, in terms of the probability that their vote will change the outcome multiplied by the benefit they gain in seeing one candidate win as opposed to another, is almost certainly less than the transaction costs involved in voting for the simple reason that the likelihood that one's vote will be decisive can be expected to be minuscule (Owen and Grofman, 1988).

Downs has changed the way political scientists think of the act of voting by problematizing that decision and making it a puzzle to be explained. Before Downs, it was essentially taken for granted that if you preferred a candidate you would vote for that candidate. Now, following Downs, it seems necessary to appeal to noninstrumental reasons for voting. One such factor is what has come to be called D, citizen duty. Yet, since time at the polls is time away from other (more profitable or more pleasurable) activities, and it requires time and effort to collect the information necessary to make a reasoned choice among available alternatives, an opportunity-cost perspective on voting suggests that few voters should bother. The fact that there is often a substantial level of political participation has led some political scientists to suggest, albeit partly tongue-in-cheek, that turnout is the "paradox that ate rational choice theory" (Morris Fiorina, quoted in Grofman, 1993), and it has led others to completely reject the rational choice perspective as something like "nonsense on stilts" (Green and Shapiro, 1994).

Downs' model of the turnout decision as one to be made in cost-benefit terms directly parallels other seminal public choice work on the provision of collective good and the avoidance of public bad, since supporters of a given candidate (or a given point of view) wish to see their candidate elected (or their point of view prevail) but would prefer to bear no cost to obtain this outcome. The large literature seeking to explain why people bother to vote thus parallels the even larger literature on "free riding." Ways to explain turnout within a generally rational choice perspective include ideas of expressive voting (Brennan and Hamlin, 1998; Glazer, Grofman, and Owen, 1998), approaches linking voting to group solidarity (Uhlaner, 1989; Morton, 1991), and work that emphasizes a comparative statistics approach to turnout whose concern is to explain variation in turnout *across* elections in terms of the factors that Downs used to explain the individual decision to vote or not to vote, such as expected electoral closeness, magnitude of candidate differences, and costs of electoral participation, along with size of the electorate, and the relative importance of the election (Hanks and Grofman, 1998; Grant, 1998).

The Median Voter Theorem

The other result most often associated with the name of Downs is the "median voter theorem." This theorem (also sometime called the Black–Downs theorem, in acknowledgement of Duncan Black's work on the power of the median voter when preferences are single-peaked (see Rowley, Chapters 1 and 3, this volume)), asserts that in a two-party competition along a single ideological or policy dimension, the candidates of each party will converge toward the location of the median voter in the electorate, i.e., we expect tweedledum-tweedledee politics. In the Downsian adaptation of the Hotelling–Smithies spatial model of economic competition, in which shops seek to locate where they can maximize their flow of customers, politicians compete to locate where (in policy terms) they can attract the most votes.

The median voter theorem has led to a huge literature on spatial models of politics. Since this topic is treated in detail in Hinich and Munger (Chapter 18, this volume) our comments will be brief. First, it would be a mistake to identify the median voter theorem as the Downsian model of party competition, since Downs offers other important but much neglected insights into party competition, e.g., on what happens with multidimensional politics and single-issue voters. Second, the fact that complete party convergence is not found in the United States (and in other countries using plurality-based elections) does not invalidate the insights that Downs had into the centripetal pressures on party platforms generated by different voting methods. The median voter theorem rests on a large number of quite specific assumptions whose violations often lead to non-convergence.

The Role of Information

A third important contribution of *An Economic Theory of Democracy* is on the role of information. Indeed, Downs deserves to be regarded as a founding figure in information economics.

First and foremost, Downs offers what might be called a model of "rational ignorance." Just as Downs' approach to voter turnout emphasizes the need to take into account whether one's vote can be expected to make a difference in the election outcome, Downs' approach to political information emphasizes the need to take into account whether new information can be expected to make a difference in the choice we make about which candidate/party to support or about whether to vote. Only if new information can be expected to improve our decision making so as to give us a higher expected payoff than before, and only if that expected gain exceeds the costs of pursuing the new information, will it be instrumentally rational to seek additional political knowledge. Second, to counteract this expectation of rational ignorance, Downs points out that information useful to political choice may be gained at a relatively low cost as a "by-product" of other activities. For example,

while voters may not know the values (or recent time path) of aggregate indicators such as GDP, inflation, or unemployment rates, they can use the information that comes to them when they buy things, and from their conversations with friends and neighbors, to assess inflation, unemployment, and the overall state of the economy. Third, Downs highlights the "signaling" power of various types of informational cues, such as party labels, interest-group endorsements, and self-professed ideology. Such low-cost cues allow voters to infer what choices are in their best interest, i.e., a vote that matches the choice the voter would make if she were in possession of complete information about the options available to her.

Conclusion

While *An Economic Theory of Democracy* is Downs' most important contribution to social science, the corpus of his other work is also quite impressive. He has written extensively in areas of public policy such as housing policy, transportation policy, and urban development, and on the politics of bureaucracy. He has been a Senior Analyst at the RAND Corporation and on the faculty of the University of Chicago, and is currently a Senior Fellow of the Brookings Institution.

References

Adams, J., Merrill, S., and Grofman, B. (2005). *A Unified Theory of Party Competition: A Cross-National Analysis Integrating Spatial and Behavioral Factors*. New York: Cambridge University Press.

Brennan, G., and Hamlin, A. (1998). "Expressive voting and electoral equilibrium," *Public Choice*, 108, 295–312.

Downs, A. (l957). *An Economic Theory of Democracy*. New York: Harper and Row.

Downs, A. (l993). "The Origins of 'An Economic Theory of Democracy'." In Bernard, G. (Ed). *Information, Participation and Choice: An 'Economic Theory of Democracy' in Perspective*. Ann Arbor: University of Michigan Press, pp. 179–193.

Glazer, A., Grofman, B., and Owen, G. (1998). "A neo-Downsian model of group-oriented voting and racial backlash." *Public Choice*, 97, 23–34.

Grant, D. (1998). "Searching for the Downsian voter with a simple structural model." *Economics and Politics*, 10(2), 107–126.

Gray, M. M. and Grofman, B. (2005). "Vindicating Anthony downs." *Political Science and Politics*, 38(4), 737–740.

Green, D. P. and Shapiro, I. (1994). Pathologies of Rational Choice Theory: A Critique of Applications in Political Science. New Haven, CT: Yale University Press.

Grofman, B. (Ed.). (l993). "Is turnout the paradox that ate rational choice theory" In *Information, Participation and Choice: An 'Economic Theory of Democracy' in Perspective*. Ann Arbor: University of Michigan Press, 93–103.

Hanks, C. and Grofman, B. (1998). "Turnout in gubernatorial and senatorial primary and general elections in the South, 1922–90: A rational choice model of the effects of short-run and long-run electoral competition on turnout." *Public Choice*, 94, 407–421.

Morton, R. (1991). "Groups in rational turnout models," *American Journal of Political Science*, 35, 758–76.

Owen, G. and Grofman, B. (1984). "To vote or not to vote: The paradox of nonvoting." *Public Choice*, 42, 311–325.

Uhlaner, C. (1989). "Rational turnout: The neglected role of groups." *American Journal of Political Science*, 33(2), 390–422.

Chapter 7
James M. Buchanan (1919-)[1]

Robert D. Tollison

Introduction

James M. Buchanan was born on October 3, 1919, in Murfeesboro, Tennessee. He grew up on a farm in this area of the United States. His post-secondary school education consists of a B.A. degree from Middle Tennessee State University (1940), an M.A. degree in economics from the University of Tennessee (1941), and a Ph.D. degree in economics from the University of Chicago (1948). He served in the U.S. Navy in the Pacific during World War II, where he received a Bronze Star. He has taught at the following universities: Tennessee, Florida State, Virginia, University of California, Los Angeles, Virginia Polytechnic Institute (VPI), and George Mason University where he still works today. He also maintains an office at VPI. He has held endowed chairs in economics at Virginia, VPI, and George Mason University. He was Department Chair at Florida State. He was Department Chair and Director and cofounder of the Thomas Jefferson Center for Political Economy at Virginia. He was General Director of the Center for Study of Public Choice at both VPI and George Mason. He has also served as a visiting professor at the University of Miami, Brigham Young University, the London School of Economics, and Cambridge University. He spent a year in Italy as a Fulbright Research Scholar. He is a former President of the Mont Pelerin Society, the Western Economic Association, and the Southern Economic Association. He has been awarded many honorary doctorates, including those from the University of Catania and the New University of Lisbon. He is a Distinguished Fellow of the American Economic Association. He received the Alfred Nobel Memorial Prize in Economics in 1986, the last year the award was tax free in the United States.

My Plan

My approach to a Buchanan biography will be linear in nature. Basically, I shall follow him in a straight line, where divisions of time are marked by his university

---

[1] This chapter is a reprint of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, pp. 139–145.

affiliations. In this way I can review his work and perhaps say a few words about his contemporaneous colleagues and doctoral students. My focus, however, will be on his intellectual work at each school, beginning with graduate school at the University of Chicago and proceeding to his present residence at George Mason University. So rather than discussing Buchanan's ideas by category, I am going to trace their evolution over time at different work stations.

I am only going to hit the high points of Buchanan's contributions in these various locales. That he is a prolific scholar is well known. The interested reader may refer to the 20 volumes of his collected works published by Liberty Fund.[2]

Graduate School

I am fortunate in that Buchanan has written an autobiographical memoir (Buchanan 1992) that provides invaluable guidance to his view of the various stages of his career. I begin with his graduate student years at the University of Chicago.

Two features of his graduate student experience stand out. One was his introduction to Frank Knight, and the other was his discovery of the work of Knut Wicksell. From what I can gather, his attraction to Knight was based on Knight's personality and his general approach to intellectual affairs and not especially on Knight's economics. Basically, he was impressed by Knight as a person. Knight came from a rural background outside the establishment. Buchanan had similar roots. To Buchanan, Knight was a truth-seeker (but there were no truths to be sought), who cared not one whit for anything else. This was what Buchanan wanted to be; this is what he took from Knight; Knight is his role model.

Buchanan (1992) also credits Knight with his conversion from socialism by teaching him how markets work. The conversion apparently came about six weeks into a price theory course taught by Knight. I am not so sure about this recounting, mostly because it is hard to conceive of Buchanan as a budding socialist. In any event, Buchanan was not much of a socialist, if at all, although residues remain, such as his antipathy towards inherited wealth.

Buchanan's second formative experience at Chicago was his happenstance discovery of Wicksell's dissertation (Wicksell 1896). This is the famous work (part of which was later translated and published by Buchanan) that emphasized the use of more qualified (stricter) voting rules in defining the efficiency of public spending proposals. Wicksell's ideas were to play a significant role in shaping Buchanan's approach to political economy as it evolved over the upcoming years. Not only had Buchanan found a calling at Chicago (economics and scholarship), but he had found some useful guides about how and where to go, in Knight and Wicksell.

---

[2] See Brennan et al. (1999–present). Note that these volumes do not include work that Buchanan has produced since 1999, work which continues unto this day.

Early Academics and Italy

Buchanan began his academic career at the University of Tennessee in 1948. He moved to Florida State University in 1951, where he was a Full Professor and Department Head from 1954 to 1956. He spent an eventful year in Italy as a Fulbright Research Scholar, after which he moved to the University of Virginia in 1956.

Over this period he wrote and published two pieces in the *Journal of Political Economy* which presaged his later work in public choice (Buchanan 1954a, b). Both papers were written in response to Arrow's famous work on social welfare (Arrow 1951). One paper (1954b) is the original and classic statement of the differences in terms of individual choice behavior between voting and the market. Voting, for example, is a more "bundled" choice than market choices. The other paper (1954a) is a fundamental critique of Arrow's analysis. Buchanan makes a variety of points here, with perhaps the most important being that there is nothing special about majority rule as opposed to a unanimity rule, where the latter will yield consistent collective choices analogous to the way that markets work. Basically, he argued that outcomes and rules were related and that Arrow ignored this linkage in his analysis.

I do not have the time and space to review these papers in detail. The point is that they clearly were important early precursors of public choice analysis. Buchanan, in this early period, was already thinking deeply about voting processes and the implications of voting processes for economic well-being. These papers were the seed corn of the public choice revolution, and clearly contained echoes of Wicksell.

The year that Buchanan spent in Italy was intellectually fruitful. He was introduced to the Italian tradition in public finance, in which an individual choice perspective was employed and spending and taxes were linked and not treated separately. This methodological insight was later to fuel many of Buchanan's contributions to the theory of public finance.

He also had an epiphany about public debt theory, which led to his major work in this area (Buchanan 1958). The latter involved the individual choice approach to fiscal analysis, in which Buchanan clearly exposited how the burden of the debt was shifted to future taxpayers. We clearly did not simplistically owe the debt to ourselves. Though the Keynesians howled in protest, time has been kind to Buchanan's analysis, as it now seems to have strong currency among present day analysts and observers of public debt policy. Paying down the debt so as not to leave a burden on our children has virtually become a political mantra in some quarters.

Charlottesville

It is hard to call one period of Buchanan's academic life more productive than another, but the amount and quality of the work he did at the University of Virginia is simply amazing. Most of this work is so well known that I need only mention it in passing.[3]

---

[3] Note also that the previously discussed book on the public debt was published in 1958 while Buchanan was at Virginia.

It was over this period that Buchanan met Gordon Tullock, and Tullock joined the Economics Department in Charlottesville.[4] An intellectual association was thus formed that would produce seminal work and carry forward for many years into the future. The seminal work was, of course, *The Calculus of Consent*, published in 1962. This is one of three or four major works in early public choice that are rightly considered classics. The book was a tour de force, covering methodological issues, constitutional economics, analyses of voting rules, and many other topics that continue to occupy public choice scholars today.

What is so amazing about this period of Buchanan's life is that he also made lasting and fundamental contributions to public expenditure theory and to the theory of taxation. He wrote his famous papers on externalities (Buchanan and Stubblebine 1962), tax earmarking (Buchanan 1963), and clubs (Buchanan 1965), each of which heavily influenced the subsequent literature of public economics. Indeed, the clubs paper by itself has created an industry of further applications to such topics as alliances and fiscal federalism.

He wrote and published a major treatise on public finance (Buchanan 1967), in which he introduced an individual choice approach to public finance theory, as well as rehabilitating and extending such concepts as fiscal illusion. This is my favorite work by Buchanan, and it still merits re-reading today. As Buchanan shows time and again in this work, understanding the efficiency of taxation and spending programs requires analyzing both sides of the fiscal account at the same time.

He wrote and published a major book on public goods theory (Buchanan 1968). This book is deceptively technical, and is still the most creative work on public goods theory in the literature. It also treats the "supply" as well as the "demand" for public goods, an aspect of analysis that makes this book unique in the area of public goods theory.

Buchanan published his little book on subjective cost (Buchanan 1969) during this time. Here, we have a prime example of Buchanan's dalliance with Austrian ideas, a link that he personally cares about, but which really is not all that important in the general context of his work. Buchanan cannot be claimed by the Austrians; his work is much bigger than their narrow methodological hiding place. And while costs may be subjective, this has not stopped Buchanan from forging ahead as a creative economic theorist.

Finally, he made major contributions to the discussion of methodology in the 1960s. For some of this work, see Buchanan (1962–1964).

On top of all this, there were numerous other papers, lectures, and academic duties. The Public Choice Society was co-founded by Buchanan in 1963, and as noted earlier, he served as President of the Southern Economic Association in 1963. This was at a time when being President of the Southern actually meant something.

Moreover, many of Buchanan's best doctoral students studied and wrote their dissertations under his direction at Virginia. These include (in no special order)

---

[4] Other members of Virginia's Economics Department at this time were Warren Nutter, Leland Yeager, and Ronald Coase.

Matt Lindsay, Dick Wagner, Charlie Goetz, Charlie Plott, Mark Pauly, Toby Davis, and Craig Stubblebine, to mention a few.[5]

This is a good place to discuss Buchanan as a teacher. In the classroom he was at his remarkable best. He was a hard teacher, who set a good example for his students. His method was to assign short papers, due every two weeks, about whatever he was working on at the time. These papers and his classes made the students feel as if they were working on the frontiers of economics and participating in an exciting discussion of ideas. Grades were based on one's originality in approaching a topic, not on technique or the derivation of results. Creativity was the key to a good grade in Mr. Buchanan's class.

Oftentimes, these class papers led to later publications by students, which, of course, helped them immensely in their careers. The best example of this is Mark Pauly's paper on moral hazard (Pauly 1968). This was a very important contribution to economic theory, and it was written and published while Pauly was a graduate student at Virginia.

Buchanan's class was transforming for students. Typically, one entered the program at Virginia (as I did) to obtain a doctorate and return to a small liberal arts college to teach. The idea of being a research economist had never really occurred to many of these students.

Yet under the tutelage and encouragement of Buchanan, they got their degrees at Virginia and headed off to Harvard, Northwestern, Cornell, Iowa State, Purdue, Illinois, UCLA, Carnegie-Mellon, and other major universities to publish or perish. And almost to a person, these young economists have emerged in their own right as important scholars.

One significant aspect of these students is that they are all different, working in different areas and approaches to economics, some of which bear little resemblance to Buchanan's work. Buchanan did not produce homogeneous graduate students who all worked in his tradition. He produced a colorful array of creative people who found their own way in the world. They were able to do this because Buchanan did not beat them down as students and make them feel as if there was nothing they could do. He rather gave them encouragement and inspiration, showing them that they too could participate at a high level in the economics profession. This is the mark of a gifted teacher. Like a Zen Master, Buchanan gave visions and aspirations to his students that he did not possess himself.

In 1969, Buchanan left Virginia to take a position at the University of California, Los Angeles. After an uneventful intellectual year there, he joined Gordon Tullock and Charlie Goetz at VPI, where they had taken up residence previously. Tullock had left the University of Virginia earlier and gone to Rice University, but was lured to Blacksburg by Goetz, where, with Goetz, he helped to entice Buchanan back to the Commonwealth of Virginia. The story of why Buchanan and Tullock left Virginia revolved around that university's failure to promote Tullock to full professor. Virginia's loss was clearly VPI's gain.

---

[5] I too wrote under Buchanan at Virginia, finishing in 1969.

Blacksburg

Buchanan's return to Blacksburg was a happy one. There, he joined Tullock and Goetz to form the Center for Study of Public Choice, where he was to work productively for the next 14 years. He was also joined in Blacksburg by Mrs Betty Tillman, who had been his Executive Assistant at Virginia and whose role in the public choice movement would grow tremendously in Blacksburg and later in Fairfax. The Center was housed in Blacksburg in the old president's house, a large mansion atop a hill overlooking the Duck Pond. Center offices were palatial by normal academic standards.

This idyllic setting attracted an array of talented scholars to the Center, both as permanent faculty and as visitors, and to my mind this period represents the high water mark of the Center in terms of the level of work and quality of faculty there. Over this period, the faculty included people such as (in no special order) Dick Wagner, Tom Borcherding, Charlie Goetz, Winston Bush, Geoff Brennan, Mel Hinich, Bob Mackay, Art Denzau, Mark Crain, Roger Faith, Dwight Lee, and Nic Tideman, and, of course, Buchanan and Tullock.[6] Visitors to the Center were commonplace, and such notable scholars as Dennis Mueller, Charles Rowley, Fritz Schneider, Peter Bernholz, Dick McKenzie, Eddie West, and many others spent time in Blacksburg over this period. Numerous doctoral students completed their degrees at the Center at this time, and went on to careers as well-known scholars. These include (in no particular order) Randy Holcombe, Carolyn Weaver, Henry Butler, Dick McKenzie, Genia Toma, Mark Toma, David Laband, Roger Congleton, and Janet Landa. Laband, Congleton, and Landa wrote under Buchanan.

Buchanan was literally in charge. He generally opened the door in the morning, and closed it at night, putting in 10 to 12 hours a day in between, Saturdays and Sundays being only partial exceptions (6 hours). I would say also that most of the external financial support that came to the Center in Blacksburg (and later in Fairfax) was due to Buchanan and his presence in these locales. That Buchanan was unfailingly generous in supporting others' research efforts ought to be noted.

Buchanan's work over this period continued his earlier emphasis on issues of constitutional economics and public finance from a public choice perspective. In addition, there were side excursions to topics in which he was interested. Let me explain.

One of his major works over this period was *The Limits of Liberty* (Buchanan 1975a). The book is dedicated to Winston Bush, a colleague in Blacksburg, who died tragically in a local car accident. Bush had attracted Buchanan's interest to the issue of analyzing how individuals act in a setting of anarchy (no government) and how individuals make the leap to civil society with rules and laws. This, of course, is precisely the constitutional paradigm that Buchanan already knew so well, but Bush's approach opened up new vistas. In *Limits*, Buchanan offers the best

---

[6] I was Professor of Economics and Executive Director of the Center from 1976 to 1981 in Blacksburg.

statement of his intellectual position. The step to civil society contains risks (Leviathan), and it should not be approached without careful thought about how to do it. In particular, Buchanan stresses the criterion of agreement on rules as being the acid test of validation for the formation of governmental institutions. Hence, Buchanan emerges in this book not only as a major voice in constitutional economics, but in contractarian philosophy as well. Space does not permit me to do justice to this work; suffice it to say that it has had a major impact in both philosophy and economics.

On the public finance side of the street, Buchanan began a collaboration with Geoffrey Brennan in Blacksburg that proved to be fruitful and important. Indeed, Brennan would become Buchanan's most prolific collaborator, and a genuine colleague and friend in all ways. Building upon the foundation laid in *Limits*, these authors pioneered a new approach to public finance based on the idea that government could be expected to act like a Leviathan and to seek to maximize whatever advantage it was given when civil society emerged from anarchy. That is, the state would be a tax-revenue maximizer, a regulatory rent maximizer, and so on. Buchanan and Brennan (1980) traced out the novel implications of this approach for taxation, spending, and the size of government, and also explored how certain rules could be designed to constrain the tendencies of the Leviathan state. This work literally flipped the existing theory of public finance on its head. Instead of using economic analysis to show governments how to collect taxes more efficiently, Buchanan and Brennan used it to show how to guard against the potential for a bloated, tyrannical public sector.

Buchanan's other book over this period was written with Dick Wagner, and it represents what I have called "an excursion into an interesting side issue" (Buchanan and Wagner 1977). They use basic public choice analysis to explain why Keynesian economic principles are abused by self-interested politicians to run perennial budget deficits. This is not a technical book, but it is a very persuasive application of basic public choice theory. Buchanan's support for a balanced budget amendment to the U.S. Constitution grew out of this earlier critique of Keynesian economics.

I have, of course, only mentioned books so far. In addition, there are numerous major journal articles. This was the period in which Tullock's earlier work on rent seeking (Tullock 1967) was consolidated and extended (Buchanan et al., 1980). Buchanan played a major role in this effort. His paper (Buchanan 1980) "Rent Seeking and Profit Seeking" remains the clearest statement in the literature of the rent seeking idea. Other major papers over this period include the one on regulation (Buchanan and Tullock 1975), on the Samaritan's Dilemma (Buchanan 1975b), and on tax limits (Buchanan and Brennan 1977b), to name only three.

Buchanan also became very interested in the work of John Rawls during the Blacksburg era, and he wrote several papers (Buchanan 1972), in which he drew parallels between the Rawlsian approach to deriving a social contract (minimax) and his own approach to constitutional choice (expected utility maximization under uncertainty about future position). In my view, the interest of Buchanan in Rawls was another excursion into a side issue, where Buchanan was looking for individuals

who shared his general interest in the problems of constitutional choice at least in a broad sense. Today, I would say that Buchanan's position on constitutional economics is purely Buchanan's, and bears little or no resemblance to that of Rawls. Indeed, I see virtually no imprint of the Rawlsian interlude in Buchanan's work.

Alas, paradise was lost. An academic civil war erupted in Blacksburg, quite unexpectedly, and after all was said and done, Buchanan actually won the war. But fearing that too much capital had been burned up in the process, Buchanan and his Center colleagues accepted an offer to move *en masse* (at given pay and rank) to George Mason University in Fairfax, Virginia. George Mason's gain was VPI's loss.

Buchanan, however, only moved his professional address to Fairfax. In Blacksburg, he had returned to a rural lifestyle with relish, and set down his roots deep in the Appalachian Mountains, where he grew his own vegetables and chopped his own wood. This is still his main residence, as he commutes back and forth to Fairfax, and he shared this dominion with his wife, Ann, and a host of cats and dogs until her death in 2005.

A final note about Blacksburg is that it was a very social place. People worked hard, but they played hard too. Jim and Ann Buchanan were at the center of this society. To be asked over to dinner by Jim meant that there was good eating and good conversation in your future. In addition, there were poker games to be played, blackberries to be picked, Super Bowls to be watched, and foozball games after work. Needless to say, Buchanan did not play foozball, but otherwise he was the center of a unique and lively little universe. Before the war, Blacksburg was fun.

Fairfax

Buchanan did not move to Fairfax because he wanted to advise the government. He moved there because he found the academic environment there congenial. Moreover, this time, George Mason's gain really was VPI's loss. Barely over two years in residence at Mason, Buchanan was awarded the Nobel Prize in Economics (1986). The Prize changes most people, but I do not think that Buchanan changed very much at all after 1986. He still worked long hours, he was still interested primarily in ideas, he remained(s) productive, and he continued to stay on course. I mostly remember the incredible surge of pride that swept through the Center and through the hundreds of friends and colleagues of Buchanan and the Center on the day of the Nobel announcement. A ragtag band of public choicers basked in the reflected light.

As I said, the Mason environment was congenial. Mason is a former community college in a suburban setting. The Center was given facilities in an old Methodist church in a copse of woods on the edge of the campus. I forget who occupied the preacher's old office.

There has been a great deal of turnover of Center faculty at Mason, but over the last 19 years, the faculty has included (in no special order) Dick Wagner, Charles

Rowley, Gordon Tullock, Roger Congleton, Mark Crain, Tyler Cowen, David Levy, Ron Heiner, Geoff Brennan, Dwight Lee, Bill Shughart, Victor Vanberg, and, of course, Buchanan.[7] The Center has also educated a slew of doctoral students over this era, far more than at any other locale. Among these students are (in no special order) Gary Anderson, Pam Brown, Brian Goff, Don Leavens, Joe McGarrity, and many, many others. Buchanan directed only two doctoral dissertations at George Mason (Frank Forman and Nimai Mehta).

After Buchanan's Nobel, the university allowed the Center to rehabilitate (at its own expense) an old house across the street from the main Center building. This is now the Buchanan House, which houses Buchanan, Mrs Tillman, and Mrs Jo Ann Burgess, the Librarian of the Buchanan House. Many Buchanan artifacts are displayed in the Buchanan House, including a replica of his Nobel medal.

Buchanan's work over the Fairfax period has continued unabated. In 1983, he and Brennan published a follow-on study to *The Power to Tax*, ingeniously called *The Reason of Rules* (Buchanan and Brennan 1985). This work is a treatise on constitutional economics that seeks to make the case for an "economics of rules" as opposed to an "economics of politics." The book stepped out of the normal box in which economics operates, and asked the question, how should we go about selecting the rules of play in the box? This is, of course, the life-long question that has held Buchanan's interest.

Buchanan published one other book over this period, a work with Roger Congleton, *Politics by Principle, not Interest* (Buchanan and Congleton 1998). This book expresses in modern analytical terms many of the ideas that Buchanan was writing about earlier. In particular, Buchanan and Congleton show how general rules of taxation, for example, increase the efficiency and productivity of government. Simply put, flat taxes may be better than progressive taxes because they reduce rent seeking and tax evasion in a post-constitutional society. This work has received several nice reviews.

Buchanan's other intellectual work at Mason has been extensive. He has issued several important collections of his papers (Buchanan 1991). He has consolidated and extended his intellectual position (Buchanan 1990). He has explored new areas of economic theory (Buchanan 1994). Moreover, he is still hard at work, pushing well beyond the 20 volumes of his *Collected Works*.

Mason was not as social as VPI had been. The urban setting raised the costs of socializing. Everyone seemed to go their own way. Buchanan instituted and funded a Virginia Political Economy Lecture Series, which served as a social occasion each March. Speakers have included many of the people mentioned in this paper.

This pretty much exhausts the Buchanan timeline. We are up to date. Buchanan is still in residence at George Mason, but he also keeps an office at VPI. They have also named a Center in his honor at George Mason, called the James M. Buchanan Center for Political Economy.

---

[7] I was Director of the Center at George Mason from 1984 to1998; I also held the Duncan Black Chair in Economics over most of that period.

Essences

The science of essences concerns those things that define the elements of what makes something smell or taste so good. What is the essence, then, of Buchanan? I shall list these by way of closing this account of his life to date. My methodology is that of the pointillists (Georges Seurat), hoping to achieve a general impression from a series of interconnected dabs of paint.

- Buchanan changed the subject matter of modern economics by stressing agreement on the rules of the game as a separate and important inquiry in its own right.
- Buchanan exposed the vacuous nature of modern welfare economics by stressing agreement and not an arbitrary social welfare function as the key to the validity of institutional choices.
- Buchanan led the way in showing scholars how to analyze political processes using the methodology of economics.
- Buchanan pioneered in bringing individual choice analysis back into public finance theory.
- Buchanan refocused economics in methodological terms on those areas (individual choice behavior) where it has the greatest value.
- Buchanan has made many contributions to positive economic analysis.
- Buchanan is primarily a normative theorist.
- Buchanan is a great teacher, who trained many good students.
- Buchanan is a good colleague, reading and commenting on thousands of papers by colleagues.
- Buchanan created and largely financed an intellectual network at three universities in Virginia.
- Buchanan's contractarianism has had a major impact on philosophy.
- Buchanan is one of the most cited scholars of his generation.
- Buchanan does not suffer fools gladly.
- Buchanan is honest, and does not hesitate to state his mind, sometimes hotly.
- Buchanan has a loyal network of friends and colleagues in this country and abroad.
- Buchanan is a prolific lecturer, having given thousands of invited lectures, seminars, and talks, in a variety of venues, from Rotary Clubs to the great universities of the world (as well as the not-so-great ones).
- Buchanan is an incessant traveler, especially to Europe.
- Buchanan is an avid reader of both fiction and non-fiction.
- Buchanan grows his own food and chops his own wood.
- Buchanan is an active correspondent, having produced thousands of pages of accumulated correspondence.
- Buchanan does not like to talk on the telephone.
- Buchanan types his own work from handwritten notes, either on an old typewriter or, more recently, on a computer.
- Buchanan's memos are on yellow onionskin paper.

- Buchanan is a regular attendee and participant in professional meetings and conferences (especially Liberty Fund conferences).
- Buchanan is one of the best writers in the economics profession.
- Buchanan is virtually a whole university by himself as well as an effective academic infighter.
- Buchanan is a hard coauthor to keep up with; he has a paper drafted and back to you before you have time to take a deep breath.
- Buchanan is a good friend to animals, especially dogs and cats.
- Buchanan is an armchair theorist with an aversion to econometrics.
- Buchanan is a social man, who loves a good joke and a good conversation.

If this reminds you of Buchanan just a little bit, my methodology has worked, and I can draw this essay to a close. They say that only poets and singers achieve immortality. But surely the work of great economists lasts long enough and reverberates across time in such a way that they are practically immortal. Anyway, what is the difference between a half-life of 250 years and immortality? Buchanan has reached this level. And it is a good guess that his ideas will grow in importance over time as young scholars reshape modern social science along the lines that he has laid out.

References

Arrow, K. J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Brennan, G., Kliemt, H., and Tollison, R. (eds.) (1999–Present). *The Collected Works of James M. Buchanan*. Indianapolis: Liberty Fund.

Buchanan, J. M. (1954a). "Social choice, democracy, and free markets." *Journal of Political Economy*, 62, 114–123.

Buchanan, J. M. (1954b). "Individual choice in voting and the market." *Journal of Political Economy*, 62, 334–363.

Buchanan, J. M. (1958). *Public Principle of Public Debt*. Homewood: Irwin.

Buchanan, J. M. (1962). "The relevance of Pareto optimality." *Journal of Conflict Resolution*, 6, 341–354.

Buchanan, J. M. (1963). "Earmarking vs. general-fund financing." *Journal of Political Economy*, 71, 457–469.

Buchanan, J. M. (1964). "What should economists do?" *Southern Economic Journal*, 30, 213–222.

Buchanan, J. M. (1965). "An economic theory of clubs." *Economica*, 32, 1–14.

Buchanan, J. M. (1967). *Public Finance in Democratic Process*. Chapel Hill: North Carolina Press.

Buchanan, J. M. (1968). *The Demand and Supply of Public Goods*. Chicago: Rand McNally.

Buchanan, J. M. (1969). *Cost and Choice*. Chicago: Markham.

Buchanan, J. M. (1972). "Rawls on justice as fairness." *Public Choice*, 13(Fall), 123–128.

Buchanan, J. M. (1975a). *The Limits of Liberty*. Chicago: Chicago Press.

Buchanan, J. M. (1975b). "The Samaritan's dilemma," in Phelps, E. S. (ed.) *Altruism, Morality, and Economic Theory*. New York: Sage, pp. 71–85.

Buchanan, J. M. (1980). "Rent seeking and profit seeking," in Buchanan, J. M., Tollison, R. D., and Tullock, G. (eds.) *Toward a Theory of the Rent-Seeking Society*. College Station: Texas A&M Press, pp. 3–15.

Buchanan, J. M. (1990). "The domain of constitutional economics." *Constitutional Political Economy*, 1, 1–18.

Buchanan, J. M. (1991). *The Economic and Ethics of Constitutional Order*. Ann Arbor: Michigan Press.

Buchanan, J. M. (1992). *Better than Plowing*. Chicago: University of Chicago Press.

Buchanan, J. M. (1994). "The economic, and the ethics of idleness." *In Ethics and Economic Progress*. Norman: Oklahoma Press, pp. 112–128.

Buchanan, J. M. and Brennan, H. G. (1977). "The logic of tax limits: Alternative constitutional constraints on the power to tax." *National Tax Journal*, 32, 11–22.

Buchanan, J. M. and Brennan, H. G. (1980). *The Power to Tax*. Cambridge: Cambridge Press.

Buchanan, J. M. and Brennan, H. G. (1985). *The Reason of Rules*. Cambridge: Cambridge Press.

Buchanan, J. M. and Congleton, R. C. (1998). *Politics by Principle, Not Interest*. Cambridge: Cambridge Press.

Buchanan, J. M. and Stubblebine, C. (1962). "Externality." *Economica*, 29, 371–384.

Buchanan, J. M., Tollison, R. D., and Tullock, G. (eds.) (1980). *Toward a Theory of the Rent-Seeking Society*. College Station: Texas A&M Press.

Buchanan, J. M. and Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor: Michigan Press.

Buchanan, J. M. and Tullock, G. (1975). "Polluters' profits and political response: direct controls vs. taxes." *American Economic Review*, 65, 139–147.

Buchanan, J. M. and Wagner, R.E. (1977). *Democracy in Deficit*. New York: Academic Press.

Pauly, M. (1968). "The economics of moral hazard." *American Economic Review*, 58, 531–537.

Tullock, G. (1967). "The welfare costs of tariffs, monopolies, and theft." *Western Economic Journal*, 5, 224–232.

Wicksell, K. (1896). *Finanztheoretische Untersuchurgen*. Jena: Gustav Fischer.

Chapter 8
Gordon Tullock (1922-)[1]

Charles K. Rowley[2]

Biographical Details

Gordon Tullock was born in Rockford, Illinois on February 16, 1922. His father, George was a hardy Midwesterner of Scottish ancestry, his mother, Helen, was of equally hardy Pennsylvania Dutch stock. He obtained his basic education in the public schools of that city, displaying from early childhood a superior intellectual ability that clearly distinguished him from his peers. In 1940, Tullock departed for the School of Law at the University of Chicago to combine a two-year program of undergraduate courses with a four-year formal law program. In fact, he completed the initial two-year program in a single year.

His law school program was interrupted by his being drafted into military service as an infantry rifleman in 1943, but not before he had all but completed a one-semester course in economics taught by Henry Simons. This course was to be Tullock's only formal exposure to economics, a fact that no doubt enhanced rather than hindered his future success in contributing highly original ideas to that discipline.

Tullock served in the U.S. military until shortly after the end of hostilities, returning to civilian life in December 1945. He took part in the Normandy landings on D-Day + 7 as a member of the Ninth Infantry. His life almost certainly was spared by the good fortune of his being left behind at division headquarters to defend three anti-tank guns. The original members of the Ninth Infantry were decimated on their hard-fought route across France and into Germany.

Following behind, Tullock eventually would cross the Rhine, he claims, while still asleep. Ultimately, he would end up in the Russian sector. Although Tullock modestly dismisses his wartime service as uneventful, this can only be with the

---

advantage of hindsight and considerable modesty. Participation in a major land war as part of "the poor bloody infantry" is never without the gravest of risks.

Following this three-year wartime interruption, Tullock returned to Chicago and obtained a *Juris Doctor* degree from the Chicago Law School in 1947. He failed to remit the $5 payment required by the University and therefore never received a baccalaureate degree.

His initial career, as an attorney with a small but prestigious downtown Chicago law firm, was controversial and, perhaps, mercifully brief. During his five-month tenure, Tullock handled two cases. The first case he won when he was expected to lose, and only after one of the partners in his firm had advised his client not to pursue the matter. The second case he lost when he should have won and he was admonished by the court for his poor performance (Brady and Tollison 1991, 1994, p. 2). Fortunately for the world of ideas, these events persuaded him to seek out an alternative career.

Prior to graduation, Tullock had passed the notoriously difficult Foreign Service Examination. He joined the Foreign Service in fall 1947 and received an assignment as vice consul in Tientsin, China. This two-year assignment included the Communist takeover in 1948. Following Tullock's return to the United States, the Department of State dispatched him to Yale University (1949–1951) and then to Cornell University (1951–1952) for advanced study of the Chinese language.

In late 1952, he joined the "Mainland China" section of the Consulate General in Hong Kong. Some nine months later he was reassigned to the political section of the U.S. Embassy in Korea, where, once again, he would briefly be overrun by communist insurgents. Tullock returned to the United States in January 1955, where he was assigned to the State Department's Office of Intelligence and Research in Washington. He resigned from the Foreign Service in fall 1956.

Over the next two years, Tullock held several positions, including most notably that of Research Director of the Princeton Panel, a small subsidiary of the Gallup organization in Princeton. Essentially, he was in transition, marking time until he was ready to make a bid for entry into academia.

Unusually, Tullock had already published in leading economics journals articles on hyperinflation and monetary cycles in China and on the Korean monetary and fiscal system even during his diplomatic service, thereby whetting his own appetite for an academic career and signaling an unusual facility for observing his environment as the basis for creative thinking. Furthermore, he had read and had been intellectually excited by the writings of such scholars as Joseph Schumpeter (1942), Duncan Black (1948) and Anthony Downs (1957), scholarship that provided the basis for reintegrating economics with political science within a strictly rational choice framework. In short, Tullock was ready to play a significant role in extending the empire of economics into the territory of contiguous disciplines.

In fall 1958, at age 36, he accepted a one-year postdoctoral fellowship at the Thomas Jefferson Center for Political Economy at the University of Virginia. Still a relatively unknown quantity at that time, Tullock nevertheless brought with him to the Center two indispensable assets, namely a brilliant and inquiring, if still-unfocused, intellect and an unbounded enthusiasm for his adopted discipline

of political economy. Quickly he forged a bond with the Director of the Center, James M. Buchanan, a bond that would result in some of the most original and important political-economic scholarship of the mid-twentieth century.

His fellowship year at the Center was productive, resulting in a pathbreaking publication on the problem of majority voting (Tullock 1959). In fall 1959, Tullock was appointed as Assistant Professor in the Department of International Studies at the University of South Carolina. Publications continued to flow (Tullock 1961a, b) while Tullock crafted a seminal draft paper entitled "An Economic Theory of Constitutions" (Tullock 1959) that would become the fulcrum for *The Calculus of Consent* (Buchanan and Tullock 1962).

On this basis, Tullock quickly advanced to the rank of Associate Professor before returning to the University of Virginia, and renewing his relationship with James Buchanan, in February 1962, just as the University of Michigan Press was publishing their seminal book, *The Calculus of Consent*. In 1966, Tullock edited and published the first issue of *Papers on Non-Market Decision Making*, the precursor to the journal, *Public Choice*. Between 1962 and 1967, Tullock published innovative books on bureaucracy (Tullock 1965), on method (Tullock 1966), and on public choice (Tullock 1967a) as well as a rising volume of scholarly papers that earned him international recognition as a major scholar.

Despite this distinguished resumé, Tullock would be denied promotion to Full Professor of Economics on three consecutive occasions by a politically hostile and fundamentally unscholarly University administration. In fall 1967, Buchanan protested these negative decisions by resigning to take up a position at the University of California at Los Angeles. Tullock also resigned to become Professor of Economics and Political Science at Rice University. With Ronald Coase having resigned for similar reasons in 1964 to take up a position at the University of Chicago, it appeared that the nascent Virginia School of Political Economy might have been deliberately nipped in the bud by the left-leaning administration of the University of Virginia.

As a result of a successful initiative by Charles J. Goetz, the University of Virginia plot failed. Goetz succeeded in attracting Tullock to Virginia Polytechnic Institute (VPI) and State University (SU) in Blacksburg as Professor of Economics and Public Choice in fall 1968. Goetz and Tullock immediately established the Center for Studies in Public Choice in 1968, as the basis for promoting scholarship in the field and as a means of attracting James Buchanan to join them at VPI. This initiative bore fruit in 1969, when James Buchanan joined the VPI faculty and assumed the General Directorship of the Center, which was immediately renamed as the Center for Study of Public Choice. Simultaneously, Tullock renamed his journal *Public Choice* and the new subdiscipline set down fruitful roots in the foothills of the Appalachian Mountains.

Henceforth, Tullock would never again look back. Over the next one-third of a century he forged for himself a reputation as a brilliant entrepreneurial scholar and a formidable debater. To this day he refuses to rest on well-earned laurels as a Founding Father of three subdisciplines of economics, namely public choice, law and economics, and bio-economics.

Universities have recognized his contributions by appointing him to a sequence of Distinguished Chairs (VPI and SU 1972–1983, George Mason University 1983–1987 and 1999–, and the University of Arizona 1987–1999). Professional associations have honored him by electing him to their presidencies (Public Choice, the Southern Economic Association, the Western Economic Association, the International Bio-Economics Society, the Atlantic Economic Society, and the Association for Free Enterprise Education). In 1992, an Honorary Doctorate of Laws was conferred on him by the University of Chicago; in 1996 he was elected to the American Political Science Review Hall of Fame; and in 1998 he was recognized as a Distinguished Fellow of the American Economic Association. These awards and honors reflect his powerful entrepreneurial contributions across three major scholarly disciplines.

A Natural Economist?

James Buchanan has described Gordon Tullock as a *natural economist*, where natural is defined as having "intrinsic talents that emerge independently of professional training, education, and experience" (Buchanan, 1987, p. 9). A natural economist, therefore, "is someone who more or less consciously thinks like an economist" (*ibid*., p. 9). In Buchanan's judgment, there are very few such natural economists and most of those who claim competence in economics as a discipline are not themselves natural. Buchanan identifies Gary Becker and Armen Alchian, along with Gordon Tullock, as prominent members of the rare natural economist species.

Buchanan correctly recognizes that all economists of repute rely upon the rational choice model as the basis for analyzing the market interactions of human beings. Human beings are depicted as self-interested, utility maximizing agents for whom social interchange is initiated and exists simply as a preferred alternative to isolated action. Even though many economists do not fully endorse this model as an accurate depiction of individuals in society, they utilize it in market analysis on an "as-if" basis.

Yet, many of them waver or object when confronted with extending the rational choice model to the analysis of nonmarket behavior especially, one might conjecture, prior to Tullock's successful contributions in the 1960s. The behavior of such agents as politicians, voters, bureaucrats, judges, preachers, research scholars, family members, criminals, revolutionaries, terrorists, and media anchors, they argue, cannot be effectively captured in terms of the rational self-interest model. The natural economist has no such inhibitions.

In this perspective of Tullock's work, individuals exist as isolated islands in an ocean of exchange, solipsist in vision and poised irreversibly on the edge of the jungle (Rowley 1987a, p. 20). Because the natural economist is imbued comprehensively with a Hobbesian vision of the world, he cannot comprehend the

consentaneous promise expounded by Hume, Locke and, the young John Stuart Mill. He cannot model man as rising above his narrow self-seeking instincts.

George Stigler once suggested that a major difference between his own scholarship and that of Milton Friedman was that whereas Friedman sought to change the world, he (Stigler) sought merely to understand it. This distinction holds with equal force with respect to the scholarship of Buchanan and Tullock. Precisely because Tullock seeks to understand—even when what he learns is sometimes unappetizing—he adopts no subterfuge in his analytical approach.

If consent exists, Tullock notes and explores its rationale. If conflict is manifest, Tullock investigates the social dilemma to the extent possible with the tools of neoclassical economics. No judgment is passed; no policy recommendations are advanced. Tullock chronicles observed events as part of the pattern of a diverse universe that he is ever eager to explore. In this sense, Buchanan's insight, as I shall demonstrate, is accurate with respect to much of Tullock's scholarship, but inaccurate in important respects.

I should close this section by noting, however, that a natural economist need not manifest extreme solipsism in his own behavior. There is no reason why those who utilize self-seeking assumptions in scientific analysis should be seduced by the assumptions that they deploy into adopting an entirely solipsist mode of personal behavior.

Certainly, Tullock does not live the life of *homo economicus*, as the many faculty visitors and graduate students who have diverted him from his writings to share his intellectual curiosity, his ideas, and his wit will readily testify. If Tullock is generous with respect to his time, he is equally generous with respect to his modest wealth, as those who have dined—and dined well—at his table and those whom he has supported financially will also testify. He may well raise *homo economicus* as his indomitable standard on the field of intellectual battle. This standard is by no means the appropriate measure for evaluating his life (Rowley 1987a, p. 22).

The Calculus of Consent

The two most widely cited of Gordon Tullock's many contributions are *The Calculus of Consent* (coauthored with James Buchanan) published in 1962, and *The Welfare Costs of Tariffs, Monopolies and Theft* published in 1967. Let us focus briefly on Tullock's contributions to *The Calculus* as a means both of assessing his insights and of teasing out the limits of the natural economist hypothesis.

The Calculus is a momentous work of scholarship, the first major foray by Buchanan and Tullock into the terrain of political science, and the cornerstone of the Virginia political economy program. The principal objective of the book was to rationalize the Madisonian enterprise in strictly economic terms and to provide a logical rational choice foundation for constitutional democracy.

Fundamentally, the book was an exercise in team production, yet with each author bringing distinctive qualities to the enterprise (Rowley 1987b, p. 45).

Buchanan brought to the task an emphasis on modeling politics-as-consentaneous-exchange under the influence of Knut Wicksell. Tullock focused on modeling all agents in the constitutional endeavor in strict self-interest terms. By resolving this tension the coauthors wrote a masterpiece. In Tullock's contributions on logrolling and its implications for the simple majority voting rule (Chapter 10) and in his contributions on the bicameral legislature and the separation of powers (Chapter 16), we see the natural economist in his most unrelenting guise.

However, Tullock's central contribution to *The Calculus* was the economic theory of constitutions (Chapter 6) written at the University of South Carolina in 1959. This economic theory provides the logical foundation for constitutional democracy and, indeed, it is the anvil on which *The Calculus of Consent* was forged. Ironically, it is a chapter in which Tullock suppresses the self-interest axiom in its most myopic form as a means of identifying the unanimity principle as a rational individual decision-making rule for effecting constitutional choices.

In Chapter 6 of the book, Tullock assumes that the domain of collective action has already been determined and that the specific institutions through which collective action occurs are already in place. On this basis, he analyzes the choice of optimal rules by any random individual in society as a function of minimizing expected costs. Tullock distinguishes between two categories of expected cost, namely the expected external costs imposed on them by collective action and the expected costs of making decisions through collective action.

By recognizing that individuals fear the imposition of external costs upon them by government, Tullock challenged head-on the Platonic model of beneficent government which then dominated the political science literature. Only a rule of unanimity can protect any random individual from the imposition of such costs. By recognizing that expected decision-making costs are a monotonically increasing function of the number of individuals who must agree in order to effect collective action, Tullock was able to check the unanimity instincts of James Buchanan and to demonstrate that only voting rules short of unanimity are capable of minimizing the combined expected external and decision-making costs of collective action.

The rational individual, at the stage of constitutional choice, thus confronts a calculus not unlike that which he must face in making everyday economic choices. By agreeing to more inclusive rules, he accepts the additional burdens of decision making in exchange for additional protection against adverse outcomes and vice versa (Buchanan and Tullock 1962, p. 72). Tullock recognizes that differences in the burden of these costs with respect to specific constitutional choices will result in the selection by rational individuals of more or less inclusive rules. This insight explains the choice of supra-majority rules for collective actions involving such fundamental collective choices as life, liberty, and property in combination with the choice of significantly less inclusive rules for collective choices involving lower perceived external costs.

At this point, however, Tullock retreats from the concept of *homo economicus* in its narrow myopic form in order to focus on the mechanism through which random individuals who have selected optimal constitutional rules for themselves translate these choices into universally endorsed constitutional rules for society. This is a

significant issue. Individuals differ in many ways and, at any specific time, such differences will obstruct the achievement of universal consent.

Agreement, according to Tullock, is more likely regarding general rules for collective choice than for later choices to be made within the confines of certain agreed-on rules, because in the former case individuals are separated from their particular interests by a veil of uncertainty. Because general rules are expected to govern choices over lengthy time periods, individuals cannot predict with any degree of certainty whether they are likely to be in winning or losing coalitions on any specific issue. Their own self-interest in such circumstances will lead them to choose rules that maximize the expected utility of a randomly selected individual.

Consent will not occur without discussion. This is not the hypothetical world depicted by John Rawls in *A Theory of Justice* (1971). The discussion envisaged in *The Calculus of Consent* can be likened to that among players determining the rules of a card game before the cards are dealt. It is in the self-interest of each player at this stage to devise a set of rules that will constitute the most interesting game for the representative player. Once the cards are dealt, of course, no such agreement is likely, as *homo economicus* re-emerges to follow his self-serving instincts.

For universal consent over rules to be feasible, Tullock recognizes that participants must approach the constitutional convention as equals in the sense that differences are accepted without rancor and that there is no discernible dominant group that holds political power. For, such a group would not rationally divest itself of its authority. Therefore, *The Calculus of Consent* has little relevance for a society characterized by sharp distinctions between social classes, religious or ethnic groupings where one such grouping has a clearly advantageous position at the constitutional stage.

In 1787, this may not have appeared to be a problem for the United States because the limited suffrage went largely unchallenged. By 1860, it clearly was sufficiently important to destroy the Union. It is very surprising that Tullock completely failed to anticipate that this problem would re-emerge in the United States during the mid-1960s as long-term minorities began seriously to question the rules that had subjugated them to the whims of a dominant majority. The complete collapse of the U.S. Constitution in 1860, and its near collapse between 1968 and 1974, in any event, strongly conform to the predictions of the economic model.

Like all original insights, Buchanan and Tullock presented *The Calculus of Consent* to its intellectual audience in an embryonic form. Some forty years after its birth, significant and unresolved problems remain as is inevitable for any theory that purports to rationalize universal consent for less than unanimous decision-making rules in the real world.

Foremost among these problems is the silence of *The Calculus* with respect to the characteristics of the state of nature in the preconstitutional environment. Written as the book was, in the late 1950s, it is reasonable to infer that the authors envisaged a Lockeian state of nature governed by natural law that allowed individuals to protect inalienable rights to life and liberty and imprescriptible rights to private property.

In such an environment, individuals predictably will consent only to a set of rules that will require government to protect their natural rights (i.e., that limit the

domain of collective action to government as a minimal state). Because government will be so constrained, individuals anticipate that decision rules will be fully enforced by government as a referee and that collective action within those rules will not be reneged upon in the postconstitutional environment.

Once collective action bursts out of this restricted domain, as occurred in the United States in 1937, in the Supreme Court judgment of *West Coast v. Parrish* that destroyed forever the primacy of liberty to contract, considerations of conflict rapidly overwhelm those of consent, and constitutional rules are reformulated in a much less promising, more Hobbesian environment. This environmental shift was recognized simultaneously in 1974 at the peak of the Watergate crisis, by both coauthors of *The Calculus of Consent*.

Tullock's response was to write *The Social Dilemma* (1974) and to focus forever after on positive public choice in a Hobbesian environment. Under pressure, Tullock's natural economist instincts have resurfaced with a vengeance as his intellectual focus has switched from the potential for "gains-from-trade" to the reality of "generalized prisoners' dilemmas" and intractable hold-out situations.

Buchanan's response, in contrast, was to write *The Limits of Liberty* (1975) striving to rationalize the survival of consentaneous decision making in a Hobbesian world. Thereafter, Buchanan has focused almost exclusively on constitutional political economy, frequently changing tack to protect limited government from the adverse consequences of the predatory state (Brennan and Buchanan 1980, 1985; Buchanan 1990; Buchanan and Congleton 1998). Under pressure, Buchanan has reached, perhaps dangerously, beyond *homo economicus* in his attempt to provide an intellectual platform through which concerned private citizens might forestall the re-emergence of Leviathan in the United States.

The Political Economy of Rent Seeking

If Tullock dips his standard in *The Calculus of Consent*, he resurrects it with a vengeance in his seminal contributions to the rent seeking literature. Here we see the natural economist in his favorite role as he analyzes narrow self-seeking by individuals in the unrelenting Hobbesian environment of the churning, wealth-redistributing state.

Economic rent is a familiar concept to economists. It is simply defined as any return to a resource owner in excess of the owner's opportunity cost. Economic analysis identifies various categories of such returns—monopoly rents, quasi-rents, infra-marginal rents—that arise in market economies as a consequence of the less-than-perfect supply elasticity of factor inputs. Within a competitive market, the search for rents is nothing more than the normal profit seeking incentive that shifts resources to their most highly valued uses and creates new products and values (Tollison 1987, p. 144). Positive temporary rents induce new entry and negative temporary rents compel exit, in both cases impacting beneficially on economic output.

Tullock's rent seeking insight focuses attention on a malignant rather than a benign phenomenon. The notion that individuals and groups dissipate wealth by utilizing scarce resources to seek rents created for them by government is a classic insight by Gordon Tullock (Tullock 1967b). The insight is of pivotal importance for Virginia political economy. Arguably, it is the single most important contribution to the public choice research program and it remains, some 35 years after its inception, a major engine motivating public choice scholarship.

Tullock's insight was first presented in 1967 in an article published by *The Western Economic Journal* following its rejection by the well-known editors of three leading economics journals. The term "rent seeking" was associated with Tullock's insight some seven years later by Anne Krueger (1974) in a paper that, by accident, failed to reference Tullock's several prior contributions to the literature.

Tullock's attention was energized by a growing tendency for 1960s' economists to dismiss the welfare costs of monopolies and tariffs as unimportant in view of the minute values associated with Marshallian deadweight loss triangles of consumers' surplus imposed by such instruments (one-tenth of one percent of U.S. gross domestic product according to one measure devised by Arnold Harberger (1954, 1959)). Instinctively, Tullock sensed that such complacency was ill founded, and noted that "the classical economists were not concerning themselves with trifles when they organized against tariffs, and the Department of Justice is not dealing with a miniscule problem in its attacks on monopoly" (Tullock 1967b).

Tullock identified the Harberger fallacy by introducing a shift of emphasis based on a classic public choice insight. Generally, governments do not impose tariffs and do not create monopolies in a political market vacuum. They must be lobbied or pressured into so doing by the expenditure of resources in political activity by those who stand to benefit from such market protections. According to Tullock, rational producers would invest resources in lobbying, say for a tariff, until the expected marginal return on the last dollar so spent was equal to its expected marginal cost. Those who opposed the transfer would expend resources similarly in the opposite direction. All such outlays dissipate the rents expected by those who lobby. In certain adverse circumstances, such dissipation constitutes a complete waste of society's resources.

Tullock went on to demonstrate that rent seeking is not limited to the lobbying of government by private interests. In his 1975 article "Competing for Aid" (Tullock 1975b), he demonstrated how rent seeking for fiscal aid from the federal or state governments occurred among lower levels of government. This insight came from Tullock's experience in China where he observed how individuals deliberately mutilated themselves to make themselves attractive as recipients of charity. Similarly, the City of Blacksburg deliberately under-maintained its own roads (may even have deliberately damaged them) in order to become eligible for road-fund support from the Commonwealth of Virginia.

One of the major activities of modern government is the granting of special privileges to various politically influential organizations. Tullock observed that,

with notable exceptions, the profit record of such groups does not differ systemati-
cally from that of unprotected sections of the economy. In part, this may be because
the rents either have been dissipated up front or eroded by new entrants. In part,
however, the phenomenon is due to the capitalization of monopoly rents so that
only the original beneficiaries of the privilege make abnormal gains. Market
capitalization gives rise to a transitional gains trap where the revoking of a
government privilege imposes capital losses on second-generation rent recipients
(Tullock 1975a). It would seem, as David Friedman has put it, that "the government
cannot even give anything away." It is also evident that rational individuals will
lobby virulently to avoid the imposition of capital losses, making it extremely
difficult for politicians to support the abolition of special privileges once they
have been bestowed.

As with *The Calculus of Consent*, so it is the case with rent seeking, that
Tullock's original insight was presented to public choice in embryonic form.
Many of the gaps have now been closed (Tullock 1993). Two significant problems
yet remain unresolved.

The first is the ad hoc nature of rent seeking theory that constrains the generality
of its predictive power and that allows critics such as Stiglitz (1991) to contend that
"while these theories share with stock market analysts the ability to provide ready
interpretations of whatever occurs, their success in predicting these political forces
is much more limited." This is a fair criticism. Following the collapse of the Soviet
Empire in 1989 and the collapse of Enron in 2001, rent seeking rationalizations
abound. However, no public choice scholar predicted either of these collapses
in advance.

The second is the marked disparity between the magnitude of rents created by
the U.S. federal government and the relatively small level of observed rent seeking
outlays. Even if the efficient rent-seeking model (Tullock 1980a) is adjusted to take
account of risk aversion and increasing returns in rent seeking, this gap by no means
is reconcilable. In his 1989 book *The Economics of Special Privilege and Rent
Seeking* Tullock ingeniously rescues the rational choice model by suggesting that
rent seekers succeed in opaque rather than transparent markets and therefore are
forced to utilize inefficient techniques in rent seeking in order to escape voter
scrutiny. Such inefficient techniques are very costly and reduce the returns to rent
seeking. Ironically, the very inefficiency of their methods reduces the total of
rent seeking in society and ultimately mitigates the loss of wealth to society.

In this context, consider two types of worlds. In one, Tullock waste is exact and
complete. Here the incentive to create monopoly is low because there are no excess
returns from so doing. However, the social cost per instance of realized monopoly is
high. In the other world, politicians succeed in converting rent-seeking costs into
transfers. There are significant excess returns to monopoly creation. Hence, there
will be many more realized monopolies and many more Marshallian triangles of
deadweight loss imposed on society. It is not clear a priori which type of world is
preferable from the viewpoint of wealth maximization.

Let me conclude this discussion with an accolade to Gordon Tullock from one of his former colleagues, Robert Tollison, much of whose career has been expended on researching the rent-seeking research program initiated by Tullock:

> The theory of rent-seeking is here to stay. As I have observed in another context the most interesting thing about Tullock's ingenious insight is how simply he put it. Like Coase, he communicated his vision in terms that every lay economist could follow. This is a criterion by which greatness in science is measured. In economics, the Tullocks of our profession are more indispensable than ever. To wit, the scarcest thing in any science is a good idea, clearly communicated. (Tollison, 1987, p. 156)

The Vote Motive

The truly original insights into the vote motive must be ascribed to Duncan Black, whose writings during the late 1940s on the median vote theorem and the problem of vote cycles make him the undisputed Founding Father of public choice, and to Anthony Downs, whose 1957 book introduced rational choice analysis to the study of democracy and representative government and defined the paradox of voting implicit in the rational abstention of voters when confronted with large-scale elections. Tullock, nevertheless, leaves firm footprints on the sand with respect to this area of public choice scholarship.

First, Tullock has focused attention on the relevance of logrolling and vote trading for majority voting in representative assemblies. In 1959, his paper "Problems of Majority Voting" demonstrated that majority voting mechanisms in the absence of logrolling and vote trading deny voters the opportunity to seek gains from trade available to them where varying minorities care more passionately than varying majorities over specific programs in the policy bundles potentially available through the political process.

However, utility-maximizing logrollers, in the absence of binding contracts among each other, typically induce excessive public provisions (in terms of median preferences) under majority rule. Only by requiring supra-majorities can this weakness be avoided. This insight provides powerful support for a constitutional requirement that legislatures should always operate under supra-majority vote-rule constraints.

In 1981, Tullock returned to his earlier work on logrolling to address a perceived paradox in legislative behavior, namely the perceived stability of policy outcomes in a spatial environment seemingly conducive to endless cycling. His innovative paper entitled "Why so much stability?" initiated a major research program on the topic now referred to as "structure-induced equilibrium". Although Tullock's contribution is generally referred to as logrolling, in truth it falls directly within the structure-induced paradigm.

Tullock's contribution is based on the recognition that most government actions have the characteristic of providing a relatively intense benefit to a small group at a small cost to each member of a large group. Such bills are passed by several small groups getting together to logroll across their separately preferred programs.

In line with his work in *The Calculus of Consent* (1962), Tullock distinguishes between two forms that logrolling can take, namely individual bargains and formal coalitions.

Individual bargains predictably involve everyone since anyone excluded can offer lower prices for his vote in order to get back in. Tullock claims that a stable equilibrium is likely in such circumstances, though it will not be a Pareto optimum. In my judgment he is incorrect. As Bernholz (1974) established, if there is a cycle in the voting, there is also a logrolling cycle, unless individuals somehow can commit themselves to a specific bargain.

Tullock recognizes the instability of formal coalitions, given that those excluded from the majority coalition can destabilize it through counter-offers, since there will be over-investment in projects favored by members of the coalition and under-investment in projects favored by the minority. Moreover, there is little evidence either of formal coalitions in legislative bodies, or of any systematic exploitation of specific minorities. Rather, as Tullock observes, the committee structure of Congress creates stability to protect itself from the chaos of endless cycles:

> One simple procedure is to have the relevant committee which will, of course, contain representatives from both parties, canvass the House and decide which particular rivers and harbors bills would, in fact, pass if implicit logrolling were used on votes on each individual bill. This collection of specific projects can then be put together in one very large bill and presented to Congress as a unit. (Tullock, 1981, pp. 199–200)

This was the first attempt to explain the observed stability of political equilibrium under conditions conducive to cycling within the framework of a strictly rational choice model.

Second, (in 1967a) Tullock re-focused the rational voter abstention model of Downs (1957) in order to take account of the phenomenon of rational voter ignorance. If information is costly and if voters rationally economize in obtaining it, then the original equation of Downs, where the expected payoff to the individual from voting in an election is:

$$R = BP - C + D$$

changes to:

$$R = BPA - C_v - C_i + D$$

where B refers to the net personal benefit expected from the victory of the voter's preferred party or candidate, P refers to the probability that the voter's vote is decisive, A refers to the voter's subjective estimate of the accuracy of his judgment, C_v refers to the cost of voting, C_i refers to the cost of obtaining additional information and D refers to the consumption benefit received from voting.

Suppose, in such latter circumstances, argues Tullock, that C_v is negative as a consequence of social pressures, in which case voting is always rational. The cost of becoming adequately informed is much more expensive. In such circumstances, it would rarely be rational for the individual voter to cast a well-informed vote. In essence, most voters will be rationally ignorant (Tullock 1967a, 114).

The fact that the average voter is rationally ignorant opens up incentives for highly motivated members of the mass media to attempt to influence others in their voting behavior. Tullock also addresses this issue (Tullock 1967a). The expected payoff associated with such behavior is:

$$R = BP_p - C_i - C_p$$

where P_p is the probability that persuasion is decisive and C_p is the cost of persuasion. For individuals working in the mass media, P_p is much larger than P and C_p is likely to be zero. Advocacy therefore is a highly predictable activity in political markets. Advocacy will be directed most heavily at rationally ignorant swing voters whose behavior typically determines the outcome of political elections.

So far Tullock discusses the provision and consumption of political information without specific reference to the important issue whether or not such information is deliberately deceptive, although he recognizes that there is a fine distinction between persuasion and lies. In a further essay, on the economics of lying, Tullock (1967a) focuses on the incentives for politicians to lie to rationally ignorant voters in the course of election campaigns.

The expected benefit associated with a political lie comes from its success in securing votes. This is the product of the probability that the lie will be believed and the probability that it will persuade individuals to switch their votes in favor of the liar. The expected cost of a political lie is the sum of any cost to conscience and the product of the probability that the lie will be detected and the loss of votes associated with such detection. According to Tullock (1967a), the rational vote-seeking politician will lie to the point where the marginal expected benefits are equated with the marginal expected cost. Predictably, politicians will lie more extensively to the rationally ignorant than to the well-informed voters.

Because competing politicians have clear incentives to expose each others' lies, explicit lies are less likely than lies by inference. Politicians are well versed in such nuances of expression. Negative campaigning, where the respective campaign staffs of competing politicians, rather than the candidates themselves, lie about each other's candidate and accuse each other of lying is an excellent example of such nuanced vote-seeking behavior.

Tullock's natural economist instincts dominate in his approach to the vote motive. The current faddish popularity of theories of expressive voting, for example, wherein rational voters are assumed to vote their conscience rather than their interest, leaves Tullock unmoved and unconvinced. If individuals go to the polls, they vote their interest, as such interest is perceived to be through the fog of rational ignorance, stupidity, persuasion, and lies.

One senses (and shares) Tullock's skepticism concerning public choice scholars who relinquish the rational choice model in this field in favor of sociological explanations of human action. If Tullock's understanding of the vote motive speaks little for the net benefits of democracy, this does not concern him, nor should it concern us. Tullock views the world as it is and not as it ideally might be. From this

perspective, democracy is a very weak reed on which to rest the well-being of a nation, save when the domain of collective action is strictly and effectively curtailed by constitutional rules (Tullock 1989, 2000).

Bureaucracy

Tullock's 1965 book, *The Politics of Bureaucracy*, is the first application of the positive rational choice approach to a field that until then was dominated by "a normative mishmash of Max Weber's sociology and Woodrow Wilson's vision of public administration" (Niskanen 1987, p. 135). In this tradition, senior bureaucrats were viewed for the most part as impartial and well-informed servants of the prevailing public good as determined by each ruling government. The one prior book on bureaucracy by an economist (Ludwig von Mises 1944) was essentially devoid of analytic content. Tullock's (1965) contribution, therefore, inevitably was a voyage of discovery that opened up a fertile field for future research by challenging the fundamental premise that dominated the political science literature. Tullock is clearly influenced by Machiavelli's *The Prince* and by Parkinson's Law in modeling the behavior of senior bureaucrats and their subordinates.

Tullock models bureaucracy as a hierarchical system in which individuals advance by merit, as determined by senior bureaucrats. Ambitious self-interest motivates the behavior of all bureaucrats. The organizational system selects against moral rectitude. A man with no morals has a marked advantage over a more moral colleague who is willing to sacrifice career opportunities, at the margin, in pursuit of moral goals.

The moral quality of senior bureaucrats, therefore, with rare exceptions, is extremely low, not least because they must respond to the amoral behavior of ambitious underlings who seek to usurp their positions. There is no market check on the harmful organizational consequences of such unbridled personal ambition. It is also pointless to train bureaucrats in ethics, since self-interest dominates moral rectitude in this perverse non-market environment.

Because bureaus are hierarchical systems in which top-down decision making is the norm, Tullock identifies two major problems that lead to organizational inefficiency. First, instructions are unlikely to pass down the hierarchy without distortion even in the absence of malevolent design. Tullock refers to this as the problem of *whispering down the lane*. Second, senior bureaucrats cannot access fully the information available at lower levels of the hierarchy. If they delegate, they lose control. If they fail to delegate, their decisions will be ill-informed. Thus, Tullock shreds the central postulates of the political science research program and sets the scene for the economic analysis of bureaucracy.

Tullock (1965) focuses entirely on the internal organization of a bureau. Later work by Niskanen (1971) and by Weingast and Moran (1983) tightened the economic analysis and identified the link between bureaus and their sponsor

organizations. This shift of emphasis opened up the path to important empirical analysis that strongly supports the rational choice approach. Tullock's insights, culled from his personal experience in the Department of State, were indispensable to establishing this research program.

The Law

Tullock, the natural economist, rarely strays from positive rational choice analysis to engage in normative discussion. His first book on the law, *The Logic of Law* (1971a), however, is an exception to this rule. Here Tullock adopts utilitarian philosophy as first outlined by Jeremy Bentham, but as modified by Lionel Robbins (1938), by Nicholas Kaldor (1939) and by Hicks (1939).

Bentham's brand of utilitarianism comprises a combination of three conditions (Sen, 1987, p. 39), namely:

1. *Welfarism*, which requires that the goodness of a state of affairs should be a function only of utility information regarding that state of affairs;
2. *Sum-ranking*, which requires that utility information regarding any state of affairs should be assessed in terms of the sum of all individuals' utilities concerning that state of affairs; and
3. *Consequentialism*, which requires that every choice in society should be determined by the goodness of the consequent state of affairs.

Tullock's only formal training in economics was the course provided in the Chicago Law School by Henry Simons, who is best known for *A Positive Program for Laissez Faire* (1934), a propagandist tract, more an essay in utilitarian political philosophy than in economics (Coase 1993, p. 240). It is not surprising, therefore, that Tullock followed in his master's footsteps, albeit modifying the utilitarian ethic to suppress the sum-ranking condition in favor of the Pareto principle.

In *The Logic of the Law*, the first book ever published in law-and-economics, Tullock explicitly refers to Bentham's failed reforms of the English legal system, and claims that: "[s]ince we now have a vast collection of tools that were unavailable to Bentham, it is possible for us to improve on his work" and "[h]opefully this discussion, together with empirical research, will lead to significant reforms" (Tullock 1971a, p. xiv). On this basis, Tullock launches a critical review of substantive law and legal procedure within the United States as they existed in the late 1960s.

Tullock recognizes the limitations posed by the ordinal nature of utility and the inability to make interpersonal comparisons of utility. To overcome these restrictions, he falls back on the approach first developed in *The Calculus of Consent* (1962), in which individuals are viewed as focusing on potential reforms from a long-term *ex ante* perspective behind a veil of uncertainty. In such circumstances, legal reforms that myopic individuals who suffer a short-term loss of utility might

be expected to veto, nevertheless satisfy the unanimity requirement of the modified Pareto principle.

Tullock's critical eye takes in most areas of substantive law in the United States–contract, tort, theft, robbery, tax, and family—but focuses most savagely on legal procedures within the Anglo-Saxon legal system, a focus that he has sharpened even more with the passage of time as he has become yet more enamored with Napoleon (the civil code) and yet more skeptical of Wellington (the adversarial procedures of the common law).

The Logic of the Law (1971a), *Trials on Trial* (1980), and *The Case Against the Common Law* (1997) all utilize a writing style more appropriate for policy makers than for lawyers, rejecting the minutiae of legal footnotes for the straightforward prose and anecdotal evidence for which Tullock is renowned. Not surprisingly, Tullock has failed to achieve the same level of influence over the legal profession as he has, with respect to public choice, over economists and political scientists.

Most lawyers are rent seekers rather than scholars, slaves to the complex details of the law that provide them with their remuneration and profoundly mistrustful of ideas that appear to threaten their monopoly rents. It should come as no surprise that lawyers and legal scholars have responded much more favorably to the sophistry of Richard Posner, a fellow lawyer who advises them that their pursuit of private wealth through lucrative adversarial litigation indubitably contributes to the wealth of society (Posner, 1973).

Undeterred by this apparent failure to influence the American legal profession, Tullock continues to launch successive assaults upon Anglo-Saxon legal procedure. In so doing, he identifies the weak link of Chicago law-and-economics. For, if litigation leads to incorrect legal outcomes and legal errors are not automatically corrected by future litigation, the assertion that the common law is efficient is extremely difficult to sustain.

In his most recent, and arguably his best, book on this subject *The Case Against the Common Law* (1997) Tullock deploys the rational choice approach to powerful effect, demonstrating that a socialistic court system, with salaried bureaucrats (judges) and below-average intelligence jurors responding to the competing arguments of self-seeking lawyers, buttressed by the paid lies of their respective expert witnesses, within a system that is designed to restrict relevant evidence, is extremely unlikely to contribute positively to the efficiency of the law and to the aggregate wealth of society.

The fact that legal scholars of all brands, from Yale and Harvard to Chicago, choose to remain silent concerning the issues that Tullock raises, rather than to attempt to refute them, is suggestive that they know just how potentially devastating is his logic of the law for the continuation of the high incomes that they earn. Lawyers and legal scholars are sufficiently well trained in the Socratic technique to recognize the importance of voiding it when confronted with such a formidable debater, so better armed than they are in the logic of the law (Goetz 1987; Rose-Ackerman 1987; Schwartz 1987).

Bio-Economics

In 1957, shortly after leaving the Department of State and while working in Princeton, Gordon Tullock became interested in social insects and in other aspects of biology. He prepared a manuscript that would be published in a much revised form only one-third of a century later, dealing with issues of coordination without command in the organization of insect societies. In this early draft, he deployed economic tools to analyze the internal structure of ants, termites and a few other insect species. Tullock's monograph was well in advance of the pioneering work of Edward O. Wilson who is formally and correctly credited with founding the field of sociobiology.

Tullock's full bibliography contains a surprising number of publications in journals of biological science as well as a number of more popular publications in this field. One of these—his paper (Tullock 1971b) that applied economic principles to explain the behavior of the coal tit as a careful shopper—inspired a doctoral dissertation that provided a supportive empirical test of the avian feeding habits of the coal tit (Goetz 1998, p. 629).

Together with Janet Landa, Michael Ghiselin, and the late Jack Hirshleifer, Gordon Tullock ranks as one of the founding fathers of bio-economics. Most of his contributions were collected into his research monograph (Tullock 1994) entitled *The Economics of Non-Human Societies*. In this monograph, Tullock analyses the extraordinary feats of cooperation and adaptation to changes in their environments accomplished by ants, termites, bees, mole rats, sponges, and (his favorite) slime molds, species that have literally microscopic or non-existent brains.

Tullock assumes that animals, plants, ameboid single-cells of sponges, and the individual cells of slime molds all possess the functional equivalent of the preference function of human beings. This preference function is extremely primitive and is not necessarily mediated by way of a nervous system. A process of Darwinian selection and inheritance determines the success of such species in social coordination. He details the behavior patterns of such primitive species in terms of this rational choice model. It must be said that anyone who is prepared to argue the applicability of the rational choice model to the behavior of slime molds is indeed a natural economist!

The Editorial Initiative

Tullock's career as journal editor began inconspicuously in 1966 when he edited the first issue of *Papers in Non-Market Decision Making*, the precursor to the journal *Public Choice* that would become the spearhead of the public choice revolution and arguably one of the most influential policy-oriented journals of the last third of the twentieth century. From the outset, Tullock displayed enormous entrepreneurial talent in launching and directing this editorial initiative (Rowley 1991).

Historians of scientific revolution (Kuhn 1970) observe that textbooks and scholarly journals serve for the most part to consolidate rather than to initiate new research programs. The scholarly journals, in particular, tend to be conduits facilitating the preoccupation with "puzzle-solving" that normal science epitomizes. In this sense, journals are vehicles of normal science constrained by the vision of the past and, at most, are reluctant agents in the process of scientific revolution.

Tullock was well aware from the outset of the preoccupations of journal editorship; indeed, he had investigated the nature of the problem in his 1966 book entitled *The Organization of Inquiry* completed prior to embarking on his own editorial career (Rowley 1991). In that book, Tullock placed *homo economicus* center stage in the non-market decision-making environment of the typical scholarly journal and deduced on this basis an economic explanation of conventional editorial predilections for normal puzzle-solving science.

To understand the behavior of journal editors, Tullock argues, it is necessary to take account of the non-market environment of the academy, the institution central to the scholarly journal's success or failure. Universities, with few exceptions, are either publicly owned socialist institutions or are nonprofit organizations, in each case offering bureaucratic services in exchange for block appropriations and grants supplemented by fee income.

The senior bureaucrats responsible for their operations have few incentives to become acquainted with the details of their institutions' outputs, particularly with respect to the nature and quality of advanced research and scholarship. Yet, they have strong incentives to utilize low-cost filters for evaluating scholarly output as a basis for appointing, tenuring, promoting, and remunerating their academic work force. As a consequence, "[t]he whole responsibility for evaluating research, in essence, is left to the editors of the learned journals" (Tullock 1966, p. 37).

Unfortunately, most editors exercise only a subordinate role in the evaluation of scholarship, essentially providing a brokerage function between scholars on the supply and the demand side of the market for ideas. As Tullock observes, "the job of journal editor, although respectable, is not of sufficient attraction to get the very best personnel" (Tullock 1966, p. 141). In the typical case, where the editor is a respected but not a leading scholar in his discipline, truly important and innovative pieces of scholarship often will lie beyond his evaluation capacity.

In such circumstances, the use of anonymous readers becomes a lifeline for the intellectually overwhelmed editor. Recourse to this lifeline predictably fails to protect the path-breaking contribution. Leading scholars often either refuse to referee papers or provide only cursory evaluations. Hard-pressed editors thus submit manuscripts "to relatively junior scientists since such men are rather flattered at the honor and are unlikely to delay and delay" (Tullock 1966, p. 143). Under the shield of anonymity, the referee "is also not under any great pressure to reach the correct decision" (Tullock 1966, p. 143).

In such circumstances, Tullock argues, editors tend to discriminate against ground-breaking articles because of risk aversion in the face of augmented uncertainty:

> The probability of error on the part of the original investigator is greater, the possibility of error by the editor in misjudging the article is also great, and it is certain that the article, if published, will be very carefully examined by a large number of specialists. Under the circumstances, the possibility that the editor's own reputation will suffer from publication of such articles is a real one. It is not surprising, therefore, that these articles are sometimes hard to place. The problem is compounded by the fact that the prestige of a journal is affected by those it accepts; it is not affected by those it turns down. This probably leads the editors to some degree, at any rate, to play safe. (Tullock 1966, p. 147)

Yet, in his own lengthy editorial career (1966–1990), Tullock did not reflect his own logic, did not play safe, did not hide behind the anonymity of referees, did not slip from the cutting edge of public choice, and did not step down from the editorship of *Public Choice* even as his reputation became assured as one of the two leading scholars in the discipline. Instead, he deployed his journal as an active agent, seeking out contributions in areas where he detected important research deficiencies—vote models, logrolling, rent-seeking, the stability of political equilibrium, demand-revealing bureaucracy, and autocracy are noticeable examples.

He placed the journal firmly behind empirical research, recognizing the problem of obtaining good data, and allowing authors scope to experiment with respect both to the use of proxy variables and to method (Tullock 1991). Variable though the quality of published papers undoubtedly was, scholars of public choice were attracted like magnets to each issue of the journal for the gems that they might find—and might find only in *Public Choice*—because its editor was a genius and because rival editors both in economics and in political science, quite simply, were not. Once again, Tullock's behavior diverged from that of the natural economist in its public-spirited, self-effacing contribution to the development of an important discipline.

Tullock's World View

In many respects, Tullock does manifest the characteristics outlined by Buchanan (1987) as defining the natural economist. However, as this essay demonstrates, Tullock is much more than this. He is a warm-hearted and deeply concerned person with a powerful vision of the good society and a willingness to explore the reforms necessary to move mankind onto a better path.

In this regard, Tullock's philosophy is utilitarian in the modified sense of the Pareto principle, further adjusted to allow for individual decision making behind a veil of uncertainty. This philosophy, first spelled out in *The Calculus of Consent*, has been applied systematically by Tullock ever since wherever he has engaged in public policy discussion. Tullock is not an anarchist. He believes that there is a positive role for the state. No doubt, that role extends in his mind beyond that of the minimal or "night-watchman" state.

However, any such extension is extremely limited. Unlike many professed classical liberals, Tullock has not allowed himself to be diverted onto a normative Hobbesian path by the events of September 11, 2001. Rather, he has maintained a principled Lockeian position that a free society should never over-react to perceived violence and that basic constitutional rights should not be trampled on. He is a true friend of liberty, always watchful and vigilant in its defense. His good sense and common decency is much needed and highly valued in this increasingly troubled world.

References

Bernholz, P. (1974). "Logrolling, arrow-paradox and decision rules: A generalization." *Kyklos*, 27, 49–72.

Black, D. (1948). "On the rationale of group decision-making." *Journal of Political Economy*, 56, 23–34.

Brady, G.L. and Tollison, R.D. (1991). "Gordon Tullock: Creative maverick of public choice." *Public Choice*, 71(3), 141–148.

Brady, G.L. and Tollison, R.D. (1994). "Gordon Tullock: Creative maverick of public choice." in G.L. Brady and R.D. Tollison (eds.) *On the Trail of Homo Economicus*. Fairfax: George Mason University Press, pp. 1–6.

Brennan, H.G. and Buchanan, J.M. (1980). *The Power to Tax*. Cambridge: Cambridge University Press.

Brennan, H.G. and Buchanan, J.M. (1985). *The Reason of Rules*. Cambridge: Cambridge University Press.

Buchanan, J.M. (1975). *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press.

Buchanan, J.M. (1987). "The qualities of a natural economist." in C.K. Rowley (ed.) *Democracy and Public Choice: Essays in Honor of Gordon Tullock*. Oxford: Basil Blackwell.

Buchanan, J.M. (1990). "The domain of constitutional economics." *Constitutional Political Economy*, 1(1), 1–18.

Buchanan, J.M. and Congleton, R.D. (1998). *Politics by Principle, Not by Interest*. Cambridge: University of Cambridge Press.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Coase, R.H. (1993). "Law and economics at Chicago." *Journal of Law and Economics*, 36, 239–254.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Goetz, C.J. (1987). "Public choice and the law: the paradox of Tullock." in C.K. Rowley (ed.) *Democracy and Public Choice: Essays in Honor of Gordon Tullock*. Oxford: Basil Blackwell.

Goetz, C.J. (1998). "Tullock, Gordon (1922-)," in P. Newman (ed.) *The New Palgrave Dictionary of Economics and the Law*, Vol. 3. London: Macmillan, pp. 628–629.

Harberger, A.C. (1954). "Monopoly and resource allocation." *American Economic Review*, 44, 77–87.

Harberger, A.C. (1959). "Using the resources at hand more effectively." *American Economic Review*, 49, 134–146.

Hicks, J. (1939). *Value and Capital*. Oxford: Clarendon Press.

Kaldor, N. (1939). "Welfare propositions of economics and interpersonal comparisons of utility." *Economic Journal*, 49, 549–552.

Krueger, A. (1974). "The political economy of the rent-seeking society." *American Economic Review*, 64, 291–303.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*, 2nd edn. Chicago: University of Chicago Press.

Mises, L. von. (1944). *Bureaucracy*. New Haven: Yale University Press.

Niskanen, W.A. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine Press.

Niskanen, W.A. (1974). "Bureaucracy," in C.K. Rowley (ed.) *Democracy and Public Choice: Essays in Honor of Gordon Tullock*. Oxford: Basil Blackwell.

Niskanen, W.A. (1987). "Bureaucracy," in C.K. Rowley (ed.) *Democracy and Public Choice*. Oxford: Basil Blackwell, pp. 130–140.

Posner, R.A. (1973). *Economic Analysis of Law*. Boston: Little Brown

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Belknap Press.

Robbins, L. (1938). "Interpersonal comparisons of utility: a comment." *Economic Journal*, 48, 635–641.

Rose-Ackerman, S. (1987). "Tullock and the inefficiency of the common law," in C.K. Rowley (ed.) *Democracy and Public Choice: Essays in Honor of Gordon Tullock*. Oxford: Basil Blackwell, pp. 181–185.

Rowley, C.K. (1987a). "Natural economist or popperian logician"? in C.K. Rowley (ed.) *Democracy and Public Choice: Essays in Honor of Gordon Tullock*. Oxford: Basil Blackwell, pp. 20–26.

Rowley, C.K. (1987b). "The calculus of consent." in C.K. Rowley (ed.) *Democracy and Public Choice: Essays in Honor of Gordon Tullock*. Oxford: Basil Blackwell.

Rowley, C.K. (1991). "Gordon Tullock: Entrepreneur of public choice." *Public Choice*, 71(3), 149–170.

Schumpeter, J.A. (1942). *Capitalism and Democracy*. New York: Harper & Row.

Schwartz, W.F. (1987). "The logic of the law revisited," in C.K. Rowley (ed.) *Democracy and Public Choice: Essays in Honor of Gordon Tullock*. Oxford: Basil Blackwell, pp. 186–190.

Sen, A. (1987). "Rational behavior," in J. Eatwell, M. Milgate, and P. Newman (eds.) *The New Palgrave: A Dictionary of Economics*, Vol. 4. London: Macmillan, pp. 68–76.

Simons, H. (1934). In H.D. Gideonse (ed.) *A Positive Program for Laisser-Faire: Some Proposals for Liberal Economic Policy*. Chicago: Chicago University Press.

Stiglitz, J.E. (1991). "Another century of economic science." *Economic Journal*, 101, 134–141.

Tollison, R.D. (1987). "Is the theory of rent-seeking here to stay"? in C.K. Rowley (ed.) *Democracy and Public Choice: Essays in Honor of Gordon Tullock*. Oxford: Basil Blackwell, pp. 143–157.

Tullock, G. (1959). "Problems of majority voting." *Journal of Political Economy*, 67, 571–579.

Tullock, G. (1961a). "An economic analysis of political choice." *Il Politico*, 16, 234–240.

Tullock, G. (1961b). "Utility, strategy and social decision rules: Comment." *Quarterly Journal of Economics*, 75, 493–497.

Tullock, G. (1965). *The Politics of Bureaucracy*. Washington DC: Public Affairs Press.

Tullock, G. (1966). *The Organization of Inquiry*. Durham: Duke University Press.

Tullock, G. (1967a). *Toward a Mathematics of Politics*. Ann Arbor: University of Michigan Press.

Tullock, G. (1967b). "The welfare costs of tariffs, monopolies and theft." *Western Economic Journal*, 5, 224–232.

Tullock, G. (1971a). *The Logic of the Law*. New York: Basic Books.

Tullock, G. (1971b). "The coal tit as a careful shopper." *The American Naturalist*, 105, 77–80.

Tullock, G. (1974). *The Social Dilemma: The Economics of War and Revolution*. Blacksburg, VA: Center for Study of Public Choice.

Tullock, G. (1975a). "The transitional gains trap." *The Bell Journal of Economics and Management Science*, 6, 671–678.

Tullock, G. (1975b). "Competing for aid." *Public Choice*, 2, 41–52.

Tullock, G. (1980a). "Efficient rent-seeking," in J.M. Buchanan, R.D. Tollison, and G. Tullock (eds.) *Towards a Theory of the Rent-Seeking Society*. College Station: Texas A & M. University Press.

Tullock, G. (1980b). *Trials on Trial: The Pure Theory of Legal Procedure*. New York: Columbia University Press.

Tullock, G. (1981). "Why so much stability"? *Public Choice*, 37(2), 189–202.

Tullock, G. (1989). *The Economics of Special Privilege and Rent-Seeking*. Boston: Kluwer Academic Publishers.

Tullock, G. (1991). "Casual Recollections of an Editor". *Public Choice*, 71(3), 129–140.

Tullock, G. (1993). "*Rent seeking*," Shaftesbury Paper No. 3. Aldershot: Edward Elgar Publishing.

Tullock, G. (1994). *The Economics of Non-Human Societies*. Tucson: Pallas Press.

Tullock, G. (1997). *The Case Against the Common Law. The Blackstone Commentaries, No 1*. Fairfax, Virgina: The Locke Institute.

Tullock, G. (2000). *Exchanges and Contracts. The Blackstone Commentaries, No. 3*. Fairfax, Virginia: The Locke Institute.

Weingast, B.R. and Moran, M.J. (1983). "Bureaucratic discretion or congressional control: regulatory policy making by the federal trade commission." *Journal of Political Economy*, 91, 765–800.

Chapter 9
Mancur Lloyd Olson (1932–1998)[1]

Melvin J. Hinich[2] and Michael C. Munger[3]

Introduction

Mancur Olson had the rare ability to explain clearly the subtle nature of how groups of human beings organize to educated readers who are not versed in economic theory and its arcane vocabulary. His first book *The Logic of Collective Action: Public Goods and the Theory of Groups*, first published by the Harvard University Press in 1965, presented the fundamental issues of group formation by self-interested individuals in a way that made sense to most readers. He made a serious and successful effort to connect his theory with the scholarly writings on groups by sociologists and other non-economists. This book influenced the thinking of many academics and non-academics.

The technical term for the principal inherent problem of group formation is called the "free rider problem". If all members of the group receive a benefit from the action of the group and are not forced to contribute to the cost of group action, a self-interested group member has no incentive to contribute. As an example of this "problem" consider the case of the classical music station KMFA in Austin, TX. This station plays classical music of all sorts 24 hours a day. The station does not receive any public funds. The existence of this station depends on voluntary contributions from people who value the existence of the station. The number of people in the Austin area who contribute is a fraction of the people who listen to the station. The bi-annual fund-raising drives attempt to shame those listeners who do not contribute to pledge monetary support.

There is no hard evidence about how the appeals to the noncontributing listeners work, but the station has been able to raise enough funds to flourish. Most standard

treatments of the free rider problem in the modern economics literature involve the use of noncooperative game theory. This literature uses language and formalism that most educated people cannot comprehend unless they have taken a college level course in game theory. A person with a solid education and an ability to understand a logical argument can understand anything Mancur Olson wrote, however, because of the clarity of his thought and expression. Olson did not "solve" the free rider problem. He explained the problem without jargon and then he addressed how various groups attempt to deal with the problem.

He also developed two other important theoretical propositions, the implications of which will be explored for generations. These propositions have to do with (1) the implications of group development for economic performance and political health of a nation, and (2) the implications of the transition from "roving bandits" to "stationary bandits" as a necessary and sufficient condition for the existence of government. As James Buchanan wrote of him, "Mancur Olson was perhaps more influential in political science than any other of his economist peers." Given Buchanan's own enormous influence, this is a telling statement.

His Life

(The Stationary Bandit)... is not like the wolf that preys on the elk, but more like the rancher who makes sure that his cattle are protected and given water. The metaphor of predation obscures the great superiority of stationary banditry over anarchy and the advances in civilization that have resulted from it. No metaphor or model of even the autocratic state can, therefore, be correct unless it simultaneously takes account of the stationary bandit's incentive to provide public goods at the same time that he extracts the largest possible net surplus for himself. (Olson, 1993).

His Life

Mancur was born on a ranch in North Dakota in 1932, and he had ample opportunity to watch the wolf stalk the elk firsthand. There was a residual Scandinavian cadence, learned from his parents, in his speech throughout his life. His manner was plain and straightforward, and he was artlessly humble in his dealings with people he met. The version of his curriculum vitae he gave out in response to requests led off with his social security number, as if he might have to identify himself.

He graduated from North Dakota Agricultural College in 1954, and then won a spot as a Rhodes Scholar at University College, Oxford. From Oxford he returned to the United States and attended graduate school at Harvard. His dissertation became the remarkably influential book The *Logic of Collective Action* whose title was selected partly on advice from James Buchanan when it was later published in 1965. Only Anthony Downs' dissertation, *An Economic Theory of Democracy*, even comes close in terms of impact for a first work in the history of public choice.

Olson was hired at the Economics Department at Princeton, and then took the position of Deputy Assistant Secretary of the U.S. Department of Health, Educa-

tion, and Welfare (HEW) in the Johnson administration. In his two year at HEW, he had ample opportunity to observe the work of government, and the actions of interest groups, that would figure so prominently in his next book. In 1969 he left the government and accepted a job as Professor of Economics in the rather blue collar environs of the College Park campus of the University of Maryland. He was never seriously tempted by offers at more glamorous institutions, and remained at College Park for the rest of his life.

His Contributions

Olson's major contribution, explaining the logic of collective action, defies easy categorization. It is problem for the Left, because it argues that the key distinction is between consumers and producers, or between producers at different levels of the supply chain, rather than between capitalists and labor. Further, the problem of free riding is a direct assault on Marx's ideas about the inevitable construction of a workers' paradise where "from each according to his ability, to each according to his need" is the norm.

But it is also a problem for conservatives, because powerful producers will face no countervailing political force from consumers, wage earners, or less concentrated industries. The fact that many people free ride may be no more than an extension of the standard Samuelsonian "public goods" problem, but its political implications would seem to require government intervention to ensure compliance with laws, even if all citizens might support the law in principle. Further, since voting is also subject to the free rider problem, the claim that voters can be relied on to police corruption is rendered suspect, and questions are raised about the campaign finance system.

Olson's second important book, *The Rise and Decline of Nations* (1982), made the seemingly paradoxical claim that long-term political stability hurts economic performance. This would appear to contradict the conventional wisdom that stability and predictable actions by government enhance growth. But Olson claimed that interests become entrenched in stable democracies, so that governments suffer from "institutional sclerosis" as lobbyists manage a system of redistribution that distorts incentives and fails to reward initiative.

The apparent policy implication ("need growth? Fight a war!') was not taken very seriously, but Olson was making a very serious point about the dynamics of economic history. The clearest example is the explanation implied for the economic "miracles" of Germany and Japan after WWII. Their success, in Olson's view, had at least as much to do with the fact that all the entrenched interests in government were swept away along with the governments that were replaced after the war ended.

The larger question of why some nations prosper and others fail to grow, and the role of government in fostering growth or blocking it, occupied Olson in the last fifteen years of his life. His views differed substantially from those of both the

Chicago and Virginia public choice schools. Virginians tend to believe that all government is evil (with the possible exception of the Pentagon, and that is located in Virginia). Chicago public choice scholars have moved toward a system of theory in which government is efficient, at least in the way that it conducts transfers. Olson argued that government was neither evil nor efficient. His "Maryland school" centered on the idea that government does some things well, if it has the right incentives. And, surprisingly, it has the right incentives more often than one might think, at least in Olson's view.

The key concept was the "stationary bandit," introduced in his celebrated 1993 paper in the *American Political Science Review*. While many in the public choice movement have argued that government is simply a device for extracting benefits from citizens, Olson claimed that even a bandit, once stationary, had strong incentives to ensure at least minimal prosperity for citizens.

But more than this, the autocrat will recognize that problems of legitimate succession will still rob citizens of incentives to invest and build. The only way to ensure a truly long time horizon is to form a democracy. Further, since democracies appear less likely to go to war with one another, this form of government will have important evolutionary advantages if it can once be implemented. So, the original roving bandit who first settles down is the true father of democracy, and he created the democratic system purely out of his own self-interest in stability and prosperity.

This work is quite controversial, since it raises important questions about the nature of transition to democracy. Was Stalin bad, or was he inevitable? Olson tries to take on this question in portions of his last book, *Power and Prosperity* (2000), published posthumously.

Though Olson studied the collective action problem, he spent most of his professional life providing public goods for free. At conferences, in his office, and in discussing ideas he was happiest when he felt like he was learning something. He came from a generation of scholars who crossed a threshold, taking up very modern techniques but keeping a focus on the classical questions of moral philosophy.

References

McLean, I. (1998). "Mancur, scourge of special interests, Died on February 19, aged 66." *Economist,* March 7.

Olson, M. (1965). *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Olson, M. (1982). *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*. New Haven: Yale University Press.

Olson, M. (1993). "Dictatorship, democracy, and development." *American Political Science Review*, 87(3), 567–576.

Olson, M. (2000). *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships*. New York: Basic Books.

Chapter 10
William H. Riker (1920–1993)[1]

John H. Aldrich[2]

Introduction

William Harrison Riker, one of the founders of the Public Choice Society, arguably transformed the discipline of Political Science more than any single individual in the last half-century, creating the possibility of a genuine science of politics. It is difficult to measure the relative importance of his own scholarship, the vision of the scientific enterprise he imposed on the discipline, the training he gave a new generation of scholars, and the integration of this new understanding of political science into the social sciences. Each on their own was a legacy few achieve. Collectively, his contributions are, like the man himself, peerless.

Bill, as he was called, was born in Des Moines, Iowa, on September 22, 1920. He died on June 26, 1993. His beloved wife, Mary Elizabeth (M.E.) whom he married in 1943, passed away on March 14, 2002. He had two daughters and two sons, one of whom died 20 years before Bill.

Bill graduated from DePauw University in 1942. He deferred an acceptance to attend the University of Chicago, the leading graduate program in Political Science in the pre-War years, so that he could work in support of the war effort at RCA. In 1944, he concluded that Harvard University had emerged as the leading program, and left RCA to enter Harvard's Ph.D. program. He received his degree from there in 1948. He took a position at Lawrence College (now Lawrence University) that year, rising to the rank of Professor before he left for the University of Rochester in 1962, his home for the rest of his life.

Bill's training at Harvard was conventional for its day, although one must credit his contact with Professor Pendelton Herring for association with a scholar who, while not "scientific" in the sense Riker came to believe in, nonetheless was systematic in his analyses (Shepsle 2002). It was therefore only later at Lawrence

---

[2] Pfizer-Pratt University Professor of Political Science, Duke University, North Carolina 27708. e-mail: Aldrich@duke.edu

College, learning and working on his own, that he developed his views on the nature of political science and its place among the social sciences.

Early Influences

In 1954, the leading journal of the discipline, the *American Political Science Review*, published "A Method for Evaluating the Distribution of Power in a Committee System," by L.S. Shapley and Martin Shubik. In it, they developed their "power index" and applied it to the bicameral U.S. Congress and to the U.N. Security Council. This Shapley–Shubik power index is a special case of the Shapley value, and the article provided citations to that original paper (1953), Von Neumann and Morganstern's *Theory of Games and Economic Behavior* (1944, 1947) and Arrow's *Social Choice and Individual Values* (1951). These works, Downs' *An Economic Theory of Democracy* (1957), Buchanan and Tullock's *The Calculus of Consent: Logical Foundations of Constitutional Democracy* (1962), Duncan Black's *The Theory of Committees and Elections* (1958) and Black and R.A. Newing's *Committee Decisions with Complementary Valuations* (1951) provided Bill the rational choice theory through which he would seek to achieve a scientific study of politics.

Short-Term Consequences

These studies had three major consequences for his work over the remainder of the 1950s. First, his studies led him to think deeply about the nature of science, resulting in two papers in the *Journal of Philosophy*, "Events and Situations" (1957) and "Causes of Events" (1958a). Second, he began to consider the potential for rational choice theory, in general, and game theory in particular, to explain politics. He first did so by applying theories of others, beginning with "The Paradox of Voting and Congressional Rules for Voting on Amendments," published in the *American Political Science Review* (1958b). Shortly thereafter, these considerations also led him to revise an introductory text on American government he had written earlier (originally published in 1953), transforming it into the first rational choice book aimed at undergraduate audiences in Political Science (1965). He would soon develop his own theory of political coalitions, based on game theory. But first came the third major consequence of his theoretical development, one that included recognition from the academy, via becoming a Fellow at the Center for Advanced Study in the Behavioral Sciences in 1960–61.

His thinking on how to study politics had now almost fully matured. In his application to the Center, he wrote (quoted in Bueno de Mesquita and Shepsle 2001, p. 8):

> I describe the field in which I expect to work at the Center as 'formal, positive political theory.' By Formal, I mean the expression of the theory in algebraic rather than verbal symbols. By positive I means the expression of descriptive rather than normative

> propositions... I visualize the growth in political science of a body of theory somewhat similar to... the neo-classical theory of value in economics. It seems to be that a number of propositions from the mathematical theory of games can be woven into a theory of politics. Hence, my main interest at present is attempting to use game theory for the construction of political theory.

His Fellowship year was devoted primarily to writing *The Theory of Political Coalitions* (1962) which served as the study that backed those hopes with results. This book marked a transition from applying the work of others to the creation of his own theory. He recognized his account as taking off from ideas in *Theory of Games and Economic Behavior*, but unlike Von Neumann and Morgenstern, he developed the "size principle" for the case of n-person, zero-sum games. At this point he believed that politics was best understood as being a contest about winning and losing, and thus about zero-sum games, although later he would see that winning and losing did not necessarily lead to the zero-sum property. *The Theory of Political Coalitions* was more than the first fully developed rational choice theory by a political scientist. The book had a major impact on traditional political scientists and was widely used by the contemporary profession in guiding their empirical work, perhaps most heavily in the analysis of governing coalitions in multiparty parliaments. It thus was the first choice-theoretic (to say nothing of being the first game-theoretic) study to shape traditional scholarship in his discipline.

The Institutional Builder

The 1960s was a time of even more dramatic changes in Bill's career. In 1962, not only was *The Theory of Political Coalitions* published but he also accepted the position of Chairman at the University of Rochester, beginning a fifteen-year tenure as department chair. His task was to create a new Ph.D. program in Political Science that reflected his understanding of what a science of politics could be. He took what was essentially a small-to-medium sized liberal arts college's department and expanded it considerably—all the way up to 13 members a decade later!He did so by adding young scholars such as Arthur Goldberg, Richard Niemi, and John Mueller, who were trained as close to the vision Riker held of the discipline as was then possible, to the more traditional scholars already on hand. Of these young scholars, Jerry Kramer most fully embodied this vision with serious mathematical capabilities, well beyond those of anyone else then in the discipline, tied to a deep interest in matters political.

Perhaps the most remarkable first achievement of his chairmanship was the ability to graft a new political science on to a standing department, and the greatest fruit of this tree was the long-running, intellectual, and collegial departmental leadership coalition of Riker and Richard Fenno. This pairing created a remarkable training ground for new scholars almost immediately upon formation. The new program went from unranked to one ranked as number 14 in the nation by the end of the 1960s, that is, in under a decade of existence, and then to a "top ten" ranking the

next time such comparisons were made a decade later. At the end of its first decade, Bill's program already had or was in the process of training scholars who would play a role in the Public Choice Society, would be elected to the American Academy of Arts and Sciences, and/or would join Riker and Fenno in the National Academy of Sciences. These included such scholars as Peter Aranson, Morris Fiorina, Richard McKelvey, Peter Ordeshook, David Rohde, and Kenneth Shepsle (as well as this author).

The Intellectual Entrepreneur

The 1960s were, for Bill, fruitful not only in institutional creation, through the Ph.D. program he created and his role in forming the Public Choice Society, but also in terms of his own scholarship. In 1964, Little-Brown published his *Federalism: Origin, Operation, Maintenance* which some consider his greatest work. He, often with his graduate students, launched what are believed to be the first laboratory game experiments in political science (e.g., his "Bargaining in Three-Person Games," [1967] and with William James Zavoina, "Rational Behavior in Politics: Evidence from a Three-Person Game," [1970] both in the *American Political Science Review*). His 1968 article with then graduate student Ordeshook, "A Theory of the Calculus of Voting" (also in the *American Political Science Review*), remains controversial, heavily cited, and, it is fair to say, seminal over three decades later.

In some ways, it could be said that this intellectual decade ended with the publication in 1973 of his and Ordeshook's *An Introduction to Positive Political Theory*. This last book has associated with the actual title on its dust jacket the informal subtitle, "A synthesis and exposition of current trends in descriptive political theory based on axiomatic and deductive reasoning." It stands as the first graduate-level text of the application of rational choice theory to political problems, reflecting how much work had been completed in the area Bill had launched in the discipline less than two decades earlier.

To be sure, much of the original work considered had been done by social scientists in other disciplines (still mostly, but not exclusively, economics), but a substantial amount was done by political scientists. More to the point, the book covered a much wider variety of topics common to politics, especially democratic politics, than would have been possible a decade earlier. These included chapters on political participation, voting and elections, legislatures, and regulation and other aspects of bureaucracies.

While Bill had a truly far ranging intellect and therefore worked on a remarkable array of topics, he made unusual contributions to the study of three more questions that seem in retrospect to evolve naturally from what he had accomplished by 1973. Rational choice theory made its first and greatest impact (largely through Riker and the department he created) in the study of various aspects of the democratic process. In 1982a, his *Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice* (W.H. Freeman) brought the increas-

ingly wide-ranging and deepening set of formal results to bear on the normative foundations of democracy. His basic claim was that the results from social choice theory essentially rendered democracy as a choice between or among competing platforms (or what he meant by "populism") meaningless. He found in the results, however, the basis for justification of Madisonian liberalism, by which he meant elections as a referendum on the incumbent office holders.

Liberalism Against Populism also included results about the second of the three topics, institutions. One of the things that made the formal study of government and politics different from the study of market economies, especially in this period of work under general equilibrium theory, was that institutions both structured political competition and were the result of that competition. In 1982b, he began to develop the theoretical underpinnings of what he considered to be about as close to a law-like regularity in politics as could be found in his essay, "The Two-Party System and Duverger's Law: An Essay on the History of Political Science," (*American Political Science Review*). This, he suggested, was an institutional equilibrium resulting from the interactions of citizens and their political leaders taking place within a particular institutional context, in this case the context of plurality voting.

While Duverger's Law was therefore an institutional equilibrium, it was not a general one, because it was quite possible to change the institutional context. Political leadership unhappy with the two-party system need only change its elections from plurality to proportional methods, as New Zaland did in the 1990s, for example. As the theory would predict, that system changed from an essentially exact two-party system, with one of the two winning majority control of their legislature, to one with two larger parties but sufficient smaller ones to deny either party majority control.

This point about the "endogeneity" of institutions is general, he argued two years earlier. In 1980, he had made a devastating argument about the consequences of the general absence of voting (and other) equilibrium in politics (making politics, he claimed, the "truly dismal science"). In "Implications from the Disequilibrium of Majority Rule for the Study of Institutions" (*American Political Science Review*, 1980), he argued that the disequilibrium of voting was "inherited" by a sort of backwards induction onto the choices of rules. As a result, institutions were themselves as problematic as voting—the same problem that undermined "populism" in his thinking undermined institutions as well. He thus viewed institutions as little more than a temporary "congealing of tastes." They were no more a general equilibrium than was any other voting outcome, and they therefore carry no more moral weight.

Bill began his journey by seeking to establish a science of politics based on game theory. That journey led him and his fellow scholars to discover that the science of politics was very different from the science of economics, from which his ideas originated. The political analogy to the market is the election, but the analogy does not lead to a general equilibrium outcome, but one in which disequilibrium (or, its essential equivalent in these terms, a seemingly infinite number of equilibria) is common place. In addition, the problem of government is that of power, and

in particular that those who choose outcomes also choose large portions of the rules under which the government operates. The selection of these institutional features is just as fraught with instability (and lack ethical justification) as is the passage of ordinary legislation. Bill therefore turned to seek a new way to think about choice under disequilibrium (or a proliferation of equilibria).

He turned to acts of political leadership, looking at what he called political "heresthetics." This term was his creation to cover instances of manipulation of the strategic context to turn uncertain outcomes in one's favor. He first made this sort of argument in his presidential address to the American Political Science Association, "The Heresthetics of Constitution Making: The Presidency in 1787, with Comments on Determinism and Rational Choice" (published in the *American Political Science Review* in 1984). It was itself a heresthetical act of Riker to put this art of strategy to the academy first in that most "sacred" of secular locations, the founding of this Republic. He then collected a series of case studies to illustrate and develop this account in his *The Art of Political Manipulation* (1986). He continued the attempt to develop this part of the scientific explanation of politics to the end of his life, ending with a posthumously published account of the passage of the U.S. Constitution in the various states, *The Strategy of Rhetoric* (1996).

While Bill succeeded in redefining the discipline of Political Science, his work began, as it ended, in the larger realm of social science more generally. He began by drawing from Economics and from game theory. In the middle, his and Ordeshook's *An Introduction to Positive Political Theory* (1973) was a rich application of the work of many social scientists, including those cited already and other prominent Public Choice scholars such as Mancur Olson. Indeed, it was often through the Society, its journal, and its annual meetings that this work grew. Twenty years later, he and his students in political science were developing theories of social phenomena of sufficient originality and importance to return the favor to Public Choice scholars in other disciplines.

Conclusions

In the end, then, William H. Riker succeeded in placing political science within the set of scientifically based social sciences. He was among the first social scientists to apply game theory systematically to any major set of problems in a sustained way. He extended this vision to the discipline through his own work and that of the students he directly trained at the University of Rochester and at the (increasing numbers of) graduate programs that have emulated his. He linked political science to the other scientific social sciences, once again through his own scholarship and through institution building, notably through the Public Choice Society. He then brought his considerable energies to bear on understanding the nature and ethical standing of democracy through implications of the scientific results of he and his students. He addressed the central problem of politics (who rules the rulers) by including the study of institutions and of leadership in that "formal, positive political theory" he had promised 40 years earlier.

References

Arrow, K. J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Black, D. (1958). *The Theory of Committees and Elections*. Cambridge, England: Cambridge University Press.

Black, D. and Newing, R. A. (1951). *Committee Decisions with Complementary Valuation*. London: William Hodge.

Buchanan, J. M. and Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor, MI: University of Michigan Press.

Bueno de Mesquita, B. and Shepsle, K. (2001). "William Harrison Riker: 1920–1993." *Biographical Memoirs. National Academy of Sciences,* 79, 3–22.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper &Row.

Riker, W. H. (1953). *Democracy in the United States*. New York: MacMillan.

Riker, W. H. (1957). "Event and situations." *Journal of Philosophy*, 54, 57–70.

Riker, W. H. (1958a). "Causes of events." *Journal of Philosophy*, 56, 281–292.

Riker, W. H. (1958b). "The paradox of voting and congressional rules for voting on amendments." *American Political Science Review*, 52, 349–366.

Riker, W. H. (1962). *The Theory of Political Coalitions*. New Haven, CT: Yale University Press.

Riker, W. H. (1964). *Federalism: Origin, Operation, Maintenance*. Boston: Little-Brown.

Riker, W. H. (1965). *Democracy in the United States*, Second Edition. New York: MacMillan.

Riker, W. H. (1967). "Bargaining in three-person games." *American Political Science Review*, 61, 342–356.

Riker, W. H. (1980). "Implications from the disequilibrium of majority rule for the study of institutions." *American Political Science Review*, 74, 432–446.

Riker, W. H. (1982a). Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice. San Francisco, CA: Freeman.

Riker, W. H. (1982b). "The two-party system and Duverger's law: An essay on the history of political science." *American Political Science Review*, 76, 753–766.

Riker, W. H. (1984). "The heresthetics of constitution making: the presidency in 1787, with comments on determinism and rational choice." *American Political Science Review*, 78, 1–16.

Riker, W. H. (1986). *The Art of Political Manipulation*. New Haven, CT: Yale University Press.

Riker, W. H. (1996). *The Strategy of Rhetoric* (published posthumously with the assistance of R. Calvert, J. Mueller, and R. Wilson). New Haven, CT: Yale University Press.

Riker, W. H. and Ordeshook, P. C. (1968). "A theory of the calculus of voting." *American Political Science Review*, 62, 25–42.

Riker, W. H. and Ordeshook, P. C. (1973). *An Introduction to Positive Political Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Riker, W. H. and Zvoina, W. J. (1970). "Rational behavior in politics." *American Political Science Review*, 64, 48–60.

Shapley, L. S. (1953). "A value for n-person games." *Annals of Mathematics Study*, 28, 307–317.

Shapley, L. S. and Martein, S. (1954). "A method for evaluating the distribution of power in a committee system." *American Political Science Review*, 48, 787–792.

Shepsle, K. A. (2002). "Political losers." Inaugural William H. Riker Lecture, Public Choice Society, San Diego, CA, March 23.

Von Neumann, J., and Morganstern, O. (1944, 1947). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Chapter 11
Milton Friedman (1912–2006)[1]

Charles K. Rowley and Anne E. Bradley

Introduction

Throughout the first 15 years following the end of World War II the economics profession throughout the Western world was characterized by a touching belief in the omniscience and impartiality of government as the servant of the public good and by a cynical belief in the endemic failure of free markets to maximize social welfare as defined by the Pareto Principle supplemented by the Kaldor–Hicks–Scitovsky potential compensation test. It was also characterized by the hegemony of Keynesian economics with its support for large and fiscally interventionist governments and its contempt for monetary instruments of macroeconomic intervention.

The behemoths who bestrode the economics profession throughout that period were Paul Samuelson, Kenneth Arrow, John Kenneth Galbraith, James Tobin, Robert Solow, and Joan Robinson. Classical political economy was dead, buried by the Great Depression and by the writings of John Maynard Keynes (1936). Free market economics was on the ropes, with few advocates and even those forced into perpetual defense in an environment in which public choice analysis played no role. The future for economic liberty and capitalism was bleak indeed.

In a world in which free markets were systematically derided, there would be no effective role for the emergence of public choice to analyze political market failure and to even the playing field between the competing advocates of government and of free markets. First, it would be necessary for some brave soul to step forward and clear the brush from the forest by reformulating the case both for capitalism and for monetary policy as the fundamental basis for individual freedom. Then, and only then, would it be possible for the public choice revolution to begin. Such a soul was Milton Friedman, arguably the most influential economist of the twentieth century, one of that century's greatest economic advocates of liberty, and certainly a necessary precondition for and harbinger of the public choice revolution.

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, pp. 146–159.

The Early Years

Milton Friedman was born on July 31, 1912 in Brooklyn, New York City, the only son and the youngest of four children. His father, Jeno Saul Friedman (1878) and his mother, Sarah Ethel Landau (1881) were both born in the small, mostly Jewish town of Beregszasz in Carpetho-Ruthenia. Carpetho-Ruthenia was then in the Hungarian part of Austro-Hungary. After World War I, it became part of Czecho-slovakia; after World War II, it became part of the USSR; after the demise of the USSR it became part of Ukraine. The town is now called Berehovo.

At the age of 16 (1894) Friedman's father migrated to the United States and settled in Brooklyn. His mother migrated to the United States when she was 14 (1895). Shortly after her arrival, Friedman's mother went to work as a seamstress in a sweatshop, a welcome opportunity for her to earn a living while she learned English and adjusted to the new country. Shortly after his arrival, Friedman's father went into business on his own, first as a sweatshop owner, later as the owner of a retail dry goods store and an ice-cream parlor. He remained self-employed for the remainder of his life (Friedman and Friedman 1998, p. 20).

In these respects, both of Friedman's parents benefited from late nineteenth century capitalism as it was practiced in the United States. Their family income was always small and uncertain and they sometimes resorted to post-dated checks. For the most part, however, the family struggled to balance its budget while investing to the best of its ability in the education of its young (Friedman and Friedman, 1998, p. 21).

In 1913, little more than a year after Milton Friedman's birth, the Friedman family moved from Brooklyn to Rahway, a small town in New Jersey that served mostly as a bedroom city for commuters to New York and Newark. For most of Friedman's youth, his mother ran the store while his father commuted to New York where he worked as a jobber or petty trader. The common language within the household was English, since this was deemed to be essential for the family to function economically in the New World. Milton Friedman never became proficient in Hungarian, but he picked up enough Yiddish to understand the conversation of adults (Friedman and Friedman, 1988, p. 21).

Until shortly before his bar mitzvah at the age of 13, Milton Friedman was fanatically religious, attending Hebrew School at the local synagogue and con-forming in every detail to the complex dietary and other requirements of Orthodox Judaism. By the age of 12, however, he decided that there was no valid basis for his religious beliefs and he shifted to complete agnosticism, becoming fanatically anti-religious, although he did go through the bar mitzvah ceremony for the sake of his parents.

Friedman's father suffered for many years from angina and died from a heart attack at the age of 49 when his son was only 15 years of age and was preparing to enter his senior year in high school. Milton Friedman inherited this genetic defect and would have died himself in his 60s had he not benefited from the perfection of

by-pass surgery techniques. Friedman's mother and sisters worked to support the family while Milton, as the male sibling, was encouraged to complete his education.

From 1924 to 1928, Friedman attended Rahway High School, graduating with a good grounding in languages, mathematics, and history. He ascribes his enduring love of mathematics to a civics teacher who put the classic proof of the Pythagorean theorem on the blackboard while quoting from the poet Keats' *Ode on a Grecian Urn*: "Beauty is truth, truth beauty—that is all ye know on earth, and all ye need to know" (Friedman and Friedman, 1998, p. 24). Although the Friedman household could afford few books, Milton Friedman became a voracious reader, all but exhausting the contents of the small local public library.

Encouraged by two of his high school teachers who were recent graduates of Rutgers University, and by his success in competing for a partial scholarship at this then small New Brunswick College, Friedman entered Rutgers University in 1928 as a residential scholar. He worked his way through college as a part-time clerk in the men's department of a local department store, and he earned his lunch by waiting tables at a restaurant across the street from his dormitory (no doubt this embedded in the young Friedman, through an experience that was not available to his more pampered liberal democrat economist peers, an understanding that there is indeed no such thing as a free lunch).

Together with a fellow Jewish student, Friedman also engaged in a profitable entrepreneurial venture, buying and selling second-hand undergraduate books within the campus community, and in so doing, bringing upon himself the wrath of the university bookstore, whose margins he undercut. During the summer vacations, Friedman covered his living expenses by selling fireworks for the Fourth of July celebration, and by setting up a summer school for failing high school students, teaching classes in a number of subjects at 50 cents per hour. From an early age, Friedman showed an interest in buying and selling for profit and a predilection for entrepreneurial activity, having learned from his parents to embrace the capitalist market economy.

Friedman's original intention when entering Rutgers University was to major in mathematics with the objective of becoming an actuary. Although his actuarial results were extraordinarily good for an undergraduate, Friedman soon discovered that actuarial work was not the only paying occupation that used mathematics. Fortunately, he discovered economics and ended his degree program with the equivalent of a double major in economics and mathematics.

This decision changed his life, primarily because of his exposure to two remarkable men, Arthur F. Burns, who later would become Chairman of the Federal Reserve System, and Homer Jones who later would become Vice-President of Research at the St. Louis Federal Reserve Bank. Burns inspired in Friedman a passion for scientific integrity, for scrupulous accuracy in the checking of sources, and for openness to criticism. Jones provided Friedman with a sound grounding in insurance and statistics, while first introducing him to the Chicago School's preoccupation with individual freedom.

In 1932, at the age of 19, Friedman graduated from Rutgers University with a degree in mathematics and economics. With the United States in the depths of economic depression and job opportunities very few, he applied for scholarships to a number of universities and received two offers of tuition scholarships, one from Brown University in applied mathematics and the other from the University of Chicago in economics. Had Homer Jones not taught at Rutgers University, thus exposing Friedman to the excitement of Chicago economics, Friedman would never have applied to Chicago, or even if he had applied there he would not have received the scholarship that made it possible for him to attend. As it was, he left Rahway for Chicago in the fall of 1932, journeying west of the Delaware River for the very first time.

Studying economics at the University of Chicago was an eye-opening experience for the young Friedman. In 1932, as for the rest of the twentieth century, the Economics Department had the deserved reputation of being one of the best in the United States. Jacob Viner and Frank Knight were the acknowledged stars of the faculty. Henry Schultz, Paul Douglas, Henry Simons (who would move to the law school in 1939), Lloyd Mints, Harry A. Millis, and John Nef constituted a talented supporting cast. Debate over economic issues was fierce, the intellectual atmosphere was open, and the search for truth dominated scholarly discourse. Friedman was exposed for the first time in his life to a brilliant group of graduate students drawn to this vibrant intellectual atmosphere from all over the world (Friedman and Friedman, 1998, p. 35).

In his first quarter at Chicago, Friedman took the course on price and distribution theory taught that year by Jacob Viner. This course revealed to Friedman the logical and coherent nature of economic theory as a set of tools to be judged primarily by its usefulness in understanding and interpreting important economic events. Because the students in that class were seated alphabetically, Friedman was also introduced to Rose Director, his future wife and future coauthor of two important books on classical liberalism.

During his year at Chicago, Friedman also took Frank Knight's class on the history of economic thought. He was greatly impressed by Knight's unrelenting commitment to the pursuit of truth, and by his unfailing suspicion of government intervention, attributes that he himself would unfailingly observe in his own career. He was also exposed to Chicago monetary theory in courses taught by Lloyd Mints that focused on the fundamentals and not on the institutional arrangements. He strengthened his technical expertise with courses drawn from the department of mathematics and with a course in econometrics taught by Henry Schultz. Friedman became acquainted during his years at Chicago with fellow graduate students George J. Stigler and Allen Wallis, both of whom would later make important contributions to economics. He received his master's degree at Chicago in 1933.

At the urging of Henry Schultz, Friedman moved to Columbia University for the second year of his graduate work so that he could study with Harold Hotelling. Hotelling would provide him with the same kind of feeling for mathematical statistics that Viner had imbued in him for economic theory. Wesley C. Mitchell introduced him to the empirical analysis of business cycles and John Maurice Clark

introduced him to institutional economics. Fritz Machlup, a fellow student, introduced him to Austrian economics.

If Chicago had provided Friedman with the powerful tools of neoclassical price theory, and a favorable regard for free market capitalism, Columbia gave him the institutional framework and the facts relevant for testing those tools (Breit and Ransom, 1998, p. 227). Having satisfied the course requirements for a Ph.D., Friedman returned to Chicago for the academic year 1934–35 as research assistant to Henry Schultz, working with him on his path-breaking *magnum opus* on demand analysis. By the year's end, Friedman had also satisfied the course requirements for a Ph.D. at Chicago.

There were few academic jobs available in the United States at this stage of the Great Depression and Friedman's prospects were further lowered by the existence of anti-Semitism within the U.S. academy. Therefore, Friedman left Chicago in the fall of 1935 to take up a well-paid New Deal position in Washington with the National Resources Committee, where he spent two years designing and working on an extensive empirical study of consumer budgets. The experience that he acquired there working with practical statistics would prove to be invaluable throughout his scientific career. The final report on the results of this study was published in two volumes in 1938 and 1939.

Drawing from his work at the Committee, Friedman published an article on a new statistical technique—the use of ranks to avoid the assumption of normality implicit in the analysis of variance—in the December 1937 issue of the *Journal of the American Statistical Association*. The work he performed at the Committee also formed the basis of a book published in 1957 *The Theory of the Consumption Function*. Friedman has claimed that this is his most important scientific work (Friedman and Friedman, 1998, p. 66).

In the fall of 1937, Friedman moved to New York to begin what would prove to be a long association with the National Bureau of Economic Research (NBER) and with his former mentor at Rutgers University, Arthur Burns. Friedman worked at the NBER under the supervision of Simon Kuznets to fill in a major lacuna on data on income and wealth, namely the distribution of income and wealth by size. He edited the first three conference volumes on *Studies in Income and Wealth* for the NBER.

He also revised and completed a preliminary manuscript that Kuznets had drafted on the incomes of professional practitioners. The final product was a book, *Incomes from Independent Practice* completed in 1940 but published only in 1945 because of a hostile reception to the authors' conclusions by the then-powerful American Medical Association (Friedman and Friedman, 1998, pp. 74–76). Friedman and Kuznets determined that approximately one half of the observed excess earnings of physicians over dentists was explained by the success of the American Medical Association in limiting entry into medicine.

In addition to working at the NBER from 1937 to 1940, Friedman was able to secure a part-time teaching appointment at Columbia Extension. This was Friedman's first experience of formal teaching. In 1938, Friedman married Rose Director, six years after first meeting her in Chicago. They would have a daughter,

Janet, born in 1943 and now an attorney, and a son, David, born in 1945 and now a well-known professor of law and economics. Rose would be an active partner in her husband's professional life, coauthoring with him two extremely influential books on political economy.

Harold Groves, who had become acquainted with Friedman at the Income Conference, offered Friedman a visiting professorship at the University of Wisconsin for the year 1940–41. Despite efforts by Groves to make this position permanent, and despite an excellent teaching performance in statistics and economics, Friedman would be denied such a position in part at least because of overt anti-Semitism within the Department of Economics (Breit and Ransom, 1998, p. 227; Friedman and Friedman, 1998, pp. 91–104).

From 1941 to 1943, Friedman joined Carl Shoup at the U.S. Treasury as a Principal Economist in the Tax Research Division. During this period, Friedman, according to his wife Rose, made the worst intellectual mistake of his career (Friedman and Friedman, 1998, p. 123). In order to raise taxes for the war effort, he helped to devise a scheme for withholding income tax at the source of the income. This innovation (for which Friedman was not solely responsible) arguably is the most important single cause of the growth of government in the United States during the second half of the twentieth century.

Tax withholding tends to obscure the full burden of federal and state tax liabilities on individual taxpayers and therefore provides an illusionary reduction in the cost of government. By developing the idea and advocating its implementation as a policy tool, Friedman showed no instinct for the likely public choice implications and no concern for the adverse impact of such a policy for individual liberty (Rowley, 1999b, p. 416).

Friedman soon got tired of his work on tax reform and moved to New York in 1943 to join a close friend, Allen Wallis, as Associate Director of the Statistical Research Group at Columbia University. There he joined a group of distinguished statisticians, including Abraham Wald, Jacob Wolfowitz, and Harold Hotelling in work directly relevant to the war effort. Together with Wallis and Wald, he would develop a method of sequential sampling designed to help the U.S. Navy in its sampling inspection of wartime production of munitions. Sequential analysis became the standard method of quality control inspection.

At the same time, Friedman at last turned to his doctoral dissertation drawing upon his NBER study. He completed the dissertation in 1945 and received his doctorate from Columbia University in 1946, more than ten years after completing his course work for the degree. Of course, the war was a special circumstance. However, it would now be virtually impossible for an American student to drag out his degree program over such a lengthy period and still be allowed to graduate. It is fortunate for economics that Columbia University was flexible. Without a doctorate, Friedman would have found it all but impossible to enter postwar American academia.

Although Friedman's early career appears to have been a patchwork of short-term appointments, it formed the basis for all of his subsequent work. Well versed in mathematics and statistics, formidably well trained in economic theory, and well

experienced in economic policy making, he would be uniquely equipped to confront a postwar economics profession obsessed with Keynesian macroeconomics, seduced by socialist dogma and aggressively hostile to classical liberal political economy.

The Allied victory over Japan in August 1945 coincided with the end of the "Wilderness Years" for Milton Friedman, years that had been immensely enjoyable and productive despite the absence of any permanent academic position. With the help of George Stigler, Friedman secured a one-year appointment for the academic year 1945–46 at the University of Minnesota, where he taught courses in statistics and economics. Long before the academic year was over, he would accept a tenured position at Minnesota with the rank of associate professor.

During his year at Minnesota, Friedman coauthored with George Stigler a pamphlet attacking rent control, with the catchy title of *Roofs or Ceilings*. The National Association of Real Estate Boards circulated 500,000 copies of this pamphlet as part of its campaign against rent controls. Friedman's first foray into classical liberalism marked him as a rising star among the small group of classical liberal scholars in the United States. It also marked the early signs of a willingness to stand firmly against mainstream economic thinking when such thinking could be shown to run counter to sound economic theory (Friedman and Friedman, 1998, p. 150).

In the spring of 1946, Stigler received an offer from the Department of Economics at the University of Chicago, contingent upon approval by the central administration after a personal interview. President Ernest Colwell vetoed the appointment on the grounds that his work was excessively empirical. Ironically, Friedman was offered the position initially offered to Stigler and the New Chicago School was born with an unrelenting emphasis on empirical analysis that has continued throughout the remainder of the twentieth century.

Friedman returned to the University of Chicago in the fall of 1946 as associate professor of economics, succeeding his mentor Jacob Viner in the teaching of microeconomic theory. In 1948, he was promoted to full professor. In 1951, he became the third recipient of the prestigious John Bates Clark medal. In 1963, he was appointed Paul Snowden Russell Distinguished Service Professor. In 1967, he was elected president of the American Economic Association. In 1976, he was awarded the Nobel Prize in Economic Sciences. Friedman retired from the University of Chicago in 1977, moving to the Hoover Institution at Stanford University as a senior research fellow, where he would remain until his death in November 2006.

The Path to Scientific Recognition

Unlike several of his classical liberal contemporaries—James M. Buchanan, Ronald H. Coase and Gordon Tullock—Milton Friedman forged his scientific reputation not by traveling less well-trodden paths but by a sequence of brilliant challenges to mainstream economics. The Royal Swedish Academy of Science, in

awarding him the Nobel Prize in Economic Sciences in 1976, cited Friedman "for his achievements in the fields of consumption analysis, monetary history and theory and for his demonstration of the complexity of stabilization policy." This section focuses on these contributions and assesses their implications for classical liberal political economy and for the public choice revolution.

The Methodology of Positive Economics

During his graduate years at Chicago, Friedman had been taught by Frank Knight who evidenced extreme skepticism towards empirical economic analysis. None of the leading scholars at Chicago during the 1930s showed any real interest in numbers. Quite possibly, Friedman would have embraced that skepticism, had he been able to move directly into an academic position in 1935. Experience at the National Research Corporation (NRC) and the NBER during his Wilderness Years, however, taught him to respect empirical analysis and led him to think deeply about the methodology of positive economics. When he returned to Chicago in 1946, he determined to make sense of the kind of work that he had undertaken with Kuznets and Burns. In so doing, he would make an important contribution to methodology that would be the defining characteristic of the new Chicago School of Economics.

During the 1930s the economics profession had become enamored of a view advanced by Lionel Robbins that the veracity of an economic model should be tested primarily by the correspondence between its assumptions and the facts (Walters, 1987, p. 423). Specifically Robbins explained: "But the final test of the validity of any such definition is not its apparent harmony with certain usages of every day speech, but its capacity to describe exactly the ultimate subject matter of the main generalizations of science" (Robbins, 1932, p. 4–5). Thus Robbin's view was that the assumptions of good science must directly reflect empirical reality.

This view encouraged significant challenges to the model of perfect competition from critics such as Joan Robinson and Edward Chamberlin who claimed that the assumptions of the perfectly competitive model failed to conform to the reality of twentieth-century markets. It also stimulated attacks on all theories that incorporated the assumption that firms maximize profits. More fundamentally, the Robbins test was being widely deployed to attack the laissez-faire model of economics (Samuelson, 1963, p. 213).

As early as 1947, Friedman circulated in draft form a radically different view of the proper methodology for positive economics than that espoused by Robbins. Six years later, in 1953, Friedman's article on the methodology of positive economics would make a controversial but long-lasting entry into the litany of economics.

In preparing his essay, Friedman benefited slightly both from a brief conversation with Karl Popper (whose great book *Logik der Forschung* was not yet available

in the English language) and from his collaboration with James Savage whose book *The Foundations of Statistics* would shortly revolutionize the philosophical foundations of statistics (Friedman and Friedman, 1998, 215). Ultimately, however, the methodology outlined in Friedman's 1953 essay is uniquely his own.

At the outset of his essay Friedman states that: "The ultimate goal of a positive science is the development of a 'theory' or 'hypothesis' that yields valid and meaningful (i.e., not truistic) predictions about phenomena not yet observed." (Friedman, 1953, p. 7). He reinforces this view in the following terms: "Viewed as a body of substantive hypotheses, theory is to be viewed by its predictive power for the class of phenomena which it is intended to 'explain'" (Friedman, 1953, p. 8). In this respect, a hypothesis can be falsified but never verified:

> The hypothesis is rejected if its predictions are contradicted ("frequently" or more often than predictions from an alternative hypothesis); it is accepted if its predictions are not contradicted; great confidence is attached to it if it has survived many opportunities for contradiction. Factual evidence can never "prove" a hypothesis; it can only fail to disprove it, which is what we generally mean when we say somewhat inexactly, that the hypothesis has been "confirmed" by experience. (Friedman, 1953, p. 8–9)

This emphasis on prediction leads Friedman to reverse the epistemic order presumed in orthodox methodology (Hirsch and Marchi, 1990, p. 76). Instead of reasoning from true causes to implications, Friedman reasons from observed implications to possible premises. In this view, the premises of a successful theory are accepted to the extent to which they yield a set of predictions that has not been falsified by the available evidence. The simpler and the more fruitful the premises involved, the more acceptable they are, given the accuracy of the predictions that they generate.

From this perspective, Friedman launched a controversial and, in retrospect, almost certainly an exaggerated attack on the ruling convention that a theory should be tested by the realism of its assumptions.

> Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumption... A hypothesis is important if it "explains" much by little, that is if it abstracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone (Friedman, 1953, p. 14).

Friedman immediately modified this startling and memorable assertion with a more cautious explanation:

> The relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic", for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions (Friedman, 1953, p. 15).

Friedman's statement of methodology did not meet with widespread early acceptance within an economics profession yet unacquainted with the writings of Karl Popper. Most of the early critiques were ad hoc in nature, more designed to buttress

the ongoing attack on neoclassical theory than to provide profound insights. In 1963, however, Paul Samuelson entered the debate with a more formal attempted rebuttal of Friedman's methodology (Samuelson, 1963).

Samuelson focused attention on Friedman's assertions (1) that a theory is vindicated if some of its consequences are empirically valid to a useful degree of approximation, (2) that the empirical unrealism of the theory itself, or of its assumptions, is quite irrelevant to its validity and worth, and (3) that it is a positive merit of a theory that some of its content and assumptions are unrealistic.

According to Samuelson (1963), this methodology is incorrect as a matter of logic. Define a theory (call it B) as a set of axioms, postulates, or hypotheses that stipulate something about observable reality. Fundamentally, this theory contains everything—assumptions as well as consequences—and is refuted or not as a whole by reference to how well it conforms to the relevant evidence. Friedman denies this and argues instead that B has consequences (call them C) that somehow come after it and assumptions (call them A) that somehow are antecedent to it. What are the implications of this separation?

According to Samuelson $A = B = C$. If C is the complete set of consequences of B, it is identical with B. B implies itself and all the things that itself implies. Thus, if C is empirically valid, then so is B. Consider, however, a proper subset of C (call it $C-$) that contains some but not all the implications of B and consider a widened set of assumptions that includes A as a proper subset (call it $A+$). Now suppose that C has complete empirical validity. Then so has B and so has A. However, the same cannot be said for $A+$. Similarly, the empirical validity of $C-$ does not of itself impart validity to A or to B.

If Samuelson is correct, Friedman's methodology is scientifically flawed. For example, it may well be the case that certain characteristics of the model of perfect competition conform to reality ($C-$ as Friedman would argue). However, other parts do not (A as Friedman would acknowledge). In such circumstances, the model (B in Samuelson's broader sense) has not been validated and economists should proceed with extreme care in making use of it even if the evidence strongly and consistently conforms to $C-$.

Samuelson's deconstruction is valid, however, only for a methodology that views theory as moving from cause to effect, the very methodology that Friedman rejected in his 1953 essay. The real question for Friedman is to gage the extent to which the assumptions of a theory are adequate for the job in hand, which is to generate predictions that conform with the available evidence. He rejects on methodological grounds the notion advanced by Samuelson (1963) that a theory must be realistic in all its aspects.

To put Friedman's central thesis in a nutshell, it is that "the ultimate test of the validity of a theory is not conformity to the canons of formal logic, but the ability to deduce facts that have not yet been observed, that are capable of being contradicted by observation, and that subsequent observation does not contradict" (Friedman, 1953, p. 300). In this respect, Friedman's 1953 views on methodology, though contentious at the time, proved to be consistent with those of Karl Popper and

provided the intellectual foundations first for the new Chicago School of Economics and subsequently for a significant section of the economics profession.

This shift of methodology proved to be very important for Friedman's subsequent empirical re-evaluation of Keynesian economics and for his empirical work on the role of money in the macroeconomy. By persuading many economists that economic science could be advanced by exposing the predictions of very simple models to the evidence, Friedman would be able to demonstrate, for example, that the quantity equation was a better predictor of economic behavior than the Keynesian income–expenditure equation. This result would have enormous implications for reining in fiscal interventions that threatened individual liberties.

Fiscal Policy is Over-Rated

In evaluating the evolution of a scholar's career, it is important not to do so from the end point of that career from the perspective of hindsight. This is particularly so when evaluating Friedman's critique of Keynesian economics. Ultimately, the success of this critique would constitute his most important contribution to classical liberal political economy and a significant assistance to the public choice revolution. However, Friedman's critique of Keynesian economics was piecemeal in nature, and certainly did not start out as a grand design.

Friedman was always more impressed with the scholarship of Maynard Keynes than with that of the Keynesians. Indeed, Friedman viewed Keynes, like himself, as a purveyor of the economics of Alfred Marshall (Hirsch and Marchi, 1990, p. 187). Keynes's *General Theory* (1936) made an indelible impression on economic thinking during the immediate postwar years, and the young Friedman was sufficiently impressed by it to allow the Keynesian model to dictate much of his research agenda during the 1940s and 1950s.

Friedman's early preoccupation with the Keynesian model was motivated not by ideological concerns but rather by empirical puzzles surrounding a relationship at the core of the Keynesian system, namely the consumption function. According to the Keynesians, current consumption expenditure was a stable function of current income. A fundamental psychological rule of any modern community dictated that the marginal propensity to consume was less than one and that the average propensity to consume declined with income.

These two conjectures became matters of policy importance. Governments seized on the first as a scientific justification for deficit spending during periods of recession. Economists seized on the latter to consolidate the secular stagnation thesis and to suggest that advanced economies would be condemned to stagnation in the absence of deficit financing. In both instances, the fallacy of a free lunch enticed the unwary into embracing the palliative of government growth, given that government apparently could exploit the consumption function, increasing

household incomes by increasing government expenditures, in order to achieve a leveraged impact on the macroeconomy through the multiplier mechanism.

In his book, *A Theory of the Consumption Function* (Friedman, 1957), Friedman addressed a number of empirical puzzles surrounding this theory. Early work using U.S. data for the interwar period had seemed to support the theory (Friedman, 1957, p. 3). However, postwar studies were more problematic. Estimates of saving in the United States made by Kuznets for the period since 1899 revealed no increase in the percentage of income saved during the past half century despite a substantial rise in real income (Kuznets, 1952, pp. 507–526). The ratio of consumption expenditure to income was decidedly higher than had been computed from the earlier studies.

Examination of budget studies for earlier periods strengthened the appearance of conflict. The average propensity to consume was roughly the same for widely separated dates despite substantial differences in average real income. Yet each budget study separately yielded a marginal propensity decidedly lower than the average propensity. Finally, the savings ratio in the period after World War II was sharply lower than that predicted by the relationships estimated for the interwar period. According to Friedman's methodology, something was seriously amiss. The Keynesian consumption function had failed a basic empirical test (Friedman, 1957, p. 4).

In his book, *A Theory of the Consumption Function* (Friedman, 1957), Friedman adapted the dynamic theory of Irving Fisher (1930) to explain some of the empirical anomalies that had arisen in attempts to test the static Keynesian model against time series and cross-section data on consumption and income (Sargent, 1987). This book is Friedman's best purely scientific contribution and the work that best reflects his methodology of positive economics (Walters, 1987).

Irving Fisher (1930) had posited that consumption should be a function of the present value of income, not of its current value. Friedman accepted the dynamic implications of this theory, but replaced Fisher's concept with the concept of *permanent income*. He posited that consumers separated their current income into two parts, namely a permanent part equivalent to the income from a bond and a transitory part equivalent to a nonrecurring windfall. In testing the theory of the consumption function against cross-section data, econometricians must resolve a signal extraction problem in order to estimate the permanent component of income from observations on the sum of the permanent and the transitory components of income.

To model the time-series data, Friedman introduced the concept of *adaptive expectations* to create a statistical representation of permanent income. Agents were assumed to form expectations about the future path of income as a geometric distributed lag of past values. The decay parameter in the distributed lag ought to equal the factor by which the consumer discounted future utility. Friedman estimated his model on time–series data using the method of maximum likelihood.

On this basis Friedman (1957) demonstrated that there exists a ratio between permanent consumption and permanent income that is stable across all levels of permanent income, but that depends also on other variables, most notably the

interest rate and the ratio of wealth to income. The transitory components of income have no effect on consumption except as they are translated into permanent income.

From the perspective of the 1950s, Friedman's analysis had very important consequences for macroeconomic policy. First, it suggested that the immediate fiscal policy multiplier was markedly lower than that posited by the Keynesians. Second, it indicated that the dynamic responses of income to fiscal policy shocks were much more complicated than those indicated by textbook IS–LM curves. Both results suggested caution in the use of fiscal policy as a stabilization device.

Although the Keynesians argued that fiscal policy should be used even-handedly across the business cycle, countering recessions with budget deficits and booms with budget surpluses, the political system confounded such naïve expectations (Buchanan and Wagner, 1977). The political incentives to maintain budget deficits during booms as well as slumps simply overwhelmed economic logic. Therefore, to the extent that Friedman's theory dampened economists' enthusiasm for an active fiscal policy, it thus helped to dampen the rate of growth of government. Friedman's book, although devoid of any notions of public choice, nevertheless provided an invaluable foundation for the later work in 1977 by Buchanan and Wagner on the political economy of deficit finance.

It is important to note that Friedman has never argued that fiscal policy is completely impotent. His own theory of adaptive expectations indeed supposes that individual responses to fiscal policy occur with lags, allowing fiscal policy to exert an influence on the macroeconomy during the period of adjustment. Friedman's crucial insight is that monetary policy typically is more effective than fiscal policy as an instrument of macroeconomic policy.

It is also important to note that New Keynesian views are now very much the mainstream in macroeconomics, albeit operating within a rational expectations framework that offers only limited scope for fiscal intervention and a much greater role for monetary policy than was envisaged by the original followers of Keynes.

Money Matters

Friedman's interest in the role of money in the macroeconomy was first sparked in 1948 when Arthur Burns at the NBER asked him to research the role of money in the business cycle. Thus began a 30-year program of research with Anna Schwartz that would demonstrate that money matters—indeed that it matters a great deal—and that would further erode the perceived empirical importance of the Keynesian model.

By 1948, Keynesian economic theory ruled triumphant throughout the academies of the Western world. The classical quantity theory for the most part had been eliminated from textbook economics; and where it was mentioned it was treated as a *curiosum*. The conventional view throughout the economics profession was that money did not matter much, if at all. What really mattered was autonomous

spending, notably in the form of private investment and government outlays. Fiscal policy was crucial; monetary policy was all but irrelevant in the sense that "you cannot push on a string."

Only the University of Chicago, through the teachings of Henry Simons, Lloyd Mints, Frank Knight, and Jacob Viner, had stood resolutely against this pervasive doctrine during the late 1930s and 1940s as Keynesian doctrine swept through the academia. Friedman was well versed in the subtle version of the quantity theory expounded at Chicago, a version in which the quantity theory was connected and integrated with general price theory and became "a flexible and sensitive tool for interpreting movements in aggregate economic activity and for developing relevant policy prescriptions" (Friedman, 1956, p. 3).

Systematically, over the period 1950–80, Friedman and his research associates would challenge the empirical relevance of the Keynesian model by demonstrating the empirical superiority of the quantity theory as expounded at Chicago. By the time his research program was complete, and prior to the rational expectations revolution, almost all economists would recognize that money did matter, and that what happened to the quantity of money had important effects on economic activity in the short run and on the price level in the long run (Friedman and Friedman, 1998, p. 228).

Before Keynes, the quantity theory of money had played an important role in classical economics. Using the behavioral equation $MV = PY$, classical theorists had argued that the income velocity of circulation of money, V, was a constant; that real income, Y, was unaffected by changes in the quantity of money (the so-called classical dichotomy); and therefore that changes in the supply of money, M, directly affected the price level, P. Keynes (1936) derided this naïve textbook version of the quantity theory, arguing instead that V was not a constant but was highly variable and that it served as a cushion to prevent any change in the supply of money from exerting an impact on either real income or the level of prices.

In conjunction with his work at the NBER, Friedman established a workshop in Money and Banking at the University of Chicago. The first product of this workshop was a book: *Studies in the Quantity Theory of Money* (1956), which Friedman edited. In retrospect, this publication was the first major step in a counter-revolution that succeeded in restoring the quantity theory to academic respectability. There is no evidence that Friedman was aware at that time of the dimensions of the impending battle. His express intent in writing the introductory essay was simply to "set down a particular 'model' of a quantity theory in an attempt to convey the flavor of the (Chicago) oral tradition" (Friedman, 1956, p. 4). Of course, the impact of his essay would be much more dramatic than he and his colleagues at that time could possibly foresee.

Friedman's introductory essay provided a subtle and sophisticated restatement of the quantity theory of money as a stable money–demand function (Breit and Ransom, 1998, p. 228). Unlike the classical economists, Friedman rejected the notion that V, the income velocity of circulation of money, was a constant. Instead, he modeled V as a stable function of several variables, since money was an asset, one way of holding wealth. Within this framework, he posited that V would respond

to nominal monetary expansion in the short run by accentuating rather than by cushioning the impact of such expansion on nominal income. This restatement became recognized as the theoretical position of the Chicago School on monetary economics.

The four empirical studies in the book—dealing with inflationary and hyperinflationary experiences in Europe and the United States—provided support for the quantity theory in its restated form by demonstrating a striking regularity in economic responses to monetary changes. The most significant finding was that velocity was a stable function of permanent income. Since money is a luxury good, the demand for which rises as income increases, velocity would tend to decline over time as income rose. The monetary authority therefore must increase the stock of money to offset this decline in velocity, if it wished to maintain price stability (Breit and Ransom, 1998, 230).

These results met with skepticism from Keynesian economists who counterclaimed that the supply of money merely accommodated demand and did not impact independently on the macroeconomy. It would take Friedman and his colleagues the better part of a decade of high-quality theoretical and empirical analysis to mount a persuasive case for the quantity theory.

One important component of this research program was the comparative test (Friedman and Meiselman, 1963), in which a simple version of the income—expenditure theory, $C = a + kA$ was compared with a simple version of the quantity theory, $C = b + vM$. For the period 1897 to 1958, using annual data, and for a shorter periods using quarterly data, the quantity theory performed better than the income–expenditure theory, implying that v was more stable than k, except for the period of the Great Depression.

More influential, ultimately, was the monumental book coauthored with Anna Schwartz, *A Monetary History of the United States, 1867–1960* (Friedman and Schwartz, 1963). This monumental piece of empirical research offered substantial support for the restated quantity theory and sent shock waves through the economics profession by explaining the Great Depression in terms of the failure of the Federal Reserve to deploy effective open-market operations that would have prevented the banking crisis that brought about a significant decline in the supply of money (Breit and Ransom, 1998, p. 239).

Subsequent research by Friedman determined (1) that the impact of a fiscal deficit on nominal income was short lived whereas, after a lag, an increased rate of growth of the nominal money supply permanently augmented the rate of price inflation; (2) that the adjustment of nominal income to an increased rate of monetary growth occurred with a long and variable lag; (3) that in the long run additional monetary growth affected only the rate of inflation and exerted virtually no effect on the level or rate of growth of real output (Walters, 1987, 425).

So successful was Friedman's empirical work in supporting the quantity theory that economists began to clamor for an explicit theory of the role of money in income determination, a theory capable of generating the propositions supported by the empirical investigations. In response, Friedman published two strictly theoretical articles (Friedman, 1970, 1971) that sparked critical reviews from leading

Keynesian scholars. The debate between the Keynesians and the quantity theorists would continue for another decade before the worldwide stagflation of the 1970s brought a close to decisive victory for Friedman's position.

The restoration of the quantity theory undoubtedly weakened the reliance of governments on fiscal policy as a means of countering the business cycle. This alone was a major contribution to classical liberalism, weakening as it did the justification for government macroeconomic intervention through fiscal policy. However, Friedman would fail to persuade the economics profession and the wider public that monetary policy also should be eschewed in favor of a nondiscretionary rate of increase in the nominal money supply at the underlying rate of growth of productivity. This was unfortunate because governments that wished to use the inflation tax to evade the real debt implications of deficit financing now knew just how to go about their business. Although Friedman was very slow to recognize it, failure in this regard reflected more the pressures of public choice than any weakness in Friedman's research on the long and variable lags in the relationship between changes in the nominal supply of money and changes in the behavior of nominal income (Rowley, 1999b, p. 419). The Federal Reserve Board and its influential staff in the United States and central bank systems elsewhere would not easily be dislodged from playing an active role in monetary policy.

Failure was also, in part, the consequence of Friedman's success in promoting free markets. Deregulation of the banking system made it difficult from the early 1980s onwards to determine just which M should be subjected to the non-discretionary rule. Perhaps most important, however, was Friedman's neglect (typical of the Chicago School) of any detailed institutional analysis of the banking sector. In the absence of such an analytical framework, the call for nondiscretionary policy could too easily be categorized as dogma rather than as science.

Fundamentally, of course, the case in favor of the nondiscretionary rule collapsed during the 1980s once it became apparent that the demand for money was unstable in the wake of banking deregulations.

The Fallacy of the Phillips Curve

An important component of the Keynesian orthodoxy during the 1960s was the notion that there existed a stable negative relationship between the level of unemployment and the rate of price inflation. This relationship was characterized as the *Phillips curve* in recognition of the celebrated 1958 paper by A.W. Phillips that plotted unemployment rates against the rates of change of money wages and found a significant statistical relationship between the two variables.

Keynesian economists had focused on this apparent relationship to persuade government that there existed a permanent trade-off between price inflation and unemployment, allowing choices to be made between alternative rates of unemployment and alternative rates of price inflation. In such circumstances, by accepting a

modest increase in price inflation, politicians, if they so wished, could lower the rate of unemployment in an economy.

Friedman had questioned the validity of the Phillips curve in the early 1960s, but without any significant intellectual impact. In his Presidential Address to the American Economic Association in December 1967 (Friedman, 1968), Friedman was able to raise the tone of this questioning, arguing convincingly that the concept of the stable Phillips curve was an illusion and that any trade-off that existed between the rate of inflation and the rate of unemployment was strictly temporary in nature. Once again, Friedman placed himself directly against the thrust of Keynesian doctrine, deconstructing it from the perspective of Marshallian economics (De Vroey, 2001).

Keynes had rendered money non-neutral and had made fiscal policy potent in its effects on output by withdrawing one equation (the labor supply schedule) and one variable (money wages) from the classical model (Sargent, 1987, p. 6). The Keynesian model was thus short by one equation and one variable by comparison with the classical model. To close that gap, the Keynesians had incorporated the Phillips curve as a structural relationship. In so doing, they had misinterpreted the true nature of labor market equilibrium.

Friedman in his 1967 Address reasserted the classical assumption that markets clear and that agents' decision rules are homogeneous of degree zero in prices. When agents confront intertemporal choice problems, the relevant price vector includes not only current prices but also expectations about future prices. This, the proponents of the stable Phillips curve had failed to recognize.

The trade-off between inflation and unemployment captured in the Phillips curve regression equations represented the outcomes of experiments that had induced forecast errors in private agents' views about prices. If the experiment under review was a sustained and fully anticipated inflation, Friedman asserted, there would exist no trade-off between inflation and unemployment. The Phillips curve would be vertical and the classical dichotomy would hold.

Friedman in his 1967 paper utilized a version of adaptive expectations to demonstrate that any trade-off between inflation and unemployment would be strictly temporary and would result solely from unanticipated changes in the inflation rate. The natural rate of unemployment, defined essentially in terms of the "normal equilibrium" of Marshall rather than in the Walrasian terms of the subsequent rational expectations school (De Vroey, 2001, 130), was a function of real forces. If monetary expansion fools the workers temporarily so that they do not recognize that their real wage has been lowered, it might stimulate a temporary reduction in the level of unemployment below the "normal equilibrium" (or natural rate). As soon as the money illusion dissipates, unemployment will drift back to the natural rate. To keep unemployment below the natural rate requires an ever-accelerating rate of inflation.

On the basis of this logic, Friedman predicted that the apparent Phillips curve trade-off evident in the data from the 1950s and 1960s would disappear once governments systematically attempted to exploit it. In the 1970s, the Phillips curve trade-off vanished from the data. Indeed, estimated Phillips curves

became positive as rising rates of inflation began to coincide with rising rates of unemployment.

Once again Friedman's positive economic analysis paved the way for a reduction in the extent of government economic intervention, now through monetary policy. Economic events would ultimately invalidate the Phillips curve hypothesis. However, by directing the attention of economists to model mis-specification, Friedman hastened the process, further weakening the economic case for government intervention in the macroeconomy.

The Reason of Rules

Friedman's views on monetary policy were greatly influenced by Henry Simons' teachings on the superiority of rules over discretionary policy (Breit and Ransom, 1998, p. 241). From the outset of his career, but with increased vigor following his empirical work on the quantity theory of money, not least his analysis of the Great Contraction (Friedman and Schwartz, 1963), Friedman argued in favor of committing macroeconomic policy to a series of monetary and fiscal rules designed to reduce the degree of discretionary power available to government agents. It should be noted, however, that this argument was not based on any knowledge of public choice. Rather, Friedman was concerned that central banks typically failed to predict the pattern of the business cycle and the distributed lags of monetary intervention, thereby destabilizing the macroeconomy.

Friedman's advocacy of rules stemmed from the recognition that monetary policy could not peg interest rates, could not generate full employment, and could not stabilize cyclical fluctuations in income (Butler, 1985, p. 177). Yet, monetary policy had a considerable power for mischief, since it affected every part of the economy. Therefore, it deserved great respect. In particular, because changes in the supply of money exerted an impact on the macroeconomy only with long and variable lags, the potential for destabilizing policy intervention was high even at the hands of a benevolent government.

At different times, Friedman advocated two comprehensive and simple plans for coordinating monetary and fiscal policies. In 1948, he advocated an automatic adjustment mechanism that would overcome the problem of the lag and that would be more likely to move the economy in the right direction than would discretionary monetary policy.

Friedman advocated (1) the imposition of 100% reserve requirements on the banks, making the supply of money equal to the monetary base and (2) a prohibition on government placing interest-bearing debt with the public. The Federal Reserve would be required to monetize all interest-bearing government debt, so government deficits would lead to increases in the monetary base, and government surpluses would lead to reductions in that base. Such a mechanism would act as an automatic stabilizer and would also assign a clear responsibility for growth in the money supply (and in inflation) to its primary determinant, the federal deficit (Sargent, 1987, p. 9).

If implemented, Friedman's proposed rule would have eliminated much of the discretionary power that enabled governments to implement Keynesian macroeconomic policy. For that reason alone, it was doomed during the era of Keynesian hegemony. In addition, it implied the abolition of the central banking institutions that determine the course of monetary policy. Such powerful pillars of the economic establishment would not easily surrender their power and wealth by stepping down in favor of an automatic rules-based system.

By 1960, Friedman pragmatically recognized that central banks, open market operations, and fractional reserve banking were here to stay, at least for the foreseeable future. In such circumstances, he advanced an alternative rules-based mechanism that was in some respects quite contradictory to his earlier, preferred ideal. Its essential element would be a legislated monetary rule designed to ensure the smooth and regular expansion of the quantity of money.

According to this mechanism, the Federal Reserve would be required by statute to follow a rule of increasing high-powered money by a constant $k\%$ per annum, where k was a small number designed to accommodate productivity growth in the economy. This rule would permanently limit the fiscal authorities' access to the printing press to the stipulated $k\%$ increase and would force them to finance current deficits only by credibly promising future surpluses (Sargent, 1987, p. 9).

Cyclical movements in real income would not be avoided by this nondiscretionary mechanism. However, the nondiscretionary nature of the rule would prevent some of the wilder swings induced by inept and ill-timed monetary measures (Butler, 1985, p. 185).

So far, this advocacy has failed, not least because of public choice pressures combined with some skepticism as to the importance of high-powered money as the key monetary variable. Nevertheless, Friedman's advocacy has not been in vain. Monetary authorities in the United States and elsewhere are now aware of the relationship between the quantity of money and the price level. Throughout the world, there is far more reliance on monetary restraint as the basis for price stability than was the case during the Keynesian era. Such monetary restraint has increased the political costs of fiscal expansion. Once again, a largely positive program of economic analysis has served the cause of liberty well.

On Liberty

Individuals are born with free will, and, if they so choose, they are able to forge judgments that are conditioned neither by their particular circumstances nor by the environment in which they find themselves. Nevertheless, particular circumstances and environments influence judgments even though, ultimately, they do not shape them. Milton Friedman's views on the nature and importance of liberty surely were influenced by his particular circumstances and environment.

Friedman was a second-generation central European immigrant and a Jew, characteristics that were not viewed favorably in the United States during the first

half of the twentieth century; characteristics, indeed, that attracted hostile discrimination from public bodies and their agents, themselves protected from the discipline of the competitive market place. Growing up in such circumstances demonstrated to Friedman in a very personal way the powerful and even-handed protection against prejudice provided by the capitalist system.

Much of Friedman's scholarly career has been played out against the international backcloth of unequivocal political evil, in the form of totalitarian fascist nightmares epitomized by Adolf Hitler's Third Reich and in the form of totalitarian socialist nightmares, epitomized by Josef Stalin's USSR. The "days of the devils" (Johnson, 1983) may now be largely over. However, their evil mark is printed indelibly on everything that Friedman wrote, said, and did.

Domestically in the United States, Friedman's career played out against a background of all-but monotonic growth in the size of government and in the reach of its interventionist tentacles. Not for him the privilege of nineteenth century British classical liberals who lived out their lives in environments that largely matched their philosophical beliefs. Circumstances and environments combined, in Friedman's case, to demand an aggressive classical liberalism, designed to roll back tyranny as well as to preserve and to protect established liberties. That demand called forth an unwavering supply.

Friedman outlined his special brand of classical liberalism very clearly in the introductory paragraphs of *Capitalism and Freedom*:

> The free man will ask neither what his country can do for him nor what he can do for his country. He will ask rather "What can I and my compatriots do through government" to help us discharge our individual responsibilities, to achieve our several goals and purposes, and above all, to protect our freedom? And he will accompany this question with another: How can we keep the government we create from becoming a Frankenstein that will destroy the very freedom we establish it to protect? Freedom is a rare and delicate plant. Our minds tell us, and history confirms, that the great threat to freedom is the concentration of power. Government is necessary to preserve our freedom, it is an instrument through which we can exercise our freedom; yet by concentrating power in political hands, it is also a threat to freedom. (Friedman, 1962, p. 2)

In three important books—*Capitalism and Freedom* (Friedman, 1962), *Free to Choose* (Friedman and Friedman, 1979) and *Tyranny of the Status Quo* (Friedman and Friedman, 1983)—as well as in many other essays (see Leube, 1987 for a representative selection) and in numerous *Newsweek* columns Friedman outlined a view of classical liberalism closely related to the philosophy of the young John Stuart Mill.

Friedman's philosophy, like that of Mill, was one in which freedom is viewed as indivisible, with economic freedoms equally as important as political freedoms. Like Mill, Friedman also held that government should be as decentralized as possible in order to allow alienated citizens to vote with their feet. Like Mill, Friedman also held that "the only purpose for which power can be rightfully exercised over any member of a civilized community against his will, is to prevent harm to others" (Mill, 1865, p. 6). It is a philosophy like that of Mill in which "[o]ver himself, over his own body and mind, the individual is sovereign" (Mill, 1865, p. 6).

This said, Friedman did not believe that most debates over economic policy are debates over value judgments (Friedman, 1967). Disagreements exist, for the most part because economists accept differing tentative hypotheses about the relationship between economic phenomena. Friedman maintained an optimistic perspective that most of these disagreements would disappear over time as competing hypotheses were subjected to empirical testing. He qualified this optimism, however, with the passage of time, recognizing in the wake of the public choice revolution, that economists and policy makers are not always driven by considerations of high moral purpose (Friedman, 1986).

In Friedman's normative ideal, government should be strong and yet severely constrained. The major function of government is to protect the freedom of the people from outside and from inside intervention (i.e., to protect negative freedom in the sense most clearly defined by Isaiah Berlin, 1969). To achieve this objective, government must be empowered to provide an effective system of defense and to provide internally for the protection of property rights, the enforcement of private contracts and the maintenance of competitive markets. These powers, however, should not be unlimited. Government itself should remain strictly subject to the rule of law.

Unlike modern anarcho-capitalists, Friedman did not believe that private forces are capable of effectively providing these indispensable prerequisites of the free society. Nor was he comfortable with restricting government to the functions of the minimal (or night watchman) state. Although he expressed a strong preference in favor of voluntary cooperation and private enterprise, he also recognized that government upon occasion may enable individuals to accomplish jointly arrangements that would be more difficult or more expensive for them to accomplish severally (Friedman, 1962, p. 2). In particular, Friedman was sensitive to the problem of poverty and argued in *Capitalism and Freedom* (1962) in favor of a negative income tax to set a limit below which no family income could fall.

In contemplating such arrangements, however, Friedman unequivocally focused on the harmful consequences of institutional arrangements that shield individuals from taking personal responsibility for their own decisions or that reflect paternalistic value judgments imposed by philosopher kings on their fellow citizens. He was driven in this presumption by a recognition that the great advances in civilization have never come from centralized government; that centralized government can never duplicate the variety and diversity of individual action; and that centralized government always substitutes uniform mediocrity for the variety that is essential for successful entrepreneurial innovation (Friedman, 1962, p. 4).

A basic human value that underpinned Friedman's philosophy is tolerance based on humility (Friedman, 1991). An individual has no right to coerce someone else, not only because of fundamental principles of classical liberalism, but also because no individual can be sure that he is right and the other person wrong. In this respect, Friedman set himself aside from Utopian classical liberals such as Ludwig von Mises (Mises, 1963) who protect their arguments from empirical criticism on a priori grounds. Friedman rejected *praxeology* of the kind advanced by Mises on the ground that it converts a body of substantive conclusions into a religion.

Friedman argued that democracy is the appropriate form of government to foster political freedom. However, the prerequisite to democracy is a capitalist economy that separates economic from political power, allowing the one to offset the other (Breit and Ransom, 1998, p. 257). In this regard, Friedman failed to take full account of public choice arguments that there is a predictable tension between democracy and free markets in the absence of self-enforcing constitutional constraints on government authority (Rowley, 1999a).

Much of Friedman's normative message is now commonplace in the debate over public policy. For example, several experimentations in the use of school vouchers are currently under way in the United States and many more are under consideration. The Great Society programs of the 1960s that attempted to provide a welfare state from cradle to the grave are systematically, if slowly, being dismantled in favor of market-based alternatives. Affirmative-action policies that rely on bureaucratic controls rather than on competitive capitalism are increasingly the subject of criticism, even among those for whom those policies were ostensibly designed. Conscription in the military has given way to the market-based volunteer force. Fixed exchange rate regimes systematically have given way to flexible exchange rate regimes throughout the large majority of the Free World. In all these areas, Friedman's once controversial ideas have begun to overwhelm the forces of mercantilism, mirroring the success of Adam Smith two centuries earlier.

It should not be forgotten, however, that Friedman's success was bitterly fought and courageously achieved against powerful forces in a Western world then dedicated to the elimination of individual freedom in favor of democratic socialism. Ranged against Friedman in this regard were eminent members of the economics profession (including Paul Samuelson, Kenneth Arrow, John Kenneth Galbraith, James Tobin, Robert Solow, and Joan Robinson) who consistently demonstrated anti-free market prejudices combined with a high regard for big government and an easy willingness to sacrifice economic freedoms on the altar of Keynesian macroeconomic policies.

When intellectual battles are won and lost, the victor rarely receives his justly earned accolades. Those whose arguments have failed, and who seek continued academic respect, shift their positions and rely on myopia to protect them from the consequences of their earlier mistakes.

Rest assured, however, that those leopards who argued so confidently in the middle years of the twentieth century for the institutional arrangements of democratic socialism would not have changed their spots to the extent that they have in the absence of Friedman's firm and convincing voice in defense of economic freedom, a voice that penetrated the citadels of coercion in the West as well as in the East, a voice that gave hope for a freer and more prosperous future during a dangerous half century for those who cherish freedom. Without that clear and convincing voice in favor of capitalism, it is doubtful whether the public choice revolution would have made the inroads that it has into the widely held postwar misconception that government is the omniscient and impartial servant of the public good.

Milton Friedman's death, on November 16, 2006, was a day of mourning for all those worldwide who support individual liberty and free markets. However, his life was well- and fruitfully lived in the honorable and always faithful pursuit of those high ideals.

Selected Works

(1945). *Income from Independent Professional Practice*. (With Simon Kuznets). New York: National Bureau of Economic Research.

(1946). *Roofs or Ceilings? The Current Housing Problem*. (With George J. Stigler). Irvington-on-the-Hudson. New York: Foundation for Economic Education.

(1953). *Essays in Positive Economics*. Chicago: University of Chicago Press.

(1956). *Studies in the Quantity Theory of Money*. Chicago: University of Chicago Press.

(1957). *A Theory of the Consumption Function*. Princeton: Princeton University Press.

(1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.

(1963). "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897–1958," in *Stabilization Policies*. New Jersey: Prentice Hall, pp. 165–268. (With Meiselman, D.)

(1963). *A Monetary History of the United States, 1867–1960*. (With Anna J. Schwartz). Princeton: Princeton University Press.

(1966). *Price Theory: A Provisional Text*. (With David I. Fand and Warren J. Gustus). Chicago: Aldine Press.

(1967). "Value judgments in economics," in Sidney Hook (ed.) *Human Values and Economic Policy*. New York: New York University Press, pp. 4–8.

(1978). *Milton Friedman's Monetary Framework: A Debate with His Critics*. (Edited by Robert J. Gordon). Chicago: University of Chicago Press.

(1980). *Free to Choose*. (With Rose D. Friedman). New York: Harcourt Brace Jovanovich.

(1984). *Tyranny of the Status Quo*. (With Rose D. Friedman). San Diego, New York and London: Harcourt Brace Jovanovich.

(1986). "Economists and economic policy." *Economic Inquiry*, 24, 1–10.

(1987). *The Essence of Friedman*. (Edited by Kurt R. Leub, with a Foreword by W. Glenn Campbell and an Introduction by Anna J. Schwartz). Stanford: Hoover Institution Press.

(1991). "Say 'no' to intolerance." *Liberty*, 4(6), 17–20.

(1998). *Two Lucky People*. (With Rose D. Friedman). Chicago: University of Chicago Press.

References

Breit, W. and Ransom, R.L. (1998). *The Academic Scribblers*. Princeton: Princeton University Press.

Buchanan, J.M. and Wagner, R.E. (1977). *Democracy in Deficit: The Political Legacy of Lord Keynes*. New York: Academic Press.

Butler, E. (1985). *Milton Friedman: A Guide to His Economic Thought*. New York: Universe Books.

De Vroey, M. (2001). "Friedman and Lucas on the Phillips curve: From a disequilibrium to an equilibrium approach". *Eastern Economic Journal*, 27(2), 127–148.

Fisher, I. (1930). *The Theory of Interest*. London: Macmillan.

Friedman, M. and Meiselman, D. (1963). "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897–1958", in *Stabilization Policies*. New Jersey: Prentice Hall, pp. 165–268.

Hirsch, A. and de Marchi, N. (1990). *Milton Friedman: Economics in Theory and Practice*. Ann Arbor: University of Michigan Press.

Johnson, P. (1983). *Modern Times: The World from the Twenties to the Eighties*. New York: Harper & Row.

Keynes, J.M. (1936). *The General Theory of Employment, Interest and Money*. New York: Harcourt, Brace and Jovanovich.

Kuznets, S. (1952). "Proportion of capital formation to national product". *American Economic Review*, 42, 507–526.

Mill, J.S. (1865). *On Liberty*. London: Longman.

Mises, Ludwig von (1963). *Human Action*. New Haven: Yale University Press.

Robbins, L.C. (1932). An Essay on the Nature and Significance of Economic Science. London: Macmillan.

Rowley, C.K. (1999a). "Five market friendly nobelists." *The Independent Review*, 3(3), 413–431.

Rowley, C.K. (1999b). "Review of—Milton and Rose Friedman, *two lucky people*". *Public Choice*, 99(3–4), 474–480.

Samuelson, P.A. (1963). "Comment on Ernest Nagel's 'assumptions in economic theory.'" *American Economic Review*, May, 211–219.

Sargent, T.J. (1987). *Some of Milton Friedman's Scientific Contributions to Macroeconomics*. Stanford: Hoover Institution.

Walters, A. (1987). "Milton Friedman: Born 1912," in John Eatwell, Murray Milgate, and Peter Newman (eds.) *The New Palgrave: A Dictionary of Economics*, vol. 2. London: Macmillan, pp. 422–427.

Part III
Perspectives

Chapter 12
The Perspective of the History of Thought[1]

Charles K. Rowley

Introduction

Public choice is a relatively new discipline located at the interface between economics and political science. Its modern founding was the achievement of Duncan Black whose 1948 (a–c) articles are widely viewed as the seminal contributions that launched scholarship in the application of economic analysis into the traditional domain of political science. Yet, it is true that the founding goes back almost two centuries in time, to the late eighteenth century contributions of two French *Encyclopedistes*, the Compte de Borda and the Marquis de Condorcet. The two French noblemen shared a conviction that social sciences were amenable to mathematical rigor, and made significant contributions to the theory of voting. These contributions form the foundations on which much of modern public choice has been crafted.

In his pioneering work on elections, Condorcet (1785) sought to 'inquire by mere reasoning, what degree of confidence the judgment of assemblies deserves' (1785, iv). In modern jargon, he posed what is now known as the jury problem or the vote problem. The starting point, well-known to the *Encyclopedistes*, is that majority voting is unambiguously the best voting rule when only two candidates are on stage. How might this rule be extended to three or more candidates? The naive but widely held answer is plurality voting, where each voter casts a vote for one candidate and the candidate with most votes is elected.

Condorcet raised doubts as to the general acceptability of the plurality vote rule. Suppose that 60 voters have opinions about three candidates A, B and C as shown in Table 12.1.

In the illustration, candidate A wins by plurality. Yet, if A is opposed only by B, he loses (25 to 35) and if A is opposed only by C he loses again (23 to 37). Thus the plurality rule does not convey accurately the opinion of the majority.

Table 12.1 Borda's method of marks

	23	19	16	2
Top	A	B	C	C
	C	C	B	A
Bottom	B	A	A	B

Table 12.2 Condorcet's paradox of voting

23	17	2	10	8
A	B	B	C	C
B	C	A	A	B
C	A	C	B	A

Using identical premises, in 1781 Borda had initiated the discussion of voting rules by questioning the effectiveness of the simple majority vote rule and by proposing the method of marks as a more appropriate rule. In this method, each candidate receives 2 points from a voter who places him first, 1 point from a voter who places him second and 0 points from a voter who places him third. Hence, by reference to Table 12.1, C is elected with a score of 78 points. Condorcet, however, in following up on the insight of Borda, sought a different solution.

Condorcet posited a simple binomial model of voter error. In every binary comparison, each voter has a probability $1/2 < p < 1$ of ordering the candidates correctly. Thus the relevant data is contained in the 'majority tournament' that results from taking all pairwise votes: B beats A 35 to 25; C beats A 37 to 23; C beats B 41 to 19. Condorcet proposed that candidates should be ranked according to 'the most probable combination of opinions' (1785, p. 125). In modern terminology, this is a maximum likelihood criterion.

In the above example, the most probable combination is given by the ranking CBA, since this agrees with the greatest total number of votes. Condorcet's ranking criterion implies that an alternative that obtains a majority over every other alternative must be ranked first. Such an alternative, if one exists, is now known as a 'Condorcet winner'. However, as Condorcet established, some configurations of opinions may not possess such a winner because the majority tournament contains a cycle. Such an occurrence is now known as Condorcet's paradox and is illustrated in Table 12.2.

In this illustration, A beats B, 33 to 27; B beats C, 42 to 18; C beats A, 35 to 25. In such circumstances, pairwise voting results in intransitivity. According to Condorcet's maximum likelihood criterion, the cycle should be broken at its weakest point, namely A over B, which yields the ranking of B over C over A. Therefore, in this case, B would be declared the winner.

Condorcet's *Essai* contains other useful insights that now play an important role in public choice. Perhaps the most important is the issue of strategic manipulation, which is hinted at in several places, although it is never systematically explored (Moulin and Young, 1987). For example, on page clxxix of the *Discours Preliminaire*, Condorcet criticizes Borda's method of marks as being vulnerable to

a *cabale*. When confronted with this criticism, Borda was merely moved to comment: 'My scheme is only intended for honorable men' (Rowley, 1987). It has since been established by modern game theory that any configuration of individual opinions that guarantees the existence of a Condorcet winner also defines a strategy proof voting rule. This remains an important argument in favor of Condorcet consistent rules designed to elect the Condorcet winner whenever it exists (Moulin, 1983).

Because the publications by Condorcet and Borda were not widely circulated in the late eighteenth century, because they were somewhat densely written and because they were written in French, their ideas disappeared for some 150 years until they were rediscovered and proselytized by Duncan Black in 1958. Since then, the ideas have strongly influenced public choice theorists and have played a central role in many of the discipline's recent developments.

The Insights of Duncan Black

Ideas that are lost do not constitute any part of the litany of science. Duncan Black essentially rediscovered ideas that had been advanced earlier by the two eighteenth century French noblemen only to be lost, then to be rediscovered late in the nineteenth century (1884) by Charles Dodgson (Lewis Carroll), then to be lost again. Since Black's discovery has not been lost, he must be viewed as the true founder of public choice (Rowley, 1991). The work of Borda, Condorcet, and Dodgson is known today only because Black researched their writings and made them available to his own generation of scholars (Grofman, 1987).

Duncan Black's vision as a young economist was that of developing a pure science of politics that would place political science on the same kind of theoretical footing as economics. All of his work was underpinned by the deceptively simple insight of modeling political phenomena 'in terms of the preferences of a given set of individuals in relation to a given set of motions, the same motions appearing on the preference schedule of each individual' (Black, 1972, p. 3).

In this search, Black rediscovered Condorcet's paradox of cyclical majorities (Black, 1948a, pp. 32–33) and thereby opened up an extremely fruitful avenue of public choice research. It is important to acknowledge Black's achievement because recognition for the rediscovery of the Condorcet paradox is frequently and incorrectly given to Kenneth Arrow. Black (1948a, b, c) raised a number of important questions and offered some preliminary answers related to this paradox (Grofman, 1981). The first question asks whether the paradox is inevitable; the second asks how frequently the paradox can be expected to occur; the third asks how easy it is to detect a paradox from the available evidence on majority rule outcomes; and the fourth asks how large a cycle will be.

Black's answer to the first question was that the paradox is not inevitable. Embedded in this answer is the famous median voter theorem that will be outlined and evaluated later in this section. In answering the second question, Black focused

attention on the special case of three voters and three alternatives for what is now known as the 'impartial culture' i.e., a committee in which all strong preference orderings are equally likely.

Black recognized the wider significance of this question. He suggested that the likelihood of the paradox of cyclical majorities occurring increased rapidly as the number of motions under consideration and the number of committee members increased (Black, 1958, p. 51). In this judgment he has been proved correct by subsequent analysis (Grofman, 1981, p. 15).

In answering the third question, how easy it is to detect a paradox, Black provided two useful results. The first result is that under standard amendment procedures, given sincere voting, a voting paradox will always be revealed if there are as many rounds of voting as there are alternatives, less one. The second result is the theorem that the voting paradox is always revealed if data is available on all paired comparisons. This is a powerful result since Black also shows that if a majority winner exists, no voter has an incentive to vote insincerely in such a complete balloting.

The fourth question, querying how many alternatives are likely to be located in a *top cycle*, was not directly addressed by Black. However, he did provide a number of insights on interrelationships between cycles. For example, he noted that if two intersecting cycles have one motion in common, it must be possible to form a cycle that includes all the motions of both cycles (Black, 1958, p. 48). He also examined the case of three nonintersecting cycles (where every motion in the first defeats every motion in the second, and where every motion in the second defeats every motion in the third). He demonstrated, in such circumstances, that every motion in the third may still defeat every motion in the first (Black, 1958, p. 50). As subsequent analysis has confirmed, winning cycles are likely to include all alternatives (McKelvey, 1976).

Black's answer to the first question, concerning the inevitability of the paradox of the voting cycle, has been left to the end because it is his most important legacy to public choice. His insight came in February 1942, while 'fire-watching' in case of air raids, late at night in the magnificent green drawing room of Warwick Castle (Black, 1972, p. 4). While playing with diagrams that represented motions as points on a line and with preferences represented as single-peaked utility functions, Black saw 'in a shock of recognition' (*ibid.*) the property of the median optimum, or what we now refer to as the median voter theorem.

The idea of single-peakedness can be defined in a number of different ways. Black provided a graphical interpretation that is illustrated in Fig. 12.1. A set of preference schedules is said to be single-peaked if there occurs an ordering of the alternative motions such that the preference schedules of all committee members can be graphed as single-peaked curves (i.e., as curves that change direction at most once, up or down). Where this condition holds, Black established that a unique alternative exists capable of attracting a simple majority in pairwise competition against all other alternatives. This majority winner is the most preferred alternative of the median voter. Hence, for single-peaked preferences, Black established that there is a stable majority choice at the median of the voter distribution. Furthermore, under this condition, majority rule gives rise to a transitive ordering of alternatives.
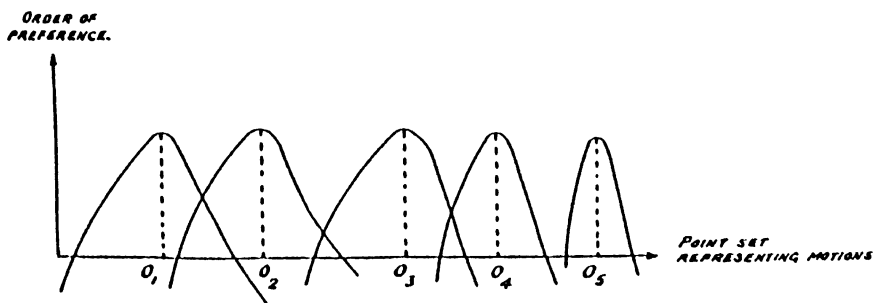
Fig. 12.1 Duncan Black's median voter model

In Fig. 12.1, the median outcome is at point O_3 in policy issue space. It is important to note that Black's theorem is restricted to the case (illustrated in Fig. 12.1) where policy issue space is defined in terms of a single dimension. As we now know (Black, 1958), where the median voter theorem holds, majority rule always selects a Condorcet winner.

The Insight of Kenneth J. Arrow

Although Duncan Black's 1948 article is best known for its derivation of the median voter theorem, Black was clearly aware of the potential for cycling, should the condition of single-peaked preferences fail to hold. In 1949, Black and Newing attempted to define the necessary conditions for the existence of a stable voting equilibrium in multi-dimensional space, focusing on the three-person case. In this contribution they clearly anticipated the contributions of Kenneth Arrow. Their paper was submitted to *Econometrica* in November 1949. The referee reported favorably on the paper.

In a stroke of misfortune, *Econometrica* delayed by some 18 months in reaching a decision on the paper. When it did so, the Managing Editor, Ragnar Frisch, informed the authors that he would recommend the article for publication 'if the interrelation-ships with Arrow's recent monograph could be brought out clearly throughout the paper' (Coase, 1981). Arrow's 1950 article and his 1951 monograph apparently had pre-empted the Black and Newing article as a consequence of inexcusable editorial delay. Black and Newing withdrew the article from *Econometrica* and published it in 1951 in a little-read booklet. By such chance events are Nobel Prizes sometimes won and lost (Rowley, 1991).

In any event, Arrow's 1951 book has exerted a significant impact on the evolution of public choice, even though its primary concern was normative rather than positive in nature, focusing as it did on the desirable characteristics of alternative mechanisms of social choice. This impact stems from Arrow's redis-covery of Condorcet's paradox of cyclical fluctuations.

Arrow (1950, 1951) responded to the apparent collapse during the 1930s of Benthamite utilitarianism as economists systematically retreated from the notion that utility is measurable on a cardinal scale and comparable across individuals. If the weak Pareto principle is all that remains of the once mighty utilitarian doctrine, what are the normative implications for the available mechanisms of effecting social choices? In his famous impossibility theorem, Arrow proved that any social welfare function involving at least three individuals choosing over at least three alternatives must violate at least one of six reasonable axioms of social choice, namely rationality, unbounded domain, the Pareto principle, non-dictatorship, non-imposition and independence of irrelevant alternatives.

Most important, from the perspective of public choice, was Arrow's proof that a social welfare function based on majority rule has the unsatisfactory property of being intransitive when at least three individuals vote over at least three alternatives, even when the preferences of each person are strictly transitive. Arrow did not infer that majority rule would always produce cycles in such circumstances. Given the assumption of an unbounded domain, it sufficed for him to demonstrate that certain configurations of individual preferences would result in the Condorcet paradox.

Although this insight is not original to Arrow, nevertheless, it is he who has gained recognition for it. Undoubtedly, Arrow's emphasis on the instability of majority rule contrasts sharply with Black's emphasis on the stability of the median voter outcome. Since these two impulses still course strongly through much of public choice analyses of the vote motive, it is convenient, if not strictly accurate, to distinguish them by reference to the two scholars.

The Insight of Anthony Downs

Both Black and Arrow analyzed the majority vote mechanism in abstract terms, deliberately seeking generality at the cost of sacrificing institutional detail. Although their contributions, especially those of Arrow, sparked an almost obsessive interest among students of social choice, perhaps because of their abstractness, they failed to make much initial inroad into political economy and political science.

In 1957, Anthony Downs filled this institutional vacuum with a book entitled *An Economic Theory of Democracy* that would become a fulcrum for public choice analysis. Downs was a student of Kenneth Arrow whose work on social choice theory clearly motivated his contributions. Surprisingly, Downs displayed no knowledge of Black's contributions despite Arrow's evident acquaintance with them. Ironically, despite the fact that most public choice scholars identify Downs with the median voter theorem, the theorem is referred to nowhere in the book.

Rather Downs adapted the spatial economic model of Harold Hotelling (1929) to demonstrate that competition between two political parties under conditions of parliamentary democracy often results in both parties converging in policy issue space to adopt identical platforms that reflect the preferences of a majority of the

electorate. Since Downs depicted normally distributed voter preference distributions, there is no way in his analysis of distinguishing between the mean, the median and the mode as the relevant point of party convergence.

The real contribution of Downs was not the median voter theorem (unequivocally the insight of Black) but rather the introduction of the rational choice approach to the study of political science. Pitting himself against the well-entrenched tradition of behavioral analysis among political scientists, Downs laid the foundations for a major research program that would apply rational choice theory to every aspect of the political market place.

By rational action, Downs meant action that is efficiently designed to achieve the consciously selected political and/or economic ends of every actor in the political market place. From this perspective he developed an economic theory of democracy designed to understand and to predict political behavior within an environment of two-party representative democracy.

From the self-interest axiom sprang Down's view of what motivates the political actions of party members. They act 'solely in order to attain the income, prestige, and power which comes from being in office' (Downs, 1957, p. 28). Politicians, in this model, never seek office as a means of promoting particular policies. Their only goal is to reap the rewards of holding office. The fundamental hypothesis of Down's model is that 'parties formulate policies in order to win elections, rather than win elections to formulate policies' (Downs, 1957, p. 28). Thus, the application of the self-interest axiom leads Downs to the hypothesis of vote-maximizing politicians.

Downs also applied the self-interest axiom to voter behavior, hypothesizing that each citizen casts his vote for the party that he expects to provide him with the most benefits. As Downs recognized, the concept of rational voting is deceptively complex, ambiguous and, hence, deserving of close scrutiny. The benefits that voters consider in making their decisions are streams of utility (referred to as utility income) derived from government activity.

Not all utility income is relevant to the vote decision, since utility income includes benefits that the recipient does not realize that he will receive and also benefits that he is aware of without knowing their exact source. However, only benefits of which rational voters are conscious at the time of the election can influence their voting decisions.

The unit of time over which voters evaluate utility income flows is the election period, defined as the time elapsing between elections. At least two such election periods enter into the calculus of the rational voter, namely, the period ending at the time of the election and the period following that election. Both periods are relevant to his determination of the expected party differential in utility income, the measure that will determine which party will secure his vote.

In placing his vote, the voter is helping to select the government that will govern him during the coming election period. His rational decision must reflect the expected future performances of the competing parties. Yet, he knows that political parties are neither obligated to honor nor always capable of carrying out their platform commitments.

In such circumstances, the most recent election period experience of the party in power is the best possible guide to its future behavior, assuming that its policies have some continuity. This performance must be weighed against the performance the opposition would have produced had it been in power. Downs asserted that it is rational for the voter to ground his voting decision primarily on current events, while applying two future-orienting modifiers to his current party differential.

The first modifier is the trend factor, an adjustment made by each citizen to his current party differential to account for relevant trends in the behavior of the government during the current election period. The second modifier is the tie breaker adjustment utilized only when the voter cannot distinguish between the parties. In such circumstances, voters cast their votes by comparing the performance of the incumbent government with that of its immediate predecessor. Voters who still cannot distinguish between the competing parties rationally abstain from voting.

Because Downs was not aware of the median voter theorem, his discussion of the basic logic of government decision-making was less precise than it might have been. In general, he suggested that vote-maximizing incumbents will follow the majority principle, subjecting each decision to a hypothetical poll and always choosing the alternative that the majority of voters prefer. He recognized that such a strategy would not guarantee victory in every election.

The opposition party might defeat a majority-pleasing government by adopting one of three possible strategies. The first such strategy is the adoption of a program identical in every detail with that of the incumbent. Such a strategy forces the electorate to decide their vote by comparing the performance of the incumbent with those of previous governments. Only rarely would such a strategy be effective.

The second such strategy is that of opposing the incumbent by supporting minority positions on carefully selected issues, building a coalition of minorities into a majority vote for the next election. Such a strategy can succeed only where the preferences of those in the minority are more intensely held than the preferences of those in the majority, i.e., where consensus is weak. In the case of passionate majorities, a sufficiently large coalition of minorities will not emerge.

The third such strategy is available to an opposition once again only when there is a lack of consensus in the electorate. In this case, the lack of consensus takes the form of the Condorcet paradox of cyclical majorities. In such circumstances, any alternative that the government chooses can be defeated in a paired election by some other alternative. As long as the government must choose first, and must remain committed to this choice, a prescient opposition can always defeat it.

Downs correctly recognized that his model appears to disintegrate at this point because of the false assumption of certainty. In reality, political parties do not fully know what voters prefer and voters do not fully know the consequences of government acts. If uncertainty is introduced into the model, the incumbents are saved from almost inevitable defeat at each succeeding election, but appear also to be freed from the grip of the majority principle. Therefore, Downs devoted a major part of his book to the effects of uncertainty on the behavior of political markets.

According to Downs, uncertainty divides voters into groups endowed with varying degrees of confidence in their voting decisions. Those who feel least well-informed

are vulnerable to persuasion by voters who are well-informed and who provide correct but biased information favorable to their own causes. Interest groups that want government to adopt policies favorable to their causes pose as representatives of the popular will, simultaneously creating public opinion supportive of their views and convincing government that such public opinion exists. Political parties, once they have formed their policies, endeavor to extend electoral support for those policies. Uncertainty thus forces rational governments to regard some voters as more important than others. By so doing, it modifies the equality of influence that universal suffrage was designed to ensure.

Uncertainty limits the ability of the voter to relate every act of the government to his own view of good policy. The rational voter, in such circumstances, may lower his information costs by identifying party ideologies as a substitute for detailed policy platforms. Each political party in turn will develop an ideology consistent with its policy actions as a short cut to gaining votes. According to Downs, all parties are forced by competition to be relatively honest and responsible in regard both to policies and ideologies.

From this economic perspective, Downs utilized the theory of spatial competition invented by Harold Hotelling (1929), as refined by Arthur Smithies (1941), to elaborate a theory of political party competition under conditions of representative democracy. His version of Hotelling's spatial market consisted of a linear scale running from zero to 100 in the usual left-right fashion. He assumed that all voters would agree on the ordering of both parties across this single dimensional left-right space, essentially by reference to the projected ideologies of the parties.

Downs further assumed that every voter's preferences are single-peaked over this left-right issue space implying that each voter always prefers a position closer to his ideal point over one that is further away and that he always votes for the political party that is closer to his ideal point. If these conditions hold, and if all voters always vote, the two parties will converge at the center of the voter preference distribution in order to maximize their respective votes. Figure 12.2 illustrates this outcome with both parties converging at point 50 in left-right space.
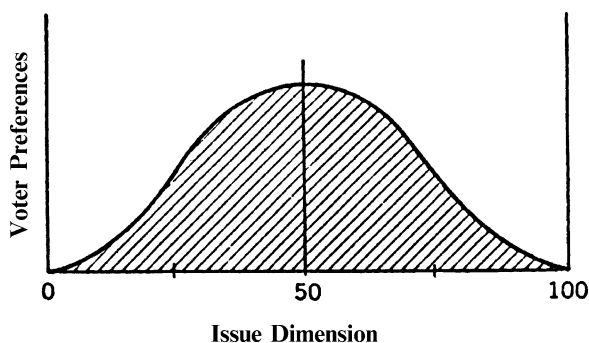
Fig. 12.2 Downs' model of convergence in spatial politics
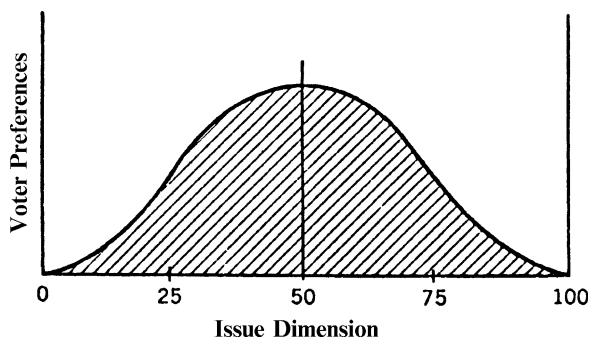
Fig. 12.3 Downs' model of spatial politics with voter abstentions
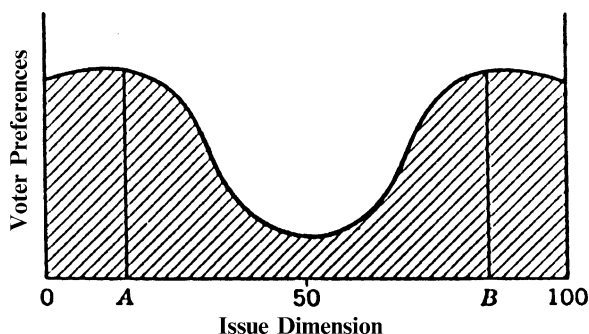
Fig. 12.4 Downs' model of spatial politics with a bimodal distribution of voter preferences

However, if voters located at the two extremes of left-right space become alie- nated as the political parties move towards the center their threats to abstain may halt this process of convergence well short of the center of the distribution. In such circumstances the ideologies of the two parties may differ sharply and political consensus may not emerge. Figure 12.3 illustrates this outcome with the two parties ending up located respectively at points 25 and 75 in issue dimension space.

If the condition of single-peaked preferences does not hold, and the distribution of voters across left-right issue space is bimodal, with modes located near each extreme, parties will locate themselves in proximity with their respective modes. The victorious party will implement policies radically opposed by the opposition. In such circumstances, Downs predicted that government policy will be unstable and that democracy may induce chaos, leading perhaps to its replacement by some form of autocracy. Figure 12.4 illustrates such an outcome with the two parties located respectively at points A and B across left-right space.

In the view of Downs, multi-party systems are likely to occur whenever the distribution of voters over issue space is multi-peaked or polymodal. In such circumstances, the Hotelling model (1929) is likely to hold with political parties
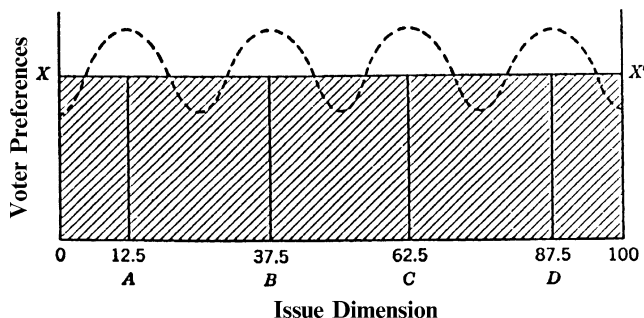
Fig. 12.5 Spatial politics with a multi-modal distribution of voter preferences

maneuvering in left-right space until the distance between each party and its immediately adjacent neighbors is the same for all parties. Figure 12.5 illustrates this outcome with the four competing parties located respectively at points A, B, C and D in the issue space.

Downs focused particular attention on the causes and effects of the rational voter abstention, recognizing that many citizens who are eligible to vote in democratic elections fail to do so. Downs assumed that a citizen's reward for voting is the value of his vote, i.e., his party differential discounted to allow for the influence of other voters upon the election's outcome. On this basis, he demonstrated that when voting is without cost, every citizen who prefers one party over the other votes and every citizen who is indifferent to both the parties abstains.

In reality, voting is always costly, because every act takes time. The cost of voting may outweigh the returns, even for citizens who prefer one party to the other. Indeed, because the expected returns of voting are often miniscule, even low voting costs may result in rational abstentions for citizens who view voting in elections solely from an instrumental perspective.

The importance of rational abstentions depends on its impact on political power. This impact in turn stems from two potential biases. The first potential impact arises from the distribution of the ability to bear the costs of voting. If the cost of voting consists primarily of poll taxes, loss of earnings and transportation costs, upper income citizens obtain a political advantage since the ability to bear such costs typically varies inversely with income. If the cost of voting primarily is the loss of leisure, no income-correlated disparity exists.

The second potential impact arises from biases in the distribution of high returns. The total return each citizen receives from voting depends on (1) the benefits he obtains from democracy, (2) how much he wants a particular party to win, (3) how close he thinks the election will be and (4) how many other citizens he thinks will vote. Since the expected return is predictably higher for the high-income than for the low-income citizen, the former has a greater incentive to become politically informed. He also has a greater incentive to vote on the basis of expected benefits.

The Insight of James M. Buchanan and Gordon Tullock

Five years after Downs' (1957) masterpiece, there followed the most far-reaching and the only philosophical founding contribution, namely *The Calculus of Consent* by James M. Buchanan and Gordon Tullock (1962). Neither author was trained formally in philosophy or in political science. Yet, this book explicitly moved the analysis to the interface between economics, philosophy and political science, applying the tools of economics and economic philosophy to a detailed and far-ranging evaluation of political institutions in an attempt to delineate the logical foundations of constitutional democracy.

Buchanan and Tullock rejected the emphasis placed by Downs on the group behavior of political parties in favor of a model of collective decision-making that is more closely analogous to the theory of private choice. Collective action is viewed as the action of individuals when they choose to accomplish goals collectively rather than individually. Government is viewed as a set of processes that allows collective action to take place. From this perspective of methodological individualism, the rule of unanimity is advanced as a weak ethical criterion for the 'good' in evaluating both new constitutions and initiatives for constitutional change. Buchanan and Tullock embedded their analysis firmly within the framework of rational choice, acknowledging albeit that *homo economicus* may not always be as narrowly self-seeking as neoclassical economics frequently assumes. They further acknowledged that in the effecting of collective choices, the individual has no way of knowing the final outcome at the time that he makes his own contribution. For this reason, individuals lose that full sense of personal responsibility inherent in the making of private choices.

The rational self-seeking individual will contemplate collective action only when such action increases his expected utility. In an environment devoid of any kind of Pareto-relevant externality, the state would have no utilitarian support. Buchanan and Tullock therefore rationalized the existence of collective action as a means for individuals to combine in order to reduce the burden of external costs imposed upon them by purely private or voluntary actions. In contemplating such collective action, the rational individual is concerned to minimize his relevant expected costs, defined as the sum of his expected residual external costs and of his expected costs of decision-making within a collective framework.

In deciding whether any particular activity belongs within the realm of collective rather than private choice, the rational individual must take into account the expected cost of voluntary cooperative arrangements. If such costs are zero, all Pareto-relevant externalities would be eliminated by voluntary private behavior (here we note an early application of the 1960 Coase theorem, itself developed at the University of Virginia).

If the environment is one of positive transaction costs, however, the choice between non-cooperative private behavior, cooperative private behavior and collective political action must rest on the relative expected costs of these alternatives. The existence of Pareto-relevant external effects of private behavior is neither a

necessary nor a sufficient condition for an individual to entrust that activity to the realm of collective choice. In this regard, Buchanan and Tullock, *for the first time in formal economic analysis*, called specific attention to the fact that the collective organization of activities must also impose expected external costs upon the individual unless the collectivity itself is constrained to make decisions through a rule of unanimity.

Thus, the expected costs that collective choices impose on the individual depend on decision-making rules that govern such choices. In such circumstances, the individual will compare the expected costs of private choice with the expected costs of the most efficient form of collective action when making his decision to submit to collective action.

Buchanan and Tullock designed a generalized economic theory of constitutions specifically to analyze the problem of individual choice among alternative collective decision-making rules. This economic theory, now widely recognized as the most important and enduring insight of *The Calculus of Consent*, is outlined in Figs. 12.6–12.9.

Figure 12.6 outlines the nature of the relationship between the present value of an individual's expected external costs and the number of individuals required to take collective action. Buchanan and Tullock suggested that the curve **CN** will slope downwards throughout its range, reaching zero only where the rule of unanimity is in place. The point **C** on this curve represents the (high) external costs that the individual expects will be imposed on him if *any* single individual in the group
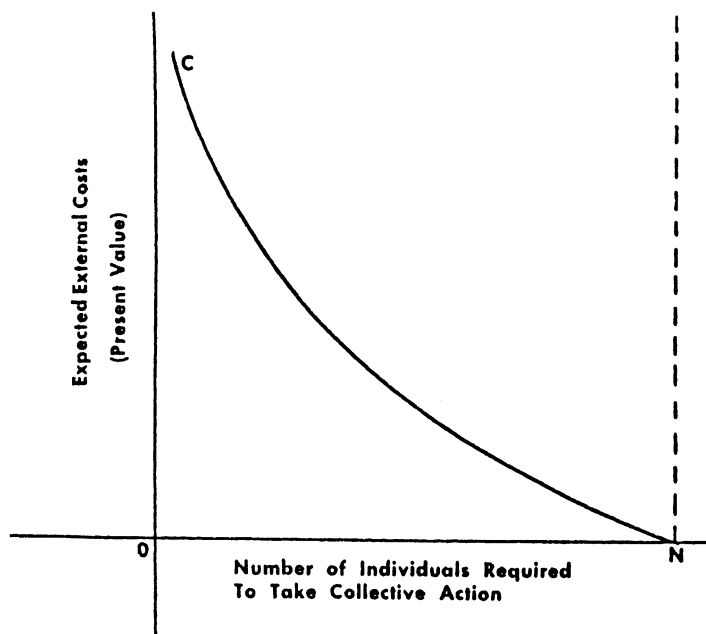
Fig. 12.6 The expected external costs of collective action
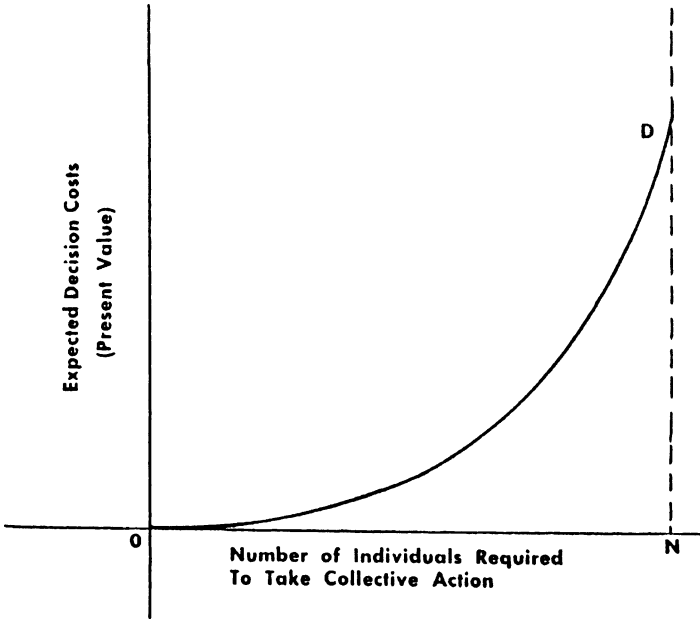
Fig. 12.7 The expected decision-making costs of collective action
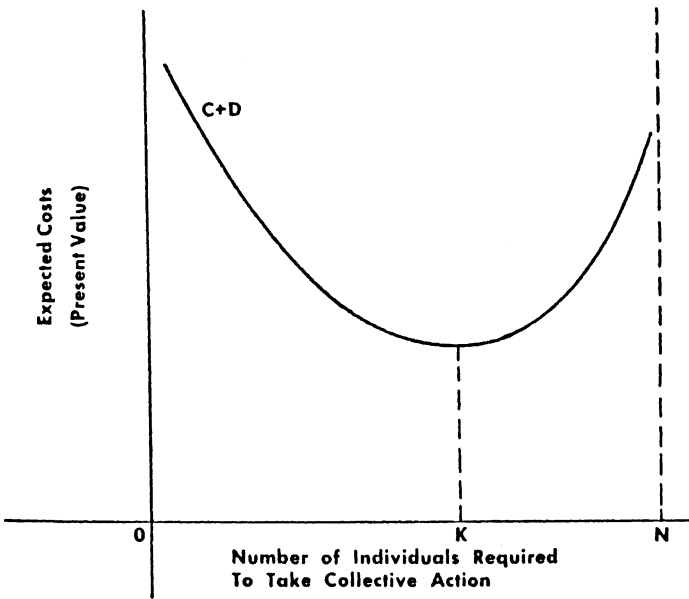
Fig. 12.8 The expected costs of collective action
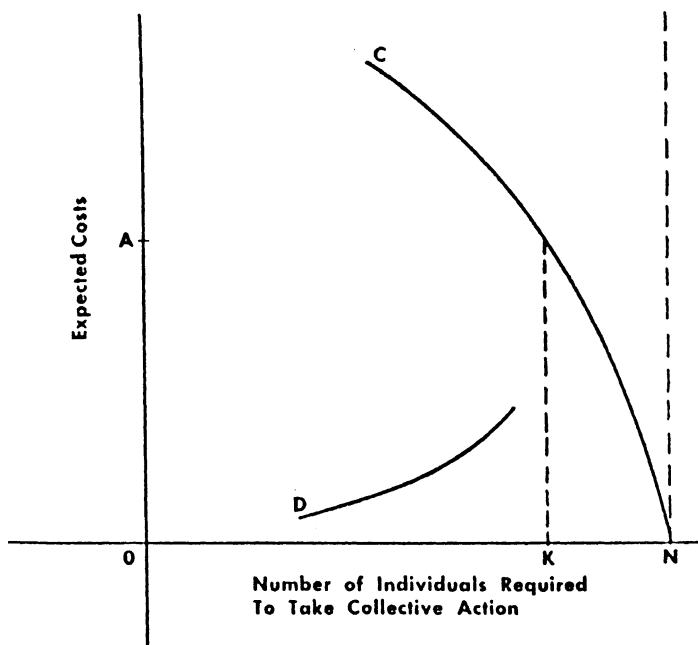
Fig. 12.9 The infeasibility of collective action

is authorized to undertake action for the collectivity. Note that **C** represents a randomly selected and not a uniquely designated individual, since the latter situation would be one of dictatorship.

As the decision rule becomes more and more inclusive, the individual's expected external costs of collective action decline, since there will be fewer decisions that the individual expects to run counter to his own desires. Only with the rule of unanimity will such costs be zero. For reasons outlined in Fig. 12.7, the rational individual will rarely choose unanimity as his most preferred rule when collective action is chosen over both private action and voluntary cooperation.

Figure 12.7 outlines the relationship between the present value of an individual's expected decision-making costs and the number of individuals required to take collective action. Buchanan and Tullock suggested that the curve **OD** will slope upwards throughout its range, reaching its highest point at **D** when the rule of unanimity is in place. At this point, the costs of strategic bargaining are so high as to render any form of agreement almost impossible.

Figure 12.8 sums the expected external costs and the expected decision-making costs functions vertically to create a curve that relates expected costs to the number of individuals required to take collective action. The rational individual will seek to minimize the expected costs of collective action by choosing the rule that requires **K/N** of the group to agree in order to act collectively. If the expected cost of private action or of voluntary cooperation is less than **OB** in Fig. 12.8, the rational individual will not endorse collective action.

As Buchanan and Tullock emphasized, the calculus of individual consent does not require that all potential collective action should be organized through the operation of the same decision-making rule. In their view, two categories of potential collective action can be separated even at this conceptual stage. In the first category are located those potential collective actions that are characteristically undertaken by government. Figure 12.8 effectively depicts the calculus of individual consent for this category.

In the second category are located those potential collective actions that modify the structure of established individual rights and property rights. The rational individual will foresee that collective action in this area potentially may inflict on him very severe costs. Figure 12.9 illustrates that the rational individual at best will require a much more inclusive rule as the basis for consenting to this category of collective actions. In the limit, the calculus of individual consent will break down entirely, and the individual will decline to enter into political society but will choose instead to protect this category of rights by private actions and/or by voluntary cooperation.

In Fig. 12.9, the expected external costs curve remains relatively high throughout the range of collective action rules short of unanimity. In such circumstances, the expected decision-making costs curve scarcely becomes a factor. In Fig. 12.9, for example, where the expected costs of private organization are depicted as **OA**, the expected external costs of collective action exceed the expected costs of private organization for all rules less inclusive than that shown by **K/N**. Given that expected decision-making costs rise exponentially in this latter range, the prospects for an individual agreeing to collective action under any rule short of unanimity are small.

Throughout this analysis, Buchanan and Tullock centered attention on the calculus of a single individual as he confronts the constitutional choice concerning rules of decision-making. What we should now perceive as a weakness in their book is the limited attention that Buchanan and Tullock devoted to dealing with the way in which individuals reach agreement concerning the rules that must govern various forms of collective action. Since individuals are aware of their own relative positions in society at the moment of constitutional choice, they are also aware that decision-making rules short of unanimity may impose different expected external costs upon them.

For the most part, Buchanan and Tullock evaded this issue, commenting that '[W]e prefer to put this issue aside and to assume, without elaboration, that at this ultimate stage, which we shall call the constitutional, the rule of unanimity holds' (Buchanan and Tullock, 1962, p. 77). In fact, they did not completely put the issue aside. They relied upon the extended time horizon of the individual in making his constitutional choices to explain his greater willingness to consent to rules that potentially are harmful to his shorter-term interests.

Constitutional rules, by their nature, are expected to be long-lived, since constitutional change is usually subject to highly inclusive rules of decision-making. The rational individual, confronted with a constitutional choice, is inevitably uncertain of his particular interest at some unspecified future time. In such circumstances, he will selfishly tend to choose rules of collective decision-making that maximize the

utility of some random individual. Such far-sightedness in constitutional decision-making differs sharply from the more myopic, sectional-based approach of the individual in the ordinary business of politics.

Buchanan and Tullock recognized that uncertainty alone would not necessarily guarantee unanimity in the prior judgment of individuals as to the rules of collective decision-making that would minimize costs. Nevertheless, they argued that any initial conflicts of opinion over rules should be amenable to reasoned compromise.

Buchanan and Tullock likened the resolution of such conflicts to the discussion that might take place between potential participants as to the appropriate rules under which a game shall be played. Since no player can anticipate which specific rules might benefit him during a particular play of the game, he will not find it difficult to concur with others in devising a set of rules that will constitute the most interesting game for the average or representative player.

Buchanan and Tullock recognized that the process of constitutional decision-making set out in their book has little relevance for any society that is deeply divided by reference to social class, race, religion, or ethnicity. Unanimity over a collective decision-making rule is especially unlikely when one of these coalitions is perceived to hold an especially advantageous position. Needless to say, this implies that *The Calculus of Consent* could not have been written in its existing form, at least with relevance for the United States, with the constitutional democracy explicitly central to their analysis, had the co-authors joined forces in the late 1960s rather than in the late 1950s.

In any event, the analysis of Buchanan and Tullock provided a number of important insights into constitutional political economy. First, it is rational to have a constitution, in the sense that there is more than one rule for collective decision-making. Second, there is no necessary justification for majority rule as the basis for collective decision-making. At best, majority rule should be viewed as one among many practical expedients made necessary by the costs of securing widespread agreement on political issues when individual interests diverge.

Third, it is rational to have a constitution that requires a more inclusive rule of collective decision-making with respect to incursion on individual rights and property rights than with respect to less fundamental issues. Fourth, the more inclusive the decision-making rule, the more willing will individuals be to the entrustment of decision-making to collective choice. The range of collective decision-making, thus, is not independent of the rules that govern such activities in societies that respect the primacy of individual choice.

Finally, the analysis of Buchanan and Tullock suggests that the over-all costs of collective decision-making are lower, with respect to any constitutional rule, in communities characterized by a more, rather than by a less, homogeneous population. From this perspective alone, a more homogeneous community would adopt a more inclusive rule for collective choice. However, the homogeneity characteristic affects expected external costs as well as expected decision-making costs. On balance, Buchanan and Tullock predict that the more homogeneous the society, the less inclusive will be the rules of collective choice and the more extensive will be the range of actions encompassed within the collective sphere.

Buchanan and Tullock deployed the rational choice model to offer a number of important insights into the logic of constitutional design. A comprehensive review of these applications is beyond the scope of this essay. However, their evaluation of the rule of simple majority voting is illustrative of the general approach.

Buchanan and Tullock grounded their discussion of the simple majority vote rule on the generalized assumption that individuals vary in the intensity of their preferences for or against specific collective actions. In such circumstances, the rule of simple majority, applied to a single issue of collective choice, may provide minor gains in utility for a majority at the cost of imposing major losses in utility upon a minority (abstracting from the issue of the problem of measuring utility across individuals). Rational individuals will recognize this danger when engaging in constitutional decision-making and will protect themselves from its most serious consequences by providing institutional opportunities for logrolling (or the trading of votes).

An institutional environment in which logrolling cannot occur is the standard referendum on a single issue determined by a simple majority vote conducted by secret ballot. The rational individual, concerned about the potential tyranny of the majority, will therefore be extremely wary about endorsing decision-making by referenda as the basis for determining collective choices.

Buchanan and Tullock noted that logrolling opportunities are prevalent in many of the political institutions of the Western democracies. Explicit logrolling is a common feature of all representative assemblies where exchanges of votes are easy to arrange and to observe. Such exchanges of votes significantly affect the political process. Implicit logrolling dominates the electoral process since the leaders of the political parties formulate complex mixtures of policies into electoral platforms designed to attract voters' support by appealing to intensely held preferences.

Buchanan and Tullock suggested that both explicit and implicit logrolling tend to improve the efficiency of the political process, even though these practices are widely criticized on ethical grounds. They demonstrated, however, that even when logrolling is possible, simple majority rule is likely to produce undesirable collective decisions, for example by over-investing in the public sector. Indeed, they further demonstrated that a system in which the open buying and selling of political votes is encouraged tends to improve the efficiency of simple majority rule as evaluated in terms of the Pareto criterion.

Recognition of the fact that preference intensities over policy alternatives differ among the electorate may encourage the rational individual to favor the bicameral over the unicameral legislature as a basis for constitutional design. A properly designed bicameral legislature, offering different bases of representation, will discriminate automatically between legislation potentially affecting intense minorities and legislation on which the intensity of desires is more or less equal. This will significantly improve the efficiency of the political process.

A further improvement in political market efficiency occurs when the constitution provides a president with veto power, effectively establishing a third house of the legislature. This third house represents the entire body of voters in one grand constituency, raising the minimum size of the logrolling coalitions and further protecting the individual voter from the excesses of the rule of simple majority voting.

In this manner, Buchanan and Tullock outlined the sort of calculus that the individual must undergo when he considers the question: Can the pursuit of individual self-interest be turned to good account in politics as well as in economics? They were able to show that, even under the behavioral assumption of extreme self-interest, something akin to the constitutional democracy conceived of by the American Founding Fathers would tend to emerge from the rational individual calculus. They concluded their epic on an extremely optimistic note, a note perhaps that some forty-five years later they would no longer feel able to hold:

> With the philosophers of the Enlightenment we share the faith that man can rationally organize his own society, that existing organization can always be perfected, and that nothing in the social order should remain exempt from rational, critical, and intelligent discussion. Man's reason is the slave to his passions and recognizing this about himself, man can organize his own association with his fellows in such a manner that the mutual benefits from social interdependence can be effectively maximized. (Buchanan and Tullock, 1962, p. 306)

The Insight of Mancur Olson

Prior to 1965, public choice had been developed with a primary emphasis on the vote motive. It is true that Downs (1957) and Buchanan and Tullock (1962) acknowledged the relevance of pressure group activities in the political process. Neither of them accorded to interest groups the central role that they evidently play in the determination of political outcomes in the Western democracies. In his important 1965 book, *The Logic of Collective Action*, Mancur Olson filled this lacuna in the public choice literature with his rigorous application of the rational choice approach to the analysis of interest groups.

Prior to Olson's book, economists, sociologists and political scientists had taken for granted the notion that groups of individuals with common interests usually attempted, often successfully, to further those interests by the application of political pressure. This notion played a central conceptual role in early American theories of labor unions, in the 'group theory' of the pluralists in political science, in John Kenneth Galbraith's concept of 'countervailing power' and in the Marxian theory of class conflict. This theory of interest group behavior essentially transferred the logic of the rational choice theory of individual behavior to that of groups.

In *The Logic of Collective Action*, Olson provided a dramatically different view of collective action. If individuals in some group share a common interest, furtherance of that common interest automatically benefits each individual in that group, whether or not he bears any of the costs of collective action to further that interest. Thus the existence of a common interest need not provide any incentive for individual action in the common interest, especially when any one member's efforts are highly unlikely to make the difference between group success and group failure.

From an analytical viewpoint, Olson demonstrated that the benefits of collective action take the form of public good in the sense that individual members of the

group cannot easily be excluded from any benefits that accrue. Economists recognize that voluntary and spontaneous market mechanisms either do not arise or seriously under-provide public good, as a consequence of the free-rider problem. This under-provision of markets is paralleled exactly by the under-provision of pressure in the case of large groups attempting to pursue a common interest.

Since many groups with common interests do not have the power to tax their members, Olson's theory predicts that many groups that would benefit from collective action will fail to organize effectively in pursuit of their common interests. This prediction is supported by evidence. There is no major country in which organizations of consumers effectively support consumer interests. There is no major country in which groups of unemployed workers are effectively organized for collective action. Neither taxpayers nor most of the poor are typically organized to act in their respective common interests.

Although the logic of collective action indicates that some groups can never act collectively, Olson suggested that other groups, with the assistance of ingenious leadership, may be able to overcome the difficulties of collective action. He posited three conditions, any of which is ultimately sufficient to make collective action possible, namely (1) that the relevant group is small in size, (2) that the group has access to selective incentives or (3) that the group can coerce the supply of pressure.

Suppose that a group is composed of a small number of members, each with identical preferences in favor of some common interest. An example of such a group would be an industry made up of two large firms that would gain equally from the provision of a government subsidy or a tax loophole. Since the lobbying activity of each firm, if successful, will exert a significant impact on profits, strategic bargaining between them predictably will result in a group optimal outcome. As the number of firms in the industry increases, however, the incentive to act collectively is eroded.

Even in an industry composed of many firms effective lobbying may occur where one firm has a differentially high absolute demand for collective action. In such circumstances, such a firm may engage in collective action, notwithstanding the inability of other firms to provide pressure of their own. This leads to the paradoxical exploitation of the great by the small. Olson illustrated the existence of this phenomenon in a variety of military alliances, in international organizations and in metropolitan areas in which collective good is provided across an entire metropolis by independent municipalities of greatly different size.

If large groups are to organize themselves effectively to supply pressure, Olson argued that they must engage in the provision of selective incentives to their members. These selective incentives are functionally equivalent to the taxes that enable governments to supply public good, except that interest group members, unlike taxpayers, cannot be coerced into accepting selective benefits.

Selective benefits either punish or reward individuals depending on whether or not they have borne a share of the costs of collective action. One example of this device is the provision of life insurance and medical policies to paid-up members of the American Association of Retired Persons at rates that would not be available to individual consumers. Another example is the mechanism whereby farm associa-

tions in the United States obtain most of their membership by deducting the dues owed by farm organizations from the patronage dividends or rebates of farm cooperatives and insurance companies associated with those organizations.

Large groups that fail to provide selective benefits may nevertheless overcome the free-rider problem associated with collective action where they are able to devise mechanisms for coercing the supply of pressure. An obvious device of this kind is the combination of the closed shop and the picket line utilized by some trade unions to make union membership a condition of employment and to control the supply of union labor during strikes. Another conspicuous example is the statutory requirement extracted by state bar associations in the United States that only paid-up members of the bar are allowed to engage in the practice of law.

Olson's application of the rational choice approach to the analysis of collective action offered disturbing insights into the political process. Since access to collective action is uneven, the application of pressure by groups in pursuit of common membership goals will be uneven across society. Legislatures that respond systematically to such uneven pressures (by taking advantage of rational ignorance among the electorate or by utilizing the campaign contributions to manipulate voters' preferences) may be able systematically to evade the centripetal pressures of two-party spatial politics while effectively providing tenure to incumbent representatives.

Conclusions

The five contributions evaluated in this essay together comprise the founding content of the public choice research program. By rejecting both the philosopher-king approach of economic science and the behavioral approach of political science in favor of the rational choice approach, the Founding Fathers revolutionized our understanding of the political process.

One important consequence of these contributions has been a dampening of the enthusiasm with which social scientists proffer policy advice to governments. A second important consequence has been the dampening of enthusiasm for active government even among scholars who still nurse strong suspicions concerning the behavior of private markets.

The Founding Fathers of public choice, in some cases by design and in other cases by accident, effectively leveled the playing field in the debate over the relative merits of governments and private markets. This playing field, by the mid-1950s, had become undeniably prejudiced in favor of an allegedly omniscient and impartial government.

In balancing this playing field, the Founding Fathers of public choice played an indispensable role in stimulating the Western democracies to abandon their mid-twentieth century flirtation with socialism, thereby paving the way for resurgence of market processes. The insights provided by the public choice research program rank among the most important of all advances in economic science during the second half of the twentieth century, when measured in terms of their contribution to the wealth of nations and to the expansion of individual liberty.

References

Arrow, K.J. (1950). "A difficulty in the concept of social welfare." *Journal of Political Economy*, 58, 328–346.

Arrow, K.J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Black, D. (1948a). "On the rationale of group decision making." *Journal of Political Economy*, 56(1), 23–34.

Black, D. (1948b). "The decisions of a committee using a special majority." *Econometrica*, 16(3), 245–261.

Black, D. (1948c). "The elasticity of committee decision with an altering size of majority." *Econometrica*, 16, 261–272.

Black, D. (1958). *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Black, D. (1972). "Arrow's work and the normative theory of committees." (unpublished manuscript).

Black, D. and Newing, R.A. (1951). *Committee Decisions with Complementary Valuation*. Glasgow: William Hodge.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Coase, R.H. (1981). "Duncan Black: A biographical sketch," in Tullock, G. (ed.) *Toward a Science of Politics*. Blacksburg: Center for Study of Public Choice.

Condorcet, M. (1785). Essai sur l'application de l'analyse a la probabilite des decisions rendues a la pluralite des voix. Paris.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Grofman, B. (1981). "The theory of committees and elections: The legacy of Duncan Black," in Tullock, G. (ed.) *Toward a Science of Politics*. Blacksburg: Center for Study of Public Choice.

Grofman, B. (1987). "Black, Duncan," in Eatwell, J., Milgate, M., and Newman, P. (eds.) *The New Palgrave: A Dictionary of Economics*. London: Macmillan, pp. 250–251.

Hotelling, H. (1929). "Stability in competition," *Economic Journal*, 39, 41–57.

McKelvey, R.D. (1976). "Intransitivities in multidimensional voting models and some implications for agenda control." Journal of Economic Theory, 18(1), 1–22.

Moulin, H. (1983). *The Strategy of Social Choice*. Amsterdam: North Holland.

Moulin, H. and Young, H.P. (1987). "Condorcet, M," in Eatwell, J., Milgate, M., and Newman, P. (eds.) *The New Palgrave: A Dictionary of Economics*. London: Macmillan, pp. 566–567.

Olson, M. (1965). *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Rowley, C.K. (1987). "Borda, Jean-Charles de (1733–1799)," in Eatwell, J., Milgate, M., and Newman, P. (eds.) *The New Palgrave: A Dictionary of Economics*. London: Macmillan, pp. 262–263.

Rowley, C.K. (1991). "Duncan Black: Pioneer and discoverer of public choice." *Journal of Public Finance and Public Choice*, 2, 83–87.

Chapter 13
The Perspective of Economics[1]

Robert D. Tollison

Introduction

I address public choice from the perspective of economics in this essay. The "perspective of economics" is taken to mean the application of the principles of maximizing behavior and demand and supply to institutions and behavior in the political world. I begin with a discussion of this familiar methodology, and then proceed to illustrate how the principles of maximizing behavior and demand and supply can be applied to the various component parts of a representative democracy, including the legislative, executive, and judicial branches, as well as interest groups, bureaucracy, and voters. This will be in no sense a review of the literature. The point is to illustrate how economic principles can be applied to political behavior in each of the above contexts. In each case a single and simple illustration will be given. In such a way, the reader can decide whether the economic perspective really adds anything to the understanding of political behavior over and above alternative analyses. For example, do we learn more about a legislator's behavior with an assumption that he acts in his self-interest or in the "public interest?" Finally, although many of the illustrations are related to U.S. political processes, I endeavor in each case to generalize the discussion to an international setting.

The Perspective of Economics

In the movie, *A Few Good Men*, a Marine officer, who is testifying at a court martial, is asked if a soldier was in danger from his colleagues. He does not answer the question, so the interrogator repeats the question, adding, "in mortal danger?" The officer responds, "Is there any other kind?" This response represents my basic

approach to the topic of this essay. When given the assignment to discuss the contributions of economics to public choice, my instinct was to echo the answer of the Marine officer, "Is there any other kind?"[2]

Public choice emerged from the maximizing paradigm of modern microeconomics, and it remains to this day within that approach. This tried and tested model colonized the traditional intellectual territory of political science. Even the key political scientists who participated in the public choice revolution, such as Riker (1962), assumed that politicians and their coalitions were maximizing some objective subject to constraints (for example, the pro rata gains to the minimum winning coalition). The simple transfer of the economist's model of individual self-interest to the subject matter of political science was the seed corn of the public choice revolution.

In this essay I discuss the transfer of economic methodology to the theory of public choice, and attempt to assess whether the application of the economist's model of human behavior has been more or less successful. First, I briefly stretch the economist's model, and then I describe what it means to argue that its application to politics has been "successful."

Any conventional textbook on microeconomics lays out the economist's model of individual behavior.[3]

Individuals are assumed to have transitive and stable preferences, which they pursue by making trade-offs among desired goods as a function of their relative costliness. The law of demand, for example, is an empirical proposition about such behavior. In effect, the economic model predicts that individuals will seek to minimize the effect of constraints, such as income and prices, on their behavior. If "price" rises, they economize on the use of the more costly "goods"; if "price" falls, they expand their use of the less expensive "goods." The quotation marks around "price" and "goods" are there to indicate that the economic model is general. The model applies to any context which includes "prices" and "goods", ranging from obvious cases like the price of chocolate to other cases, such as a market for legislation, in which "prices" and "goods" may not be so obvious. Any subject is fair game for the application of the economic model, including the world of politics. The only thing that limits the expansion of the economic approach to other areas is the creativity of the analyst. Economics, of course, may not explain or predict behavior very well in these applications, but there is no subject matter to which economic reasoning cannot be deployed. Arguably, there is nothing in the world that is "non-economic."[4]

---

[2] The reader may want to contrast this approach to that given in Cohn (1999), in which political science, specifically Riker's Rochester School of Political Science, is given the credit for the invention of public choice. While I do not quarrel that Riker and his students have been important figures in modern public choice theory, they are most surely not the only ones when one considers such names as Downs, Buchanan, Tullock, Stigler, Niskanen, and others too numerous to mention, all of whom wrote as economists.

[3] See Silberberg (1995) for an excellent discussion.

[4] Politics is not the only area of study that has been colonized by the economic approach. Other areas include the family, crime, religion, and law.

The economic model is a simple model of behavior, but not a simplistic model. Preferences, as noted, are assumed to be given and stable. This places preferences outside the purview of economists. People want what they want; economists cannot say why people like chocolate. Taking preferences as given, the maximizing model is quite general. Sometimes individuals will maximize utility, and sometimes they will maximize wealth (a special case). Individuals are "selfish" only to the extent that they pursue their goals purposively. The goals can be anything the individual chooses, be it piggish wealth accumulation or some notion of a higher life including service to others. The economic model says to the analyst, 'Give me the constraints or "prices" and I will give you predictions about how individuals will respond'. All behavior is economic; if the "price" of altruism falls, individuals will be more altruistic. Even the altruist will seek to help others in the most effective manner, given the "price" of altruism.

The stability of preferences is an empirical issue. Typically, economic analysis proceeds on the basis that individuals reach "equilibrium" states of behavior. That is, a constraint or price changes, individuals rearrange their behavior so as to minimize the effect of the change on their lives, and then they settle down into a new equilibrium mode of behavior. Obviously, unstable preferences would undermine the explanatory value of the economic model, which is based on tracing the effects of constraint changes in the face of given preferences. This does not mean that preferences never change or evolve, only that they are stable enough for the economic approach to make reliable predictions. In both markets and politics, equilibrium behavior seems pervasive. Consumption decisions are repetitive; political transactions are durable and last for a long time (for example, the Interstate Commerce Commission or Social Security).

How does one evaluate the "success" of the economic model in analyzing politics? The primary criterion is how well the economic model explains or predicts political behavior relative to competing models, say in the sense of a statistical test or an R^2. This criterion cannot always be applied because it is not feasible to test all theories empirically. In some cases we have to use our judgment about what is going on or about what actually "explains" events. Is the pattern of predictions consistent with economizing behavior or with some other model in the absence of a defining empirical test? In the discussion of the success of economic models in this paper, however, I shall primarily adhere to the testability criterion for success; that is, how well have these models fared in empirical tests.[5]

The key point, then, to keep in mind as I proceed is that the economic content of public choice is taken to mean that political actors, like private actors, pursue their ends effectively, but the constraints they face in the process are different. Hence,

---

[5] It should not go unnoted, however, that the public choice paradigm has had great acceptance in the larger sense of being a useful way to think about politics and political institutions. Political actors are generally seen today as self-interested and not disinterested agents; government is no longer treated as an exogenous, unexamined institution in economic and political models ($\bar{G}$); and public choice analyses permeate the work of modern economics and political science. Public choice is no longer an interloper; it is a paradigm.

political actors (bureaucrats) will behave differently than private actors (corporate executives) for this reason, and not because they are different types of people. My effort in this regard will be to cover selected areas of public choice analysis in order to assess how well the economist's model has performed in explaining political behavior and institutions. I will not try to be copious in the sense of a literature review; I will rather try to be concise in offering an example of how to apply the economic model to selected areas of public choice analysis, beginning with the legislature.

One final proviso is in order. There is no doubt that the economic approach has come under heavy assault in recent times (Thaler, 1992; Green and Shapiro, 1996). For the most part, in my view, economic methodology has withstood these attacks. For every anomaly, there is a rational choice explanation. Nonetheless, this debate will continue in the literature, but in the meantime, this essay will offer an unashamedly thick rationality approach to the subject matter of public choice.[6]

The Legislature

The legislature is the most analyzed institution of representative democracies in modern public choice analysis. From the perspective of economics are the principles of demand and supply relevant to the legislature? To explore this question the labor market for legislator services is analyzed. Specifically, I address the problem of how legislators are paid, using U.S. state legislators as the example to be analyzed. My explanation of legislative pay will seem familiar to economists. Nonetheless, it will contrast markedly with the explanations and approaches to the same problem offered by other observers of such matters. For example, "Most states fail to pay their lawmakers anything approximating a living wage" (Straayer, 1973, p. 3).

In effect, I view legislators as participants in a labor market, and I try to explain differences in the legal (above board) pay of legislators by factors that affect the supply of and demand for their services. The supply of legislative services is analogous to the supply of any service where labor is extensively used in (roughly) fixed proportions to other inputs. The quantity-supplied of legislative services (which I measure in man-years per year) is therefore determined by the relative wage, the price of inputs other than labor, and technology. Each state has a separate supply function, but I do not expect the conditions of supply to vary greatly across states. Potential legislators are never a significant fraction of the available labor in a state, and the occupational composition of legislatures is similar across states. These positions are held primarily by members of professions who can capitalize (through extra-legal pay) readily on certain aspects of being a legislator. Lawyers

---

[6] Thick nationality is a term used by Green and Shapiro (1996), which means rationality in the sense of wealth maximization rather than the more general case of utility maximization.

often continue to draw a wage from their law firms while serving. Farmers can be legislators where sessions are held between growing seasons. The reasons that banking, insurance, and real estate people gravitate to these offices are not hard to discern.

In each state there is some demand for legislative influence. The demand for legislative influence implies a derived demand for legislators. The technical relationship between influence and legislators is not one of proportionality because an excessive number of legislators would dilute the influence of each and might not be able to pass any laws. I further expect that, given the lack of low-cost substitutes for legislative action within a state, the elasticity of the demand for representation with respect to the legislative wage rate must be close to zero over the relevant range. Across states, in contrast to the relative invariability of supply in this market, I expect that the demand for representation will shift as a function of state income, population, budget size, and so forth.

With this background in mind, note that wage determination takes essentially two forms across states. In some states legislative pay is set in the constitution and is difficult to change. A new wage requires the passage of a constitutional proposal. Such proposals typically emanate from the legislature under relatively strict voting and quorum rules and must be signed by the governor and passed in a statewide referendum. In other states pay is set by a statute passed by both houses of the legislature and signed by the governor. These pay bills are subject to legislative consideration under normal voting and quorum rules and do not require a statewide referendum.

I contend that legislative determination of pay by statute amounts to a strong form of union power. Unions typically achieve higher relative wages by restricting entry. In this case entry is somewhat more loosely controlled through constitutional limitations on the size of the legislature and on the procedures for gaining a seat, and legislators are given a direct hand in wage determination. I would expect to observe the impact of this monopoly power in higher relative wages for legislators in these states.

The conditions in the legislative labor market for a single state are depicted in Fig. 13.1.

Each legislature is treated as a separate labor market. A measure of legislative output (Q_L) in terms of man-years per year is on the horizontal axis, and annual legal pay (W_L) measured as dollars per man-year is on the vertical axis. The competitive supply curve for successful applicants for these seats is given by S. This relationship represents the wage that must be forthcoming for a given level of output to persuade prospective legislators to run for and to accept office. Following the previous argument, I draw a completely inelastic demand curve over the relevant range for the services of legislators. In the absence of any contrary evidence I assume that existing wages clear the market for the given constraint on legislative size in both union and nonunion states. That is, there is no excess supply.

In states where the legislative wage is constitutionally determined, some given wage, W_C, will prevail. Candidates will adjust to the given wage, and supply or
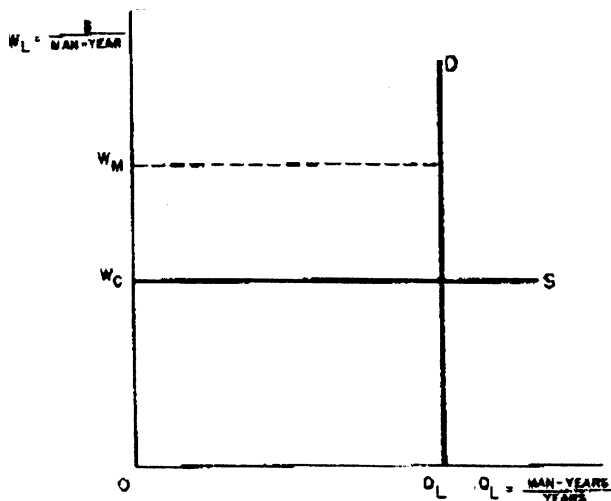
Fig. 13.1 Determination of legislative pay

marginal opportunity costs will shift accordingly as more- or less-qualified indivi-
duals seek election, so that the market clears. In states that allow legislative control
over pay, the wage is adjusted by legislators to maximize the present value of a seat.
This wage is, for the moment, arbitrarily drawn in Fig. 13.1 at W_M.

The main issue confronting this theory concerns the forces that constrain legis-
lators from setting an infinite wage in Fig. 13.1. Since I argue that the demand for
legislator time is completely inelastic over the relevant range, this pay problem
reduces to a question of what limits the wage-setting ability of the legislature under
these conditions.

Basically, the present value of a seat will be inversely related to the wage rate
after some point, because higher wages will attract new entrants and alienate voters,
both of which dampen reelection prospects and offset the effect of increasing the
wage on the present value of seats. Incumbents must thus trade off union wage gains
and other benefits from being in office against the extra costs associated with
increased competition to retain seats. There is thus a determinate upper bound on
the monopoly wage in the problem.

As a result of monopoly power in this labor market, then, wages in states where
legislators can set their own wage will be higher on average (W_M) relative to states
where the wage is set in the constitution (W_C). The legislative union predictably
will have a substantial impact on relative wages because the demand for legislator
services will be quite inelastic, as suggested earlier. This condition follows from the
rules of derived demand in two related senses. First, there is only one legislature per
state, so there is not a nonunion sector from which to buy output. Second, there are
in general poor substitutes for the services of legislators (for example, legal versus
private cartels).

This model of legislator pay offers a robust explanation of state legislator pay in the United States. In fact, the amount of relative wage-setting power ceded to the set-your-own-pay legislatures is higher than for any known labor union (300 to 400%).[7] It should therefore be clear that the principles of supply and demand can be readily adapted to the public choice context of the legislature. At the core of the legislative process are markets and allocation mechanisms familiar to modern economics and a great distance removed from the view that legislators are under-compensated.[8]

Moreover, this lesson applies with appropriate modifications to the legislatures of other countries. Stigler (1976), for example, discusses the determinants of legislative size across countries, and finds that such factors as population provide a common explanation for legislative size in different national legislatures. And the work of Marvel (1977) on the British Factory Acts clearly puts the British Parliament into an interest-group context as early as the 1830s. So too does Weck-Hanneman's (1990) work on direct democracy in Switzerland suggest that using the voters as the legislature is no insurance against protectionist outcomes. Public choice analysis of the legislature and related institutions is not confined to the United States.

The Judiciary

No other institution of democratic government is more insulated from the political process than the judiciary. In the American political system, constitutional rules provide the courts with a high degree of independence from the other branches of government. At the federal level, for example, judges are granted life tenure; their nominal salaries cannot be reduced; and they can be removed only by means of impeachment for high crimes and misdemeanors. While most state judges typically serve more limited terms, their offices are generally much more secure than those of their counterparts in the legislative and executive branches. Judicial independence limits the ability of these other branches of government to sway courts' decisions, and because judges face heavy sanctions in cases of detected corruption, they are arguably unlikely to be influenced by the economic interests of the parties before them. In short, the standard view is that the judiciary is—and, indeed, should be—above the fray of interest-group politics.

Given their effective independence from ordinary political pressures, what motivates judges to behave in any particular way? There are three major hypotheses regarding the nature and consequences of judicial independence. One view holds

---

[7] See McCormick and Tollison (1978, 1981) for empirical results.

[8] There are related issues here concerning the potential for extra-legal (below board) compensation to legislators, which are linked to the legislator's occupation. Lawyers, for example, are more effective at combining legislative service with making money on the side, so that there will predictably be more lawyers in low-legal-pay legislatures. Again, see McCormick and Tollison (1981).

that an independent judiciary operates as a necessary counterweight to the legislative and executive branches. The judiciary acts to protect society from unconstitutional encroachments by the other government branches, and judges are therefore motivated by their concern for the public's interest. A second view regards the independent judiciary as an agent not of the general public's interest, but of the interests of groups that otherwise are unrepresented (or under-represented in other political forums). Whereas the legislature faithfully responds to the wishes of the majority, judges interpose their wills to protect the interests of politically vulnerable minorities. Finally, the independent judiciary may actually be something of a loose cannon. Posner (1986), for example, argues that because judges are insulated from interest-group politics and receive no monetary payoffs from deciding a case in a particular manner, the economic self-interest of judges cannot explain judicial decision making. He suggests instead that judges seek to maximize their own utility by imposing their personal preferences and values on society.

In an important contribution to public choice theory, Landes and Posner (1975) contend that these popular models of the functioning of the independent judiciary are ad hoc and unconvincing. They propose an alternative economic theory in which the courts increase the durability of wealth transfers purchased from the legislature by interest groups. By reason of its effective independence from the sitting legislature and practice of interpreting laws on the basis of original legislative intent, the judiciary confers on legislation something of the character of a binding long-term contract. By construing statutes in this manner, the judiciary increases the durability of legislative contracts and, hence, raises the price interest groups are willing to pay for wealth transfers in their own favor.

In the interest-group theory of government, legislatures are modeled as firms that supply wealth transfers in the form of special-interest legislation. Legislatures assign property rights in wealth transfers to the highest bidder by means of legislative contracts, i.e., statutes. Domestic producers purchase tariff and non-tariff barriers to protect them from import competition, farmers purchase production-restricting marketing orders and price subsidies to increase their incomes at consumers' expense, and so on.

But while there are many similarities between legislative markets and ordinary markets in this regard, the two differ in at least one important respect: the mechanisms available for enforcing contracts once they have been negotiated. There are basically two contract-enforcing mechanisms in private markets. One is enforcement by a third party. In this case the contracting parties agree to rely on an independent arbitrator or the courts to resolve disputes and sanction noncompliance. Alternatively, when explicit agreements are absent or incomplete by reason of being costly to negotiate, self-enforcing mechanisms help maintain a contractual relationship. Each party relies upon the threat of withdrawal of future business to provide assurance that implicit agreements will be honored (Klein and Leffler, 1981).

In political markets, however, the legislature can, in principle, break its legislative contracts at any time, and leave any "injured" party with no immediate avenue of redress. An interest group cannot bring suit against the legislature for modifying

or repealing an existing legislative contract simply because of shifts in the political winds. Landes and Posner (1975, p. 879) provide an example in which the dairy industry buys a tax on the sale of margarine in one session of Congress, but the margarine producers buy the removal of the tax in the next session.

This example illustrates the dynamic insight that contract negotiations between legislatures and interest groups will be thwarted if political fortunes are easily reversed. Uncertainty with respect to the length of time over which an interest group can expect to collect the benefits it has purchased will tend to lower the present value of the transfer, and therefore reduce the price it is willing to pay. Given that individual legislators face a limited time horizon owing to frequent electoral challenges, resulting in unpredictable shifts in the composition of the legislature, markets for legislative wealth transfers would not function very efficiently in the absence of institutional constraints capable of mitigating this source of contractual instability.[9] Interest groups are not likely to expend time and treasure to secure the passage of legislation if, once enacted, it tends to be easily amended or repealed. It should therefore not be surprising that wealth-maximizing legislatures have adopted various measures designed to enhance the stability of legislative contracts and thereby increase the demand prices for legislative output.

Landes and Posner divide these institutional arrangements into two categories. The first is composed of the constitutive rules of the legislature itself. Procedural norms on such matters as bill introductions, committee hearings, floor action, and filibusters serve to increase the continuity, regularity, and stability of the legislature's operations. By making it more difficult to enact legislation in the first place, such measures also make it more difficult to amend or repeal existing laws.

The existence of an independent judiciary also enhances the durability of legislative contracts. Legislation is not self-enforcing; recourse to the courts is necessary to give effect to often vague or ambiguous statutory language. If judges act at the behest of the sitting legislature in interpreting previously enacted legislation, decide cases with an eye toward protecting otherwise under-represented groups, or simply indulge their own personal preferences, they might refuse to enforce the bargained-for statute. Such behavior would render earlier contracts null and void.

In contrast, if independence means that judges can be relied upon to interpret and enforce legislation in accord with the original legislative intent, judges will tend to protect the integrity of the legislature's contracts with interest groups. By providing such durability, the courts enhance the value of present and future redistributive legislation and facilitate the operation of the market for wealth transfers. On the other hand, if the legislative marketplace more closely resembles a Hobbesian jungle, such legislative contracts will be worth little, and governmental wealth transfer activity will greatly diminish.

---

[9] In the limit such wealth transfers would tend toward zero.

In the Landes–Posner model, the judiciary is part of the institutional structure that induces equilibrium in the market for wealth transfers. By virtue of its independence, and by interpreting legislation on the basis of original intent (i.e., a reversion point), the judiciary functions to limit cycling in majority rule decisions. This judicial function tends to increase the present value of legislative wealth transfers to special interest groups. As Landes and Posner explain, however, the value of the courts to the legislature in this regard and, not coincidently, the ability of the judiciary to maintain its independence, depend on how well the courts play their assigned role.

What motivates judges to behave in the ways predicted by the Landes–Posner model? Landes and Posner provide a theoretical reason why legislatures might benefit from the existence of an independent judiciary, but not why judges themselves would benefit from enforcing legislative contracts with interest groups in the face of political pressure. Legislative procedural rules may make it costly for margarine producers to buy the repeal of a tax enacted at the dairy industry's behest, but what prevents the courts from declaring the tax unconstitutional? Subsequent empirical tests of the Landes–Posner model have furnished two possible answers to these questions. One is that judges are rewarded for behaving independently. The other is that alternative contract-enforcement mechanisms exist that tend to be relied on more heavily in jurisdictions where the judiciary is less independent. An independent judiciary is only one of several institutions of democratic government that play complementary roles in promoting the durability of legislative wealth transfers.

First, in a direct test of the Landes–Posner model, Anderson et al. (1989) examined the relationship between the annual salaries of judges serving on state courts of last resort, measures of their opportunity costs for serving on the court, prospective workloads, measures of judicial independence, and the courts' propensities to overturn legislation on due process grounds. The goal was to determine whether judges are in fact rewarded by legislatures (in the form of higher pay or budgets) for behaving independently in the Landes–Posner sense. In sum, the evidence from due process challenges to legislative acts suggests that "self-interested judges can be shown to behave in manner consistent with the functioning of efficient markets for coercive wealth transfers for the same reasons that other participants in those markets participate—wealth maximization" (Anderson et al., 1989, p. 3).

Second, in any principal–agent relationship the optimal amount of judicial discretion depends on the configuration of the costs and benefits of delegating decision-making authority to that branch. Some judicial independence is beneficial to the sitting legislature (i.e., judges enforcing contracts with respect to their original meanings), but too much independence (judges indulging their own personal preferences) may inhibit the well-ordered functioning of the market for wealth transfers. These observations suggest the existence of an optimal amount of judicial independence and, hence, an optimal mix of institutional constraints for promoting the durability of contracts with interest groups in particular circumstances.

Constitutional provisions, or what Landes and Posner term "legislation of a higher order", represent an alternative institution in the interplay between the legislative and judicial branches. Such provisions are worth more than ordinary legislation to interest groups because they are more durable. They are also more costly to obtain in the first place. Whereas the enactment of ordinary laws typically falls under the normal majority voting rules of the legislature, constitutional amendments are subject to stricter procedures, typically requiring approval by legislative super-majorities and subsequent ratification by popular vote.

Whether an interest group will pursue the more costly route of constitutional amendment to secure a wealth transfer in its own favor consequently depends on the expected durability of wealth transfers secured through the normal legislative processes. Crain and Tollison (1979) used data from U.S. states to test the Landes–Posner model in this context. The model assumed that because interest groups could depend on the courts to enforce legislative contracts in jurisdictions where judicial independence is high in the Landes–Posner sense, they would rationally tend to rely more on normal legislative processes in those jurisdictions. On the other hand, constitutional amendment would be worth more to interest groups in states with less independent judges. At the margin, interest groups will demand ordinary legislation or extraordinary constitutional change to the degree of a state's particular judicial independence. The results of this empirical model support the predicted trade-off of the Landes–Posner theory. The frequency of constitutional amendment tends to be higher in states with lower judicial independence, other things being equal.

Other institutions of democratic government also appear to substitute for judicial independence in ways predicted by Landes and Posner. For example, as legislator tenure and the size of the voting bloc controlled by the legislature's majority party increase, the value of an independent judiciary declines because legislators will be less likely to renege on the bargains they strike with interest groups. Reputations for honoring commitments are as valuable to politicians and political parties as they are to suppliers of more ordinary goods and services. Evidence from the states adduced by Crain et al. (1988) suggests that the sizes of legislative majorities trade off with measures of judicial independence in ways consistent with the functioning of a well-ordered market for wealth transfers.

Two final points about the public choice analysis of the judiciary should be noted. The empirical evidence supporting the Landes and Posner theory is scanty at best, especially the evidence presented by the authors themselves. Other work, as cited above, has proved more supportive, but, still, the empirical evidence is weak. Moreover, when one moves to the international arena, it is clear that the separation of powers is important. Rowley (2000), for example, details differences between the United States and England, in which this point is highlighted with respect to the budgetary process. It is also apparent that the type of legal system (civil vs. common law) plays an important role in economic growth and development, with common law being the growth-friendly legal system (Wessel, 2001). For international comparisons, these important points must be kept in mind.

The Executive

Previous work on the U.S. presidency has examined the president's formal and informal powers. Neustadt (1960) focused on the president's informal power and his ability to persuade or bargain with Congress in an institutional setting which places the two branches in conflict. The formal powers of the president (vetoes and appointments) have been examined using the structurally induced-equilibrium (SIE) models introduced by Shepsle and Weingast (1981).

Although economists and political scientists have derived equilibrium results from the bargaining game, and SIE models by including a presidential preference set, the content of this preference set has remained a black box. Since these models do not specify the policies preferred by the president, few predictions can be made about the bills the president will veto, the budget he will propose, the people he will appoint, or the regulations he will promulgate and enforce.

The few works that have advanced positive theories of presidential behavior make the essential point that the U.S. President is not a popular vote maximizer but an electoral college vote maximizer. Wright (1974), in an important early paper, showed that New Deal spending in the 1930s could be explained as a function of a measure of electoral votes across states. Anderson and Tollison (1991a) found this same result while controlling for measures of congressional influence. Grier et al. (1995) argued that winner-take-all voting in states and the unequal distribution of electoral votes across states in presidential elections make incumbent presidents rationally place more weight on the preferences of voters in closely contested, larger states when making policy decisions. They tested this hypothesis by examining whether presidential veto decisions are influenced by the floor votes of senators from these electorally crucial states. In a pooled sample of 325 individual bills from 1970 through 1988, they found significant evidence of this behavior by incumbent presidents; that is, the more senators from electorally important states oppose a bill, the more likely the president is to veto it, even when controlling for a wide variety of conditioning variables, including the overall vote on the bill.

Several basic points should be kept in mind here. First, the behavior of the executive branch of government is among the least studied parts of modern public choice analysis. This literature is in its infancy. Second, more so than other areas, this literature is tied exclusively to U.S. political institutions, namely, the Electoral College system of electing presidents. Third, the literature is rife with measurement issues. Some authors use electoral votes per capita, some use raw electoral votes (a proxy for population), and some use closeness-weighted electoral votes (either per capita or raw).

Nonetheless, in keeping with the central point of this essay, presidential behavior in this approach is modeled as maximizing electoral votes subject to constraints. Essentially, the president is analyzed as a careful shopper for electoral votes in his effort to be elected or reelected. States in which the incumbent president or candidate expects to win or lose by a wide margin can safely be ignored in this process. States that are predicted to be close will be the recipients of presidential

largesse and visits. The constraints on this activity include time, campaign resources, congressional influences over federal pork, and so on.[10] Such a model has thus far provided a strong predictive theory of presidential behavior in a variety of areas. It also represents a core example of how simple economic theory can add to our understanding of political behavior.

The basic operation of the theory is simple. In the United States, presidents are not elected by the popular vote but by an Electoral College. Each state has a number of electoral votes equal to its number of representatives and senators. A simple majority of the popular vote in a state usually suffices to win all its electoral votes. The winner of the majority of electoral votes (270) is elected president, a fact which raises the odd, but thankfully rare, prospect that a candidate could lose the overall popular vote and still be elected president (Gore vs. Bush, 2000).

An economic model of presidential behavior and presidential candidate behavior maps into this situation easily. When faced with a choice among states with respect, for example, to new funding initiatives, the president will estimate the possibility that he will win the state times the number of electoral votes. States with higher expected values will receive the funding, following an equi-marginal rule of funding allocation. States that are not expected to be close (win or lose) or small states are left out in the cold in this calculation.[11] All forms of presidential behavior, and not simply funding, can be analyzed with this model. The relevant constraints on the president are the obvious ones—time and money.

This approach has been successfully employed, as noted above, to explain the allocation of New Deal spending across states, presidential vetoes, campaign stops by presidential candidates (Brams and Davis, 1974; Colatoni et al., 1975), and still other aspects of presidential decision making (Anderson and Tollison, 1991a, b). Though still in its infancy, this approach, at least for the United States, has the potential to fill in the black box of presidential preferences and to offer a positive economic explanation of presidential behavior. It also clearly finds its roots in the basic economic methodology of maximizing expected value subject to constraints.

The chief executive outside of the U.S. setting, especially in parliamentary democracies, is coincidential with the leader of his party in the legislature. In this context parties represent coalitions of interests that are not necessarily driven by the same type of geographic imperatives as in the United States. There is also the problem of forming coalitions in the parliament in order to fashion a governing majority. Rowley (2000) provides a clear discussion, for example, of how the office of prime minister functions in England. Again, however, these chief executives are vote-maximizers, only in a more complex and less geographically oriented system

---

[10] Even if the candidate is a lame duck and cannot run for reelection, the party has strong incentives to control shirking so that the lame duck behaves as if he were actually running for reelection.

[11] Note that closeness is more than just a previous victory margin in a state. Volatility of the vote also matters. A state with a previous victory margin of seven points and a standard deviation of 2% is safer than a state with a previous victory margin of 12 points and a standard deviation of 5%.

than the United States. Moe and Caldwell (1994) outline the relevant public choice consequences of the presidential and parliamentary systems.

Interest Groups

The economic analysis of an interest-group economy is relatively straightforward, and can be stated in more or less conventional demand and supply terms (McCormick and Tollison, 1981). The demand for transfers is based upon the organizational costs facing potential interest groups. Net demanders of transfers will be those groups that can organize for collective action in a cost-effective fashion. In other words, net demanders will be those groups that can organize to lobby for $1 for less than $1. Net "suppliers" are simply the inverse of the demand function for transfers, namely, those for whom it would cost more than $1 to organize to resist losing $1 in the political process. "Suppliers" is in quotation marks because individuals clearly would not engage in such a "transaction" voluntarily without being coerced by the state.

The equilibrium amount of transfers is determined by the intersection of the demand and "supply" curves, and this equilibrium is facilitated by the actions of the agents of the political process, such as elected officials. The incentives of these agents are to seek out "efficient" transfers by targeting "suppliers", who will generally be unorganized with low per-capita losses from transfers and regulation (why spend $1 to save $0.10?), and by targeting demanders who will be well organized and active in the political process. If political agents miscalculate and transfer too much or too little wealth, the political process will discipline them, for example, through elections.

There are various testable implications of this framework, which boil down to predictions about the costs and benefits of lobbying. When the benefits of lobbying are higher and the costs lower, there will be more transfers and more lobbying (and lobbyists). Cross-sectional empirical research based on data from the American states (McCormick and Tollison, 1981; Shughart and Tollison, 1985; Crain and Tollison, 1991) and on the Organization of Economic Cooperation and Development (OECD) countries (Mueller and Murrell, 1986) have illustrated many such results. For example, larger legislatures have been shown to be more costly environments in which to lobby, as well as bicameral legislatures with more disparate house and senate sizes (McCormick and Tollison, 1981).

Bureaucracy

Bureaucracy, in a sense, constitutes a fourth branch of government. The public choice approach to bureaucratic behavior has evolved over time, dating from Niskanen's (1971) seminal work on the subject. In *Bureaucracy and Representative Government*, Niskanen argued that because of its superior information, a bureau had greater bargaining power with regard to its budget than did the bureau's

oversight committee. Thus, the economic content in this approach is that the bureau maximizes its budget subject to all-or-none demand curve for its output, and this budget tends to be about twice as large as it "ought" to be (under the assumption of linearity). Much of the subsequent work on the economic theory of bureaucracy has been in this tradition. Wintrobe (1997) offers a masterful summary of these developments.[12]

It is worth noting, however, that different bureaus may reflect differing circumstances. For example, Niskanen wrote on the basis of his experience in the U.S. Department of Defense. He also later moved away from the budget-maximizing model and allowed the possibility that bureaus may pursue the maximization of the discretionary budget, in which case excessive bureau outputs disappear (Niskanen, 1975). Nonetheless, Weingast and Moran (1985) offered an alternative to Niskanen's theory, which predicts that the oversight committee (the principal) has most of the relevant bargaining power, including the ability to remove or to hamper the career of the bureau head (the agent). They tested this theory successfully with data concerning the Federal Trade Commission (FTC).

The issue raised in this debate is an important one. Are government bureaus out of control and bloated in size or are they merely docile agents following the commands of voters as expressed through their elected representatives on the relevant committees? The Weingast and Moran approach suggests that political incentives should be compatible between the legislature and the bureaucrat. The legislator observes a particular political trade-off in the election. Imposing that trade-off on his bureaucratic agent is in the legislator's self-interest. That is, the bureaucrat's role is to transfer wealth or to implement legislation and policy in the direction of the legislator's preferred trade-off. In this approach, bureaucracy is not out of control but is closely monitored and controlled by Congress. Bureaucrats who cannot be made to behave in accordance with the legislature's wishes are moved out of power.

The agent-principal problem is an economic problem. The principal is a residual claimant who holds an "ownership" right in the activities that his agent performs. The problem of the principal is to devise contractual and monitoring arrangements so that his interest is reflected in the labors of the agent. This stylized economic setting has stimulated a great deal of economic research and interest among economists because it obviously applies to many activities in an economy, such as the corporation, the labor union, the not-for-profit firm, and so on.

The agency problem has also had an impact on the economic theory of regulation and legislation. The issue can be explained as follows. A bureau head, say a regulatory bureau head, is the agent. Members of Congress serving on an oversight committee are the principals. The members of Congress evaluate and set political trade-offs by reading their election returns. The issue is how effective the

---

[12] Niskanen's heavy use of conventional price theory in presenting his theory of bureaucracy should be noted here.

politicians are in seeing to it that the bureaus under their jurisdiction make the appropriate political trade-offs and transfers.

As suggested above, the answer, in an emerging literature pioneered by Weingast and Moran (1983), appears to be that bureaus are quite attuned to the preferences of their overseers. Weingast has studied the FTC and the Securities and Exchange Commission (SEC), and in both cases he found strong support for such a hypothesis. Contrary to the common impression, then, government agencies do not appear to have a lot of discretion or to be out of the control of the voters through their elected representatives. They appear to heed the electoral trade-offs perceived by their political overseers when it comes to supplying wealth transfers and public policies.

These same principles of bureaucratic behavior also apply across countries in an international context. Wintrobe (1997) makes this clear in his survey article. Nonetheless, "international organizations" per se may represent a particularly nettlesome case of agencies "out of control" (Frey and Gygi, 1990). The moral of such analyses is simply that the relevant controls on the behavior of international bureaucrats are much laxer than those on their domestic counterparts. Hence, their carpets are thicker, and their lunches are longer and more expensive.

What is at stake here for students of regulation and government is to pierce the black box of bureaucracy and understand its inner workings better. How does one explain the process of economic regulation and, more generally, bureaucratic performance? The agent—principal framework offers a sensible route by which to develop a better understanding of such issues. Moreover, the agent—principal framework represents modern economic theory at work in public choice analysis. Subject to the costs of monitoring bureaucratic behavior, legislators are able to influence the goals and purposes of public policies in directions that maximize their reelection prospects.

Voters

So far, it is clear that the major components of democratic government can be fruitfully approached using economic methods. It is tempting to stop here and rest my case. However, voters represent a basic unit of public choice analysis because voters are the ones who convey the property rights to the rational agents in the foregoing analysis that empower these actors to run the government. Unfortunately, the behavior of voters in public choice analysis has been characterized as being only loosely related to the operation of thick rationality. Fortunately, there is a fairly easy resolution for this problem.

Public choice analysts customarily discuss voting behavior in terms of the paradox of voting; that is, on straight economic grounds (a comparison of the personal costs and benefits of voting) voting is not worthwhile yet turnouts in most elections are nontrivial. Hence, voting behavior is rationalized as consumption-type rather than investment-type behavior. People vote, for example, to express their patriotic duty rather than to express their self-interest in legislation. In contrast with other

parts of public choice theory in which behavior is modeled with maximizing, self-interested agents at the helm, the economic role of voters is comparatively unarticulated in the conventional version of public choice theory. In the standard approach, voters maximize utility rather than narrow economic self-interest, so that their behavior in the ballot box is less predictable. Needless to say, this is a weakness of public choice theory, wherein rational economic agents are assumed to gain their property rights to run the government from unpredictable voters. There are two basic routes out of this problem.

First, Stigler (1972), in particular, has questioned the consumption approach to understanding voter behavior. He argued that in politics a little more or a little less plurality matters. In this world, votes will matter to politicians and parties at the margin, and they will invest rationally in a supply of votes in order to have an impact on political and legislative outcomes. In such an instance, the paradox of voting is a moot issue. Interest groups will invest in a supply of votes for politicians in exchange for a higher probability of seeing a favorite bill passed. Such investments will be made on cost—benefit grounds—e.g., if it takes 1% more plurality to ensure the power to put a bill through, the interest group will compare the costs of turning out voters in this amount with the benefits of the legislation. In such a way voting behavior can be incorporated into the economic theory of government. In other words, the management of votes supplied by interest groups provides an alternative way to view the voting process, a way that is consistent with the general drift of the economic theory of legislation.

Second, the Stigler approach has not had much impact on the literature. Rather, an alternative argument is made. Although the investment motive is weak, this does not challenge the rational choice model. Voters are rationally ignorant after all, which opens up opportunities for interest groups. In other words, the standard concentrated benefits/diffused costs model of interest-group legislation rests on the rational ignorance and abstention of voters. Otherwise, such legislation would not be possible. In this more plausible approach to voting behavior, the rational choice model is seen to be consistent with and strongly complementary to the interest-group theory of government. Moreover, this latter theory of voter behavior applies across countries, so that there is no difficulty in generalizing this aspect of public choice analysis to an international context.

Conclusion

It is thus fairly easy to see how economic methodology permeates the modern theory of public choice. In each case examined above, the use of economic methods leads to a general result; that is, it leads to an organizing principle that offers an explanation for the behavior of a particular set of governmental actors. Moreover, in each case there is empirical support for the economic approach as outlined.

Obviously, I have only touched upon modern public choice analysis lightly. My examples are meant to be explanatory and illustrative and not at all comprehensive in covering modern public choice analysis. Needless to say, other scholars work in

other public choice traditions, and the purpose here is not to slight these traditions. Modern public choice analysis has a unified methodology regardless of whether the analyst adheres to an interest-group approach to explaining government (as I do) or to some other approach. This methodology finds its origin and home in the maximizing paradigm of modern economics. Public choice analysis descended from economic analysis, so that when asked about the influence of economics on public choice, I find it reasonable to answer, is there any other kind?

References

Anderson, G.M., Shughart, W.F., and Tollison, R.D. (1989). "On the incentives of judges to enforce legislative wealth transfers." *Journal of Law and Economics*, 32, 215–228.

Anderson, G.M. and Tollison, R.D. (1991a). "Congressional influence and new deal spending, 1933–1939." *Journal of Law and Economics*, 34, 161–175.

Anderson, G.M. and Tollison, R.D. (1991b). "Political influence on civil war mortality rates: The electoral college as a battlefield." *Defence Economics*, 2, 219–234.

Brams, S.J. and Davis, M.D. (1974). "The 3/2's rule in presidential campaigning." *American Political Science Review*, March, 113–134.

Cohn, J. (1999). "Irrational exhuberance." *The New Republic*, 25, 25–31.

Colatoni, C.S., Levesque, T.J., and Ordeshook, P.C. (1975). "Campaign resource allocation under the electoral college." *American Political Science Review*, March, 141–160.

Crain, W.M., Shughart, W.F., and Tollison, R.D. (1988). "Legislative majorities as nonsalvageable assets." *Southern Economic Journal*, 55, 303–314.

Crain, W.M. and Tollison, R.D. (1979). "Constitutional change in a interest-group perspective." *Journal of Legal Studies*, 8, 165–175.

Crain, W.M. and Tollison, R.D. (1991). "The price of influence in an interest-group economy." *Rationality and Society*, 15, 437–449.

Frey, B.S. and Gygi, B. (1990). "The political economy of international organizations." *Aubenwintschaft*, 45, 371–394.

Green, D.P. and Shapiro, I. (1996). Pathologies of Rational Choice Theory: A Critique of Applications in Political Science. New Haven: Yale Press.

Grier, K.B., McDonald, M., and Tollison, R.D. (1995). "Electoral politics and the executive veto." *Economic Inquiry*, 33, 427–440.

Klein, B. and Leffler, K.B. (1981). "The role of market forces in assuring contractual performance." *Journal of Political Economy*, 89, 615–641.

Landes, W. and Posner, R. (1975). "The independent judiciary in an interest-group perspective." *Journal of Law and Economics*, 18, 875–901.

Marvel, H.P. (1977). "Factory regulation: A reinterpretation of early English experience." *Journal of Law and Economics*, 20, 379–402.

McCormick, R.E. and Tollison, R.D. (1978). "Legislatures as unions." *Journal of Political Economy*, 86, 63–78.

McCormick, R.F. and Tollison, R.D. (1981). *Politicians, legislation, and the economy*. Boston: Kluwer.

Moe, T.M. and Caldwell, M. (1994). "The institutional foundations of democratic government: A comparison of presidential and parliamentary systems." *Journal of Institutional and Theoretical Economics*, 150, 171–195.

Mueller, D.C. and Murrell, P. (1986). "Interest groups and the size of government." *Public Choice*, 48, 125–145.

Neustadt, R.E. (1960). *Presidential Power*. New York: Wiley.

Niskanen, W.A. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine Atherton.

Niskanen, W.A. (1975). "Bureaucrats and politicians." *Journal of Law and Economics*, 18, 617–644.

Posner, R.A. (1986). *Economic Analysis of the Law*. Boston: Little Brown.

Riker, W.H. (1962). *The Theory of Political Coalitions*. New Haven: Yale Press.

Rowley, C.K. (2000). "Budget deficits and the size of government in the U.K. and the U.S.: A public choice perspective on the Thatcher and Reagan years", in K.C. Chrystal and R. Pennant-Rea (eds.) *Public Choice Analysis of Economic Policy*. London: St. Martin.

Shepsle, K.A. and Weingast, B.R. (1981). "Structure-induced equilibrium and legislative choice." *Public Choice*, 37(3), 503–519.

Shughart, W.F. and Tollison, R.D. (1985). "Corporate chartering: An exploration in the economics of legal charge." *Economic Inquiry*, 23, 585–599.

Silberberg, E. (1975). *Principles of Microeconomics*. Englewood Cliffs: Prentice Hall.

Stigler, G.J. (1972). "Economic competition and political competition." *Public Choice*, 13, 91–106.

Stigler, G.J. (1976). "The sizes of legislatures." *Journal of Legal Studies*, 5, 1–16.

Straayer, J.A. (1973). *American State and Local Government*. Columbus, OH: Merrill.

Thaler, R.H. (1992). The Winners Curse: Paradoxes and Anolomies of Economic Life. New York: Macmillan.

Weck-Hanneman, H. (1990). "Protectionism in direct democracy." *Journal of Institutional and Theoretical Economics*, 146, 389–418.

Weingast, B.R. and Moran, M.J. (1983). "Bureaucratic discretion or congressional control?" *Journal of Political Economy*, 91, 765–800.

Wessel, D. (2001). "The legal DNA of good economies." *Wall Street Journal*, September 6, 1.

Wintrobe, R. (1997). "Modern bureaucratic theory", in D.C. Mueller (ed.) *Perspectives on Public Choice: A Handbook*. Cambridge: Cambridge Press, pp. 429–454.

Wright, G. (1974). "The political economy of new deal spending: An econometric analysis." *Review of Economics and Statistics*, 56, 30–38.

Chapter 14
The Perspective of Philosophy[1]

Hartmut Kliemt

Introduction

The ascent of public choice theory[2], including the public choice society and its journal, is clearly one of the great success stories of post-war social and political theory. Nevertheless, to look at it from the point of view of philosophy poses a task of critical appraisal rather than uncritical applause. Public choice theory as interpreted here essentially forms an economic approach to politics and to public law (the most comprehensive presentation of this view being Mueller, 2003). Therefore I shall focus on the role of economics' core assumption of opportunistically rational and selfish behavior in public choice. As shown in the first part, (1) the views expressed by some classics of philosophy are at the root of modern public choice theory. Beyond that, some of them should be taken seriously as systematic contributions to present discussions about public choice in general and the role of the basic behavioral assumption of the rational economic man in particular. (2) On the basis of the discussion of the first part, the second assesses the public choice account of constitutional democracy in terms of individual rational choice. Philosophically promising amendments from the "neo-classical repair shop" that might conceivably solve some of the "paradoxes" of the standard economic approach to constitutional democracy are discussed. (3) The third part summarizes and concludes what may be seen as a "philosopher cum economist's" view of public choice.

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 235–244.

[2] In my discussion of the relationships between public choice and philosophy I shall generally use the term "public choice" for "public choice theory" as well as for the thing itself (but often I shall also differentiate explicitly between the two).

From Classics to Moderns

Knave-Proof Institutions

David Hume said that in politics "every man must be supposed a knave" (Hume, 1985, VI, 42). Many modern public choice theorists, in particular James M. Buchanan, endorse the Humean view. Yet, "it appears somewhat strange, that a maxim should be true in *politics* which is false in *fact*" (Hume, 1985, VI, 42–43). The Humean suggestion seems incompatible with the fundamental methodological norm that the facts should be accepted for what they are and be reported truthfully (Albert, 1985). The criterion for evaluating the quality of an empirical theory is its truth. In empirical science per se no good is to be expected from using models that are known to be false. So, can we defend the advice of *counter*-factually adopting a point of view according to which everybody is looked at *as if* being a knave?

A typical modern public choice theorist will tend to interpret the term "knave" as the "opportunistically rational economic man".[3]

Doing so he may want to adapt Dennis Robertson's (Robertson, 1956) social functional justification of economics in general to public choice theory in particular and suggest that the task of normative public choice theory is to "economize on love". In particular if we engage the task of "mechanism design" (Myerson, 1991) in public choice under the behavioral assumption that all individuals behave like "(rational) knaves" then we should end up with proposals that can work properly without requiring "love" as an "input". The mechanisms will work even if all pursue their narrow self-interest in opportunistic ways. They will therefore be "knave-proof" in the specific sense of allowing for opportunistically rational behavior throughout and make it "the interest of even bad men, to act for the public good" (Hume, 1985, III, 16).

Contrary to what is often suggested, the assumption of opportunistically rational selfish behavior does not necessarily lead to the worst case scenario in politics. In politics the worst is to be expected from misguided unselfishness (Arendt, 1951). From the supporters of Hitler to those of Stalin, to the attack on the world trade center, human unselfishness and sacrifice rather than opportunism and selfishness are at the root of much of large-scale evil.[4] Therefore, if the Humean principle is seen to aim at worst case scenarios, we cannot interpret it as implying opportunistically rational behavior in the sense of the standard model of the rational economic man. To put it otherwise, the public choice theorist who is relying on the model of

---

[3] For Hume's own justification see Hume, (1985). *Essays. Moral, Political and Literary*. Indianapolis: Liberty Fund.

[4] There have been public choice analyses of the selfish rent-seeking motives of people who profited, for instance, from the persecution of the Jews by the Nazis; yet these arguments, though very useful and illuminating; cannot explain the full scale of engagement for the "good cause". For the rent seeking argument see Anderson, and Tollison, 1993. "Wealth Maximization in Hell.:" 13. Fairfax, VA: Center for Study of Public Choice.

the rational economic man in her explanations of public choice processes may be too optimistic. Engaging the task of understanding "politics without romance" (Buchanan, 1979) by using the model of the rational economic man as a universal behavioral assumption may indeed systematically underestimate the evil influence of romanticism based on superstition and enthusiasm (Hume, 1985, X).

Though the assumption of opportunistically rational selfish behavior may not lead to the worst case scenario in politics, rendering an institution "knave-proof" may be the only way to eliminate the incentive of rational but risk-averse individuals who are not knaves to behave as if they were. Seen in this light it is not the case that individuals are selfishly seeking their own advantage—for example out of "greed"—it is rather their resentment or fear of being exploited by others that motivates their actions. Hobbes came close to such a view when insisting that the lack of certainty rather than any base motive on the side of actors in a state of nature (not necessarily a stateless situation but one that ensues beyond the limits of norms enforced by selective sanctions) induces them to engage in pre-emptive action against other individuals. As long as individuals are not certain that others will behave well—or that others have at least a selfish motive to behave well—the risk that others might not behave well justifies and induces rational actors to act in ways that they would deem unacceptable otherwise.

For instance, in a setting of 'collective good' individuals might be willing to contribute their own due, voluntarily. They would resist the temptation to choose the dominant free-riding strategy if they knew that others would not free-ride. For them what is a prisoner's dilemma game in objective or material payoffs is an assurance game in subjective payoffs. Their most preferred choice is not the unilateral exploitation of the good conduct of others. Yet being afraid that others might choose the strategy that is dominant in objective payoffs, risk-averse individuals might rationally choose the free-ride strategy themselves. More specifically, should such individuals know that all others or a "sufficient" number of them had contributed already, they themselves would prefer to contribute (see more or less in this vein Jasay, 1989; Margolis, 1982; Taylor, and Ward, 1982), but they strongly resent the risk of being exploited. In such a world "defection" would not anymore be a dominant alternative and the theoretical game paradigm would not be an n-person prisoner's dilemma but rather an n-person assurance game.[5]

It is a fundamental insight of public choice theory that the mere awareness that knavish behavior may exist, together with loss aversion (in the sense of (Kahneman, D. and Tversky, 1984)) may induce individuals who are not knaves to behave as if they were. If these individuals believe that too much institutional trust in the good conduct of people may in fact "invite" or induce others to behave badly, they may "react" with knavish acts to the anticipated knavish behavior of others (or the mere suspicion thereof). This is a factual behavioral assumption that in turn may justify

---

[5] Possibly, of course, an iterated prisoner's dilemma in which the dominance properties of defection are also altered and cooperation according to folk theorem logic may become sequentially rational may be studied; see, for a philosophically particularly useful study of this kind, Taylor, (1987). *The Possibility of Cooperation*. Cambridge: Cambridge University Press.

the use of the counter-factual assumption of opportunistically rational selfish behavior in institutional design. Such design may minimize the risk to trigger knavish behavior in otherwise well-intentioned individuals.

Normative public choice theory, at least in the Buchanan and Tullock (1962) variant, suggests that institutions be chosen so as to minimize risks of too much trust. However, insufficient trust may have negative effects too. Behaviorally, treating those who are not knaves as if they were, may crowd out their trustworthiness (Frey, 1997). Treating individuals as if they were knaves may be a major policy mistake in all realms in which behavior is not fully controllable. For instance, honesty in taxpaying may actually decrease if the controls on taxpaying behavior are tightened. In a business firm intrinsically motivated employees who are providing "trust goods" may be driven out if their superiors start to control their behavior too tightly by specific incentives, rewards and punishments. More generally speaking, social order could not exist without non-opportunistic behavior and without some amount of trust in individuals' good conduct. If public choice theory suggests a solution of the problem of social order that is in this respect mistaken, it may undermine the foundations of the very order on which public choice practice rests.

Public Choice Theory and the "Hobbesian Problem of Social Order"

Robinson Crusoe on his island is one of the heroes of modern economic story telling: "Let us look at an 'islandic' story about Robinson Crusoe and Man Friday which illustrates the 'Hobbesian' problem of social order" (Parsons, 1968).[6] Crusoe and Friday live "before public choice" (Buchanan, 1972). In that state of nature, Crusoe may legitimately take action against Friday (even take Friday's life) to eliminate any uncertainty originating from the mere presence of Friday. If Crusoe lets Friday live—as he in fact does in Defoe's original story—he must incur that risk for a reason as, for instance, his hope to gain from the division of labor and exchange as made viable by Friday's presence.

Once certainty prevails: Hobbes expects individuals to behave in ways that are quite the opposite of what we expect from rational opportunists. For instance, imagine that in a bilateral exchange Friday has acted as a first-mover and delivered his goods as agreed in a preceding mutual promise (Hardin, 1982). After Friday has acted there is no uncertainty anymore about Friday doing his part. Friday has executed his promise. Therefore—according to Hobbes—Crusoe as a second-mover is under an obligation to reciprocate. According to Hobbes, Crusoe must

---

[6] For an "Icelandic" story on anarchy and law one may turn to Njal's Saga. For an account of law without the state it is interesting to consult Benson, (1990). *The Enterprise of Law. Justice Without the State*. San Francisco: Pacific Research Institute for Public Policy.

do his due even in the absence of the "shadow of the future" (Axelrod, 1984); and more generally on the underlying folk theorem (Fudenberg, and Tirole, 1992). He must perform even without a rational incentive. However, without such an incentive, Crusoe's reciprocal act is not opportunistically rational (nor is it sub-game perfect in the sense of Selten, 1965, 1975). Yet it is, according to Hobbes, what we (legitimately) can expect from a second mover who is *not* uncertain about the first mover's actions.

Hobbes clearly smuggles in norm-oriented behavior here.[7] He not only believes that individuals are sometimes under an obligation not to act opportunistically but also thinks that they will in fact behave non-opportunistically sometimes. The emergence of social order becomes more intelligible by this Hobbesian concession but at the price of rendering a Hobbesian approach to norm compliant behavior less stringent and clearly incompatible with a strict economic and public choice approach.[8]

In a Hobbesian approach, some of the sociologists' claims that men would not always act opportunistically are already accepted. To avoid this, the hard-nosed economist and public choice theorist should turn to Benedikt de Spinoza rather than to Thomas Hobbes. Spinoza endorses an "economic" approach to norm-guided behavior according to which everybody will in each instance of choice act opportunistically rational. Because of their centrality for the relationship between philosophy and public choice Spinoza's views deserve to be cited at some length.[9]

Now it is a universal law of human nature that no one ever neglects anything which he judges to be good, except with the hope of gaining a greater good, or from the fear of a greater evil; nor does anyone endure an evil except for the sake of avoiding a greater evil, or gaining a greater good. That is, everyone will, of two good, choose that which he thinks the greater; and of two evils, that which he thinks the lesser. I say advisedly that which he thinks the greater or the lesser, for it does not necessarily follow that he judges right. This law is so deeply implanted in the human mind that it ought to be counted among the eternal truths and axioms.

As a necessary consequence of the principle just enunciated, no one can honestly forego the right which he has over all things, and in general no one will abide by his promises, unless under the fear of a greater evil, or the hope of a greater good. Hence though men make promises with all the appearances of good faith, and agree that they will keep to their engagement, no one can absolutely rely on another man's

---

[7] That even the folk theorem does so has been convincingly argued as well. See, Güth, et al., 1991 "On supergames and folk theorems: a conceptual analysis," in R. Selten (ed.) *Game Equilibrium Models. Morals, Methods, and Markets*. Berlin et al.: Springer, pp. 56–70.

[8] For a philosophical account of Hobbes that may be particularly interesting for the public choice theorist see Hampton, 1988. *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.

[9] Buchanan in his appendix to the calculus of consent which is devoted to the topic of "reading political philosophy" also emphasizes the importance of Spinoza, though for a different reason (see appendix to Buchanan, and Tullock (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press).

promise unless there is something behind it. Everyone has by nature a right to act deceitfully, and to break his compacts, unless he is restrained by the hope of some greater good, or the fear of some greater evil. (Spinoza, 1951, pp. 203–204)

This contains in a nutshell an economic theory of norm-guided behavior. In this theory, in particular, the norms of law are seen exclusively as external constraints to individual action. The opportunistically rational individual will show norm-compliant behavior if and only if in the instance of choice the expected future consequences of norm-compliant behavior are preferred to non-compliant behavior. The "should" attached to norms derives from self-interest and opportunism not from using the norm itself as a standard. Norms do not have any motivational force of their own but serve as instruments in predictions of sanctions (positive or negative).

The preceding view is clearly as mistaken as the "over-socialized" model of man according to which people follow internalized norms no matter what. Human beings can and sometimes do make opportunistically rational choices but we must not assume away the human faculty to "internalize" norms.[10] They sometimes follow norms because of an "intrinsic motivation" to do so. They can and do in fact adopt an "internal point of view to norms or rules". They can accept norms or rules as standards and guidance of their behavior rather than treating them merely as external constraints to which they try to adapt in an opportunistically rational fashion.[11]

Standard public choice theory has no room for an internal point of view to norms. However, without an adequate understanding of the faculty to adopt an internal point of view to norms we cannot adequately understand power and authority and thus an essential aspect of all public choice (excellent on this, Barry, 1981). The "easiness with which the many are governed by the few; and the implicit submission by which men resign their own sentiments and passions to those of their rulers" (Hume, 1985, IV, 32) cannot plausibly be explained without some element of non-opportunistic behavior. Convictions and beliefs are not merely indicating knowledge of what will happen but express "opinions" on what should be chosen to happen.

Since a wider model of human behavior than that of Spinoza or, for that matter standard public choice theory, is needed, economists and in particular public choice theorists run into problems. They cannot have it both ways: on one hand, bash the

---

[10] A rich literature on experiments on ultimatum bargaining shows this; see for an overview Roth, (1995). "Bargaining Experiments," in J. H. Kagel and A. E. Roth (eds.) *The Handbook of Experimental Economics*. Princeton: Princeton University Press, pp. 253–348. And originally the seminal paper Güth, et al. (1982). "An experimental analysis of ultimatum bargaining". *Journal of Economic Behavior and Organization*, 3: 367–388.

[11] The leading legal philosopher Herbert Hart developed his basically adequate theory of how a legal order works and can manage to exist in criticism of "economic" or Spinozist accounts. He criticized legal philosophers as John Austin in ways that directly apply as well to many views expressed by public choice theorists; see Hart, (1961). *The Concept of Law*. Oxford: Clarendon Press; Austin, John (1954). *The Province of Jurisprudence Determined*. London.

sociologists for their assumption of non-opportunistic behavior and on the other hand wheel in non-opportunistic behavior whenever that seems convenient.[12] Either we strive to explain each and every single act of norm-compliant behavior in terms of the expected causal consequences, in particular the future rewards and punishments that the single act is expected to bring about, or we have already given up the core assumption of the model of the economic man and thus are not any longer pursuing a purely economic approach to politics.

For instance, an economist who "explains" non-subgame perfect (non-opportunistic) behavior of an individual by demonstrating that it is in the self-interest of the actor to be so committed gives the game away. Drawing attention to the fact that it is in the selfish interest of the committed individual to become committed is no explanation at all. Using such an "explanation" in constitutional political economy amounts to the same as to explaining Crusoe's use of a can opener by pointing out that it is in Crusoe's self-interest to have one (Brennan, and Kliemt, 1990). But self-interest per se does not explain the existence of the tin opener. Without a can opener being there Crusoe cannot use it—likewise, without commitment power he, as a rational economic man, cannot commit.

More generally speaking, a rational actor cannot give up her rationality and opportunism merely because this would be in her self-interest. In game theory, which has taken this aspect of rationality most seriously and thereby to its extreme, cases in which rationality and self-interest are in conflict abound. For instance the whole problem of credible threats and promises emerges only because individuals cannot suspend their faculty to act opportunistically at will. Smuggling commitment power in where it is not part of "the rules of the game" amounts to violating the very premise on which the economists' own criticism of other approaches (in particular sociological) to human behavior is based (Kliemt, 1987 and 1993).[13] Still, if truth rather than coherence or elegance is the aim of theorizing, then a modified model of human behavior is needed. It must accommodate behavior that is boundedly rational in the sense of being subject to limits of information processing as well as in the sense of being intrinsically bound by or to norms.

The philosophical classics used "opinion" to accomplish this task. Hobbes, for instance, said: ". . . the power of the mighty hath no foundation but in the opinion and belief of the people. . ." (Hobbes, 1682/1990, p. 16). Even more significantly Hume spoke of "opinion of interest" and of "opinion of right" (Hume, 1985, IV, 33) as central determinants of political behavior. Turning to public choice theory's account of the essential institutions of constitutional democracy, and in particular of voting, it will become obvious that including the role of "opinion" and the related distinction between low and high cost situations leads to a fruitful modification of

---

[12] See also Güth, and Kliemt (1998). "Towards a fully indirect evolutionary approach". *Rationality and Society*, 10(3), 377–399. The paper programmatically indicates how both the shadow of the past and that of the future can be integrated into a single model systematically.

[13] Since the utility functions are formally also part of the rules it might be added that using appropriate modifications of the utility function whenever needed is ad hoc in an unacceptable way, too.

the basic behavioral model underlying public choice theory (see for early hints (Tullock, 1971) and generalizing (Brennan, and Buchanan, 1984), originally 1982; Brennan, and Lomasky, 1983; Brennan, Geoffrey and Lomasky, 1984; Brennan, Geoffrey and Loren Lomasky, 1985, Brennan, Geoffrey and Loren Lomasky, 1989).[14]

Constitutional Democracy and Public Choice

Constitutional democracy is commonly associated with the "voting procedures for selecting rules and rulers" on one hand and with the "rule of law and the protection of individual rights" on the other hand. Both of these central aspects of the political process in a constitutional democracy cannot be understood adequately if we stick to the standard neo-classical model of opportunistically rational and selfish behavior of individuals. Since the likelihood that a single vote will be decisive in a general election is smaller than the probability of being hit by an asteroid on the way to the voting booth, participation can hardly be explained in terms of expected causal effects on the collective outcome (see Tullock, 1967; Downs, 1957). Though policies impose significant costs on citizens, voters, in view of the low probability to be decisive, decide behind a "veil of insignificance" (Kliemt, 1986). So why do they participate at all?

From the Old to the New Public Choice Account of Voter Participation

As is well known, some public choice theorists argue that people go to the polls because they are extremely risk averse, because they systematically misperceive the probability that their own vote is decisive or because they fear damage for their general reputation if they are recognized as non-voters. Taking these factors into account, voter turn out at the polls might be somewhat higher than the mixed equilibrium to be expected otherwise (Palfrey and Rosenthal, 1984). However, the theoretically plausible values are significantly below any level of participation observed in the real world. To explain the low rates of abstention that we, as a matter of fact, do observe, it is necessary to wheel in intrinsic motivation or a kind of preference for voting as formed by internalization of values and norms.[15]

---

[14] The distinction between high and low costs has been around in ethical theory, and in common sense ethics, since antiquity. According to this view we may legitimately demand and expect the other-regarding or generally norm-oriented behavior if costs are low while high costs of performance form an excuse and a motivationally sufficient reason to violate normative requirements (see Heyd, (1982). *Supererogation. Its Status in Ethical Theory*. Cambridge et al.: Cambridge University Press; Urmson, (1958). "Saints and heroes", in I. Melden (ed.) *Essays in Moral Philosophy*. Seattle/London: University of Washington Press, pp. 198 ff.).

[15] If people are asked why they go to the polls they say that they feel obliged to act that way.

The account of expressive voting behavior that involves merely insignificant costs of alternative voter choices, as fully developed in Geoffrey Brennan's and Loren Lomasky's "Democracy and Decision" (Brennan, and Lomasky, 1994, Brennan, and Lomasky, 1989), may in the last resort come directly from the shelf of what has been ironically characterized as "the neo-classical repair shop". Nevertheless, if we intend to stick reasonably close to the economists' home turf, it is a major step towards a more convincing and coherent (modified) neo-classical account of voting. In Brennan's and Lomasky's explanation of voting behavior, causal effects of voting on public outcomes or choices are separated from the private consequences for the choosing individual. For instance the private consequence of fulfilling "her duty" towards the poor will emerge with certainty if a rich woman votes for higher tax redistribution favoring the poor. She can "consume" the "warm glow of fulfilling her duty" while her vote itself will be insignificant for whether or not she will indeed have to pay higher taxes. Casting her vote according to the interests of the rich, she would forego with certainty the satisfaction derived from acting according to a standard of morals that puts a high self-esteem-premium on unselfish behavior while gaining only a completely insignificant decrease in the likelihood of having to pay higher taxes. So, in view of the intrinsic motivation to maintain the self-image as a moral person, the rich woman will rationally choose to go to the polls—provided that it is not too inconvenient and that she does not expect her own vote to be decisive with any significant likelihood.

The spatial model of voting according to which individuals vote for those parties that are located closest to their own position on the ideological field will in principle stand after this modification. Distance in the "opinion" space matters. But that it matters is not explained by the expected causal effects of individual voting behavior on the collective outcome. Voting being "expressive" the spatial model must be seen from a completely different angle. Though it is still framed in terms of individually rational choice, the utilities and expectations on which the whole argument is based are much closer to the traditional philosophical views of a government. Elections resemble opinion polls in which individuals express their views of what they deem right or wrong for the collectivity at large rather than strategic situations in which individuals seek to bring about collective results by exerting some strategic influence.

If we look at things with a philosophical eye, we may note here more generally that it is perhaps harder to explain how particular interests can in fact be pursued at the polls effectively than to explain how generalized moral aims are pursued through voting. That this difficulty does indeed exist seems to be supported by the observation that practically all particular interests in public presentation camouflage themselves as general ones. One could, of course, argue here that this serves the purpose of increasing information costs for those who might otherwise be stirred up to take action against the pursuit of special interests. However, a much more plausible view might be that individuals at the polls will tend to prefer what they think is "right" for the public at large rather than what is in their direct particular interest. Those who are trying to further particular interests at the polls must be in a position to argue that the public at large should endorse those interests from a moral or impartial point of

view. This is necessary, not because the public must be deceived, but rather to mobilize and coordinate the particular interest to be served (this is also an amendment to another great book of public choice, namely Olson, (1965)). The particular interest must be presented as a legitimate interest to those who have the particular interest but would not pursue it without moral license at the polls.

If responses of "morally motivated" individuals to this "trick" or "re-framing" lead to results that would also be predicted as the outcome of rational strategic choices in the classical model then this—in view of the veil of insignificance and its effects—is an astonishing phenomenon requiring some explanation or other. What is seen as an explanation in public choice theory (an *explanans*) may become something to be explained (an *explanandrum*) if expressive voting and the like are factored in. For example, if farmers are particularly successful in serving their particular interests then the strength of these interests and the relative ease with which they are organized is perhaps not the explanation of political phenomena but rather the phenomenon to be explained. If voting is low cost and if expressive components will in all likelihood dominate strategic considerations, why is it that groups seem to be able to act strategically as groups at the polls? The vote of the farmer is as insignificant as that of any consumer. So, why would a farmer (or more generally producers) go to the polls and vote according to their private interests while their customers cannot co-ordinate in the same way?

Of course, information costs do play a role but they cannot explain in full what we observe. More easy access to information on the income than on the spending side may be a factor but there must be some motive to act on information even at low costs. Within the expressive voting framework as laid out here some ideas come forward easily. For instance there is a "natural" claim to legitimacy going along with producers' interests and this must be taken into account in public choice theory. For it is not interest per se that is guiding individuals in their choices as voters but rather their "opinion" that they are expressing "legitimate interests" at the polls.

Now, one might be tempted to say that it is not important whether interests directly and strategically dominate voting behavior or only indirectly through expressing opinion of interest or right as filtered by criteria of legitimacy. But this is quite obviously mistaken if the dimension of influencing voter choice is taken into account. For traditional models have no room for legitimacy concerns as a *factual* determinant of choice. The philosophical or methodological criticism of public choice theory here is not a normative one but rather that the theory leads to inadequate explanations, predictions and also misguided policy recommendations if it comes to influencing "opinion" as expressed in voter choice.

Is There "Public" Choice at All?

In the proper sense of the term only individuals can choose. Individuals can make their choices "in public" but it is not the "public" that chooses. We only describe it—quite misleadingly—as a choosing entity. For instance, imagine again the simplest case of a general majority election selecting one of two candidates

who run for an office. In such a case we tend to describe our act of voting as "choosing one of the candidates". However, literally speaking no person is performing the act of "choosing the candidate" (as opposed to, say, choosing one's bride). The only act that is performed is that of choosing to vote for or against a certain candidate—or not to vote at all.

Nobody makes public choices: they rather emerge, as Buchanan has always insisted (Buchanan, 1954). If it comes to overall results of the process there is no difference between market and political choice. The results are emergent rather than chosen. Of course, we can look at the market, too, *as if* it were a choosing entity. We then describe outcomes that emerge on the market as "collective or public choices". We can always describe something as if it were something else.[16] But it seems strange and certainly not fruitful to look at the collective results of market interaction as collective choices. But then it should seem strange as well to describe the outcomes of politics as collectively chosen. At least, we use a "systematically misleading expression" (Ryle, 1931–32) whenever we speak of "public choice".[17] The older term "non-market decision making" may have been more suitable than the term "public choice" in this regard.

To insist that there are only individual and no collective choices as emerging either from market or non-market decision making does not deny that there are other relevant differences. For instance, in voting the collective result is clearly intended. Voters say yes or no with the vision that the collective result is to be brought about while on the market the collective result is typically neither on the agenda nor on individuals' minds. Unless consumers do things like buying "politically correct coffee" they are not intentionally trying to bring about or influence overall market results. Nevertheless the differences in the costs of choice making rather than an alleged difference between collective as opposed to individual choices is essential to "public choice". Cost asymmetries are also essential for deriving a more adequate account of the very existence of the rules that transform individual choices into public outcomes.

How do Public Choice Rules Emerge and Manage to Exist?

As has been noted before, constitutional democracy besides being characterized by voting is also characterized by rule of law and the respect of individual rights. Again the model of opportunistically rational, selfish behavior is in no way sufficient to explain how rules manage to exist or are brought into existence. The Spinozist or standard public choice account cannot conceivably be true in full. As a specific example, think of the behavior of judges. True enough there are aspects of

---

[16] An instructive argument against the philosophical practice of the same kind can be found in Jasay, (1996). "Justice as something else". *The Cato Journal*, 16(2):161–173.

[17] It may also be noted in passing that, aspects like cycling in voting are game theoretically and strategically interesting with respect to the working properties of systems of rules but not with respect to some concept of a genuinely collective choice of results.

the judges' behavior that can be accounted for in terms of opportunistically rational and selfish behavior. Judges do not like it if higher courts reject or revise their verdicts since this may be detrimental to their career prospects. They economize on their own effort etc. In short, judges like everybody else can and do act opportunistically rational some of the time. They respond to incentives and external constraints. Yet it is quite strange to assume that they would not often simply try to form their opinion on what the law requires and act upon this. Besides being extrinsically motivated by external incentives judges—at least some, some of the time, are intrinsically motivated to find according to the law.

In principal and agent relationships we generally utilize institutionally the notion that individuals can be expected to show genuine norm-following behavior if the costs of doing so are relatively minor. Hume, whose directly related remarks on opinion were already mentioned, uses the instructive example of the emperor of Rome who could treat his subjects like "wild beasts" but had to lead his Praetorian guards "like men" by their "opinion". Since the physical strength of the emperor would in no way suffice to overcome resistance of any greater number of individuals it is surprising, initially, to see how easily he can govern his subjects. Yet, if we take into account how the division of labor is extended to the enforcement and creation of rules, most of the surprise vanishes. Those who are policing rules like the Praetorian guards should in general bear only minor costs when doing so. At the same time the emperor must try to feed both their "opinion of interest"—namely that on average they will be better off obeying his orders—as well as their "opinion of right"—namely that he legitimately claims obedience.

The Praetorians must be induced to stick to opinions that bring so-called "power conferring rules" into existence. Rules create power if they are accepted as standards and applied by sufficiently many sufficiently influential individuals in a non-opportunistic manner to single out those who are seen as authorized to give orders or to enact rules. The mighty are mighty because sufficiently many, sufficiently influential individuals tend to follow their orders in a non-opportunistic manner. At the same time others obey the mighty because they are mighty. The latter obey out of a rational calculus in particular because they are afraid of encountering sanctions imposed voluntarily by those who follow orders to impose the sanctions.

To understand the role of self-interest here it is important to note that a rule can be adopted out of self-interest while the within rule choices are afterwards made by applying the rule as a standard rather than according to a calculation of the consequences of the singular acts. That individuals apply rules in a non-opportunistic manner, which goes beyond case by case considerations, changes the character of social interaction fundamentally even if the genuine rule following behavior is shown only if the rules are in the interest of those who follow them and the opportunity costs of that behavior are relatively low. Constitutions and the individual rights they may confer exist because power conferring rules are followed.[18] Like all collective

---

[18] Otherwise a variant of the ancient problem "quis custodiet custodes ipsos?" would emerge or "who is going to control the controllers themselves?"

structures they must be created by individual behavior and must be explained as results of individual choices which cannot be opportunistically rational throughout.

Conclusions

In all social contexts the same "entities" (human beings) are acting. Therefore predictions and explanations of human market behavior like buying and selling should be based on the same model as predictions and explanations of human political behavior like voting or rent-seeking. Likewise the so-called within rule choices should be explained according to the same logic as the choices of rules (Brennan, and Buchanan, 1985). Norm-guided behavior should be explained by the same model as opportunistically rational behavior—the one or the other triggered by different circumstances. If markets can fail as instruments of coordinating individual choices then politics can fail as an instrument of coordination as well.

Human behavior is to a large extent purposeful action under constraints regardless of the institutional settings or natural circumstances imposing the constraints. It is, however, very doubtful whether the standard model of opportunistically rational choice that is egoistically motivated by "objective payoffs" leads to a realistic account of purposeful action under constraints. All people are sometimes opportunistic. Perhaps all are sometimes even behaving like knaves. Still, social interaction would be quite a different ball game if all people were behaving opportunistically all the time. If that were true neither the very existence of a normative or legal order nor such central public choice behavior as voting in a large electorate could be adequately explained. People at least sometimes simply do what they deem right rather than advantageous for themselves. Without taking this into account our view of "public choice" is distorted and the very possibility of "public choice" as a rule-guided enterprise of determining collective outcomes is a riddle.

The modifications of the neo-classical behavioral model suggested in particular by Brennan and Lomasky on the basis of older ideas of Buchanan and Tullock may lead us a long way. But as we know and as experimental economics indicates in some high-stake experiments, humans do show non-selfish behavior even in situations of high costs.[19] The assumption of forward-looking rational choice throughout is, however, doubtful in any event. In the end, going to the neo-classical repair shop will not rescue public choice theory as a purely economic theory in the traditional sense of a discipline based on rational choice throughout. In public choice it may be even less rewarding than in more traditional branches of economics since from a philosophical methodological point of view it seems even more doubtful that without major revisions of the underlying behavioral model, public choice theory can in fact lead to true explanations of public choice.

---

[19] For ways of incorporating even this into neo-classical economics see for instance Frank, (1987). "If homo economicus could choose his own utility function, would he want one with a conscience?" *The American Economic Review*, 77(4): 593–604. Frank, (1988). *The Passions within Reason: Prisoner's Dilemmas and the Strategic Role of the Emotions*. New York: W.W. Norton.

On the other hand, traditional public choice theory very well serves our philosophical desire for a rational account of society. An account of the workings of society in general and of public choice in particular in terms of opportunistically rational choice is the only one that really appeals to our favorite self-image as rational beings.[20] James M. Buchanan's economic philosophy represents this self-image in its most mature "contractarian public choice" variant (see generally Buchanan, 1999, in particular Buchanan, 1999a, b). But psychology and cognitive science as well as experimental economics (Kagel and Roth, 1995), are knocking at the door of public choice and political philosophy (Frohlich, and Oppenheimer, 1992). Once this door opens in full, new insights into public choice and how it might be improved in theory and practice will emerge. But we should see to it that philosophy and public choice theory do not learn the wrong things from each other (Sugden, 2004).

References

Albert, H. (1985). *Treatise on Critical Reason*. Princeton: Princeton University Press.

Anderson, G. M. and R. D. Tollison. (1993). *Wealth Maximization in Hell*. Volume 13. Fairfax, VA: Center for Study of Public Choice.

Arendt, H. (1951). *The Origins of Totalitarianism*. New York: Harcourt.

Austin, J. (1954). *The Province of Jurisprudence Determined*. London: Macmillan.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Barry, N. (1981). *An Introduction to Modern Political Theory*. London: Macmillan.

Baumol, W. J. (1972). *Economic Theory and Operations Analysis*. London: Prentice Hall.

Benson, B. (1990). *The Enterprise of Law. Justice Without the State*. San Francisco: Pacific Research Institute for Public Policy.

Brennan, H. G. and J. M. Buchanan. (1984). "Voter choice: Evaluating political alternatives." *American Behavioral Scientist*, 28(2), 185–201.

Brennan, H. G. and J. M. Buchanan. (1985). *The Reason of Rules*. Cambridge: Cambridge University Press.

Brennan, H. G. and K. Hartmut. (1990). "Logo logic." *Journal of Constitutional Political Economy*, 1(1), 125–127.

Brennan, H. G. and L. E. Lomasky. (1983). "Institutional aspects of 'merit goods' analysis." *Finanzarchiv, N.F.*, 41(2), 183–206.

Brennan, H. G. and L. E. Lomasky (1984). "Inefficient unaminity." *Journal of Applied Philosophy*, 1(1), 151–163.

---

[20] It should be noted that to the first meeting of the small group that formed the nucleus of what later on was to become the Public Choice Society, besides obvious suspects like J.M. Buchanan, J. Coleman, W. Mitchell, M. Olson, V. Ostrom, W. Riker, and G. Tullock, the philosopher J. Rawls was invited and actually participated in the second meeting. Rawls always insisted that for normative political theory we need to know how political institutions in fact work and that therefore philosophy has to come down from the skies. Still, he took the counterfactual economic model as acquired from his reading of the economic literature as factual at least in part and in that sense just switched skies. In Rawls' "A Theory of Justice" there are many traces of the introductory text by Baumol, (1972). *Economic Theory and Operations Analysis*. London: Prentice Hall International.

Brennan, H. G. and L. E. Lomasky. (1985). "The impartial spectator goes to Washington." *Economics and Philosophy*, 1, 189–211.

Brennan, H. G. and L. E. Lomasky. (1989). *Large Numbers, Small Costs—Politics and Process—New Essays in Democratic Thought*. Cambridge: Cambridge University Press.

Brennan, H. G. and L. E. Lomasky. (1994). *Democracy and Decision*. Cambridge: Cambridge University Press.

Buchanan, J. M. (1954). "Individual choice in voting and the market." *Journal of Political Economy*, 62, 334–343.

Buchanan, J. M. (1972). "Before public Choice," in G. Tullock. (ed.) *Explorations in the Theory of Anarchy*. Blacksburg, VA.: Center for Study of Public Choice, pp. 27–37.

Buchanan, J. M. (1979). "Politics without romance: A sketch of positive public choice theory and its normative implications." *IHS Journal, Zeitschrift des Instituts für Höhere Studien—Wien*, 3, B1–B11.

Buchanan, J. M. (1999). *The Collected Works of James M. Buchanan*. Indianapolis: Liberty Fund.

Buchanan, J. M. (1999a). *The Logical Foundations of Constitutional Liberty*. Indianapolis: Liberty Fund.

Buchanan, J. M. (1999b). *Moral Science and Moral Order*. Indianapolis: Liberty Fund.

Buchanan, J. M. and Gordon T. (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

de Jasay, A. (1989). *Social Contract—Free Ride*. Oxford: Oxford University Press.

de Jasay, A. (1996). "Justice as something else." *The Cato Journal*, 16(2), 161–173.

de Spinoza, B. (1951). *A Theologico-Political Treatise. A Political Treatise*. New York: Dover.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Frank, R. (1987). "If homo economicus could choose his own utility function, would he want one with a conscience?" *The American Economic Review*, 77(4), 593–604.

Frank, R. (1988). *The Passions within Reason: Prisoner's Dilemmas and the Strategic Role of the Emotions*. New York: Norton.

Frey, B. S. (1997). *Not Just For the Money. An Economic Theory of Personal Motivation*. Cheltenham: Edward Elgar Publishing.

Frohlich, N. and J. A. Oppenheimer. (1992). *Choosing Justice. An Experimantal Approach to Ethical Theory*. Berkeley: University of California Press.

Fudenberg, D. and T. Jean. (1992). *Game Theory*. Cambridge, London: The MIT Press (Massachusetts Institute of Technology).

Güth, W. and K. Hartmut. (1998). "Towards a fully indirect evolutionary approach." *Rationality and Society*, 10(3), 377–399.

Güth, W., S. Rolf, and S. Bernd. (1982). "An experimental analysis of ultimatum bargaining." *Journal of Economic Behavior and Organization*, 3, 367–388.

Güth, W., L. Wolfgang, and S. Günther. (1991). "On supergames and folk theorems: A conceptual analysis," in R. Selten (ed.) *Game Equilibrium Models. Morals, Methods, and Markets*. Berlin Heidelberg New York: Springer, pp. 56–70.

Hampton, J. (1988). *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.

Hardin, R. (1982). "Exchange theory on strategic basis." *Social Science Information*, 2, 251

Hart, H. L. A. (1961). *The Concept of Law*. Oxford: Clarendon Press.

Heyd, D. (1982). *Supererogation. Its Status in Ethical Theory*. Cambridge: Cambridge University Press.

Hobbes, T. (1682/1990). *Behemoth or The Long Parliament*. Chicago: Chicago University Press.

Hume, D. (1985). *Essays. Moral, Political and Literary*. Indianapolis: Liberty Fund.

Kagel, J. H. and A. E. Roth (eds.) (1995). *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Kahneman, D. and T. Amos. (1984). "Choices, values and frames." *American Psychologist*, 39, 341–350.

Kliemt, H. (1986). "The veil of insignificance." *European Journal of Political Economy*, 2(3), 333–344.

Kliemt, H. (1987). "The reason of rules and the rule of reason." *Critica*, 19, 43–86.

Kliemt, H. (1993). "Constitutional commitments," in P. H. Herder Dorneich (ed.) *Jahrbuch für Neuere Politische Ökonomie*. pp. 145–173.

Margolis, H. (1982). *Selfishness, Altruism, and Rationality. A Theory of Social Choice*. Cambridge: Cambridge University Press.

Mueller, D. C. (2003). *Public Choice III*. Cambridge: Cambridge University Press.

Myerson, R. B. (1991). *Game Theory—Analysis of Conflict*. Cambridge, London: Harvard University Press.

Olson, M. (1965). *The Logic of Collective Action*. Cambridge, MA: Cambridge University Press.

Palfrey, T. R. and R. Howard. (1984). "Participation and the provision of discrete public goods: a strategic analysis." *Journal of Public Economics*, 24, 171–193.

Parsons, T. (1968). *Utilitarianism. Sociological Thought*. New York: Macmillan.

Robertson, D. H. (1956). *Economic Commentaries*. London: Staples Press.

Roth, A. E. (1995). "Bargaining experiments," in John H. Kagel and Alvin E. Roth (eds.) *The Handbook of Experimental Economics*. Princeton: Princeton University Press, 253–348.

Ryle, G. (1931–32). "Systematically misleading expressions." *Proceedings of the Aristotelian Society*, 32, 139–170.

Selten, R. (1965). "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit." *Zeitschrift für die gesamte Staatswissenschaft*, 12, 301–24, 667–689.

Selten, R. (1975). "Reexamination of the perfectness concept for equilibrium in extensive games." *International Journal of Game Theory*, 4, 25–55.

Sugden, R. (2004). "What philosophy and public choice should not learn from one another", *The American Journal of Economics and Sociology*, Vol. 63, 207–211.

Taylor, M. (1987). *The Possibility of Cooperation*. Cambridge: Cambridge University Press.

Taylor, M. and Hugh W. (1982). "Chickens, whales, and lumpy goods: alternative models of public-goods provisions." *Political Studies*, 30, 350–370.

Tullock, G. (1967). *Toward a Mathematics of Politics*. Ann Arbor: University of Michigan Press.

Tullock, G. (1971). "The charity of the uncharitable." *Western Economic Journal*, 9, 379–392.

Urmson, J. O. (1958). "Saints and heroes," in I. Melden (ed.) *Essays in Moral Philosophy*. Seattle: University of Washington Press, 198 ff.

Chapter 15
The Perspective of Law[1]

Francesco Parisi

Introduction

The enlightened conception of the separation of powers holds that law should be made by the legislature, interpreted by the judiciary, and enforced by the executive branch of government. Public choice theory provides a solid foundation for the appraisal of this traditional formula. The findings of public choice theory, while supporting much of the traditional wisdom, pose several challenges to the theoretical foundations of these constitutional principles. In the following pages, I shall revisit these important questions considering the issue of institutional design through the lenses of public choice theory.

Sources of Law and the Institutional Design of Law-Making

According to a fundamental principle of constitutional design, powers should be allocated to the branch and level of government or society that can best exercise them. This principle can be applied to the question of law-making in order to select sources of law that will exploit the comparative advantage of different legal and social institutions in the production of legal rules.

I consider three main criteria for evaluating the relative advantages of alternative sources of law, focusing on the political economy of production of ordinary (i.e., non-Constitutional in nature) law.

Minimization of Agency Problems

First, the mechanisms for law creation should be able to reflect the underlying preferences of the individuals subject to the law.

---

[1] This paper is a reprint of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published by Kluwer Academic Publishers, Volume I, 214–235.

For the case of the political processes of law formation, this requires the choice of collective decision making procedures that will promote the alignment of the incentives of political representatives and the incentives of the represented citizens. In the presence of perfect incentive alignment, agency problems in political representation will disappear.

Likewise, in an ideal world judge-made law should approximate the rules that private parties would have chosen if engaging in an *ex ante* choice of applicable law. This claim, known as the efficiency of the common law hypothesis, constitutes an important premise of the law and economics movement. According to this hypothesis, the common law (i.e., judge-made law) is the result of an effort—conscious or not—to induce efficient outcomes. The same proponents of this hypothesis suggest that common law rules enjoy a comparative advantage over legislation in the creation of efficient legal rules because of the evolutionary selection of common law rules induced by adversarial adjudication.

The case of customary law is quite different from those of the other sources of law. Customary law avoids the interface of third party decision makers (such as legislators and judges) and is directly derived from the observation of the behavioral choices of individuals in society. In a customary law setting the group of lawmakers coincides with the subjects of the law and agency problems are generally absent from such a process of law formation. In the following discussion, we will consider the different groups of problems that however affect the process of customary law formation.

In all the above cases, the institutional design of lawmaking should induce incentive alignment in order to minimize the extent of agency problems, with a minimization of rent seeking and a resulting optimal supply of public good.

Minimization of Rulemaking Costs

The second criterion for evaluating alternative sources of law is that of cost minimization of collective decision-making. According to this criterion, the mechanisms for law creation should be chosen in order to minimize the transaction costs of collective decision making and political bargaining.

This cost minimization problem involves the evaluation of two different costs, namely (a) the direct costs of decision-making, such as the costs of reaching a majoritarian consensus in a political context, or the cost of litigation or adjudication in a judicial context; and (b) indirect or external costs, such as the cost imposed on a minority group by the rules chosen by a majority coalition.

Different levels of transaction costs of types (a) and (b) are inherent in the different processes of law formation.

Direct Costs of Lawmaking

In a legislative process, individual preferences are captured by the collective decision making process through the imperfect interface of political representation. Bargaining among political representatives is costly, due to the strategic behavior

of a large number bargaining (i.e., free riding, hold ups, and other collective action problems) and the absence of legal enforcement mechanisms for political bargains. In this dimension, lawmaking through politics is likely to impose the highest level of transaction costs among the alternative sources of law that we consider.

Transaction and information costs are also present in the case of judge made law. The process of judicial creation of legal rules faces the obvious constraint given by the costly access to information regarding alternative legal rules. If we analogize the law-making process to a production process in the marketplace, the common law may indeed appear as a quite inefficient production process. The common law process, when shifting some of the law making functions to the judiciary, entrusts courts with the task of conceiving and refining legal rules while adjudicating specific cases.

From a production point of view, such a process foregoes the economies of scale and scope that might be exploited by a specialized legislative process. On the other hand, the common law process, by relying on the adversarial efforts of the parties, utilizes information available to the parties. Parties have direct information on the costs and benefits of alternative rules and courts may be regarded as having an informational advantage over central legislative bodies, given the opportunity of judges to infer the litigants' preferences from the choices they make during the case.

Courts have a further informational advantage in observing the revealed preference of the parties with respect to applicable law. Modern legal systems generally provide a set of default rules that apply if the parties fail to choose alternative provisions to govern their relationship. When parties opt out of the default rules (through *ex ante* choice of differing provisions or *ex ante* choice of law), they reveal their preferences over alternative legal rules. If courts observe a large number of parties routinely opting out of the default rules, it becomes evident that such rules have failed their cost-minimization task under the circumstances and do not approximate the will of the majority of the contracting parties. In these cases, courts would have a comparative informational advantage over legislators in designing and revising default legal rules.

For the case of customary law, we should distinguish two distinct costs: (a) the cost of decentralized creation of a customary legal rule; and (b) the cost of judicial finding of an existing rule of customary law.

The costs of creation are relatively minimal. Most rules of customary law are derived from the observation of widespread practice followed by individuals in society. In this context, customary rules are a costless byproduct of the economic and social interactions of individuals in society. Such practices are not being carried out with the objective of giving birth to binding rules of customary law and the legal recognition of such practices as binding customs adds no cost to the activities involved.

The costs for courts to identify a rule of customary law may, however, be considerable. Customs are intangible sources of law and their content does not enjoy any objective articulation in written law. The identification of custom thus requires knowledge of past practice and investigation of the beliefs shared by those who engaged in the practice: a process that can be costly and difficult to carry out.

A point of advantage of customary sources of law is related to the fact that custom is formed through the independent action of individuals in society, without the need for their express agreement to the emerging rule. Since most rules of custom require a very high level of participation without necessitating a unanimous consensus, hold up problems and other transaction-associated costs are generally avoided in the formation of customary legal rules. No single individual in society can prevent the emergence and recognition of a general custom.

External Costs of Lawmaking

The various sources of law also have different levels of external costs. As public choice theory has shown, in the case of political decision making direct costs and external costs of lawmaking are negatively correlated (Buchanan and Tullock, 1962). The tradeoff between direct and external costs is easily illustrated by the consideration of the two limit cases of unanimity and dictatorship in a voting context. If deliberations require a unanimity vote, the risk of external costs disappears, since unanimity gives every voter a veto power against undesired proposals. Transaction costs are instead very high under a unanimity rule. In the opposite case of dictatorship, the risk of external costs is much higher, since a dictator can single-handedly impose costs on all other individuals. Conversely, the direct costs of lawmaking are lowest under dictatorship, given that no consensus and political bargaining is necessary under a dictatorial decision rule.

Analogous tradeoffs between direct and external costs exist for the other sources of law, but the content and interpretation of such costs differ substantially in each case.

Thus, for example, rules of customary law require a very high level of participation and consensus. This reduces the risk of external costs imposed on unwilling minorities, but, as a result of such high threshold of required participation, customary laws are relatively slow in their emergence and evolution.

In evaluating the various sources of law, it will be necessary to give careful consideration to the different performance of alternative lawmaking processes from the vantage point of this criterion of cost minimization.

Stability and Transitivity of Collective Outcomes

The third problem of institutional design is to minimize the cost of instability and ensure rational and transitive collective choices. As it has been observed in the literature (e.g., Cooter, 2000; Stearns, 1994; Parisi, 1997), when political cooperation fails and the lawmaking mechanisms do not generate Condorcet winners, several legal institutions and doctrines come to the rescue to minimize instability and select among cyclical alternatives. In particular, Cooter (2000) explains how

democratic constitutions pursue these goals of stability by separating powers among the branches of government, by guaranteeing individual rights, and creating a framework of competition for political office. Parisi (1998) considers the role of logrolling as an instrument of stability in a legislative setting. With reference to judge-made law, Stearns (1994) considers the role of standing doctrines and *stare decisis* as evolved institutions aimed at reducing instability in the absence of a Condorcet majority consensus. In the different setting of customary law, Parisi (1997) discusses the process of formation and evolution of customary law, unveiling the ability of customary law to generate stable rules in different game-theoretical situations.

Law through Politics: the Political Economy of Legislation

Comparative differences in legal systems often reflect different ideologies and conceptions of the political economy of lawmaking. In recent years, all countries of the modern world have been giving written statutes increasingly greater importance among the sources of law. The supremacy of written law over other sources of legal order is not, however, a universal characteristic of all modern legal systems.

Comparative legal scholars usually distinguish between civil law and common law systems. The distinction is based on a dichotomous conception of legal traditions. Systems of the civil law tradition give greater weight to written and statutory sources of law. Generally speaking, these systems are historically derived from a legal tradition that recognized the authority of a comprehensive body of written law (e.g., the Roman *Corpus Juris*) and were not relying on the casuistic evolution of case-by-case decision making in the absence of a coherent skeleton of codified law. This dichotomous distinction, while useful as a preliminary classificatory tool, should not be overestimated.

During the last several decades, legal systems of the world have converged toward a middle ground. In the civil law tradition, the dogmas of supremacy of legislation over case-law have gradually given way to a more balanced conception of sources of law, where statutes and case-law more or less happily coexist with one another. Likewise, in the common law tradition, the proliferation of legislative intervention has gradually corroded the traditional dominance of judge-made sources.

Lawmaking and Political Representation

During the nineteenth century, the enlightened conception of democratic governance and a renewed trust in political decision-making fostered an increased importance of statutory law. Ideals of democratic legislation gradually replaced the historic conception of statutory law as a written articulation of laws of a higher and older origin.

Laws were not the mere expression of preexisting natural or fundamental rights, but rather they were the primary source, if not the sole origin, of individual rights. Rights were derived from laws, rather than laws being derived for the protection of individual rights. Legislative bodies were making (i.e., creating) law as opposed to finding (i.e., recognizing) pre-existing legal norms. With the exception of some minimal Constitutional constraints on law making, national Parliaments and Congresses acted as sovereign lawmakers. Such unbounded legislative powers were justified by the alleged function of legislative organs as faithful agents and political representatives of the people.

The unfolding of history has, however, revealed the true face of democratic decision making and the limits of the mechanisms of political representation in lawmaking.

There are two theoretically distinct problems that affect the mechanisms of political representation. These problems have become the respective focus of several important contributions in the public choice and social choice literature.

Within the public choice tradition, we learn that political representatives are agents of the individuals they represent. Such political representation is often affected by pervasive agency problems. The correction of these problems requires the choice of collective decision making procedures that promote the alignment of the incentives of political representatives with the incentives of the represented citizens, or else an effective monitoring and accountability of political agents. If incentives are effectively aligned, agency problems of this type do not affect political representation. Much of the public choice and the constitutional design literature addresses these fundamental problems.

The second problem emerges even in the absence of agency problems in representation. This problem is one of selection of appropriate criteria for aggregating individual preferences. If the interests of politicians align with the interests of the people whom they represent, politics can be viewed as a framework for bargaining among political agents of the various factions in society. The question is whether political bargaining can successfully yield a consensus among the various political factions, such that political outcomes can be legitimately and unambiguously identified with the "will of the people".

As the social choice literature has often pointed out, even if we contemplate a world of perfect incentive alignment between political representatives and the represented citizens (i.e., even if we assume away agency problems in political representation), there is no assurance that the mechanisms of law creation are responsive to the underlying preferences of individuals in society.

Political Decision Making and the Market for Votes

One of the main insights from social choice theory is that the correlation between preference and choice is weaker for groups than for individuals (Shubik, 1987: 124). According to Arrow's (1951) possibility theorem, it may indeed be too much

to expect methods of collective decision making to be, at the same time, rational and egalitarian. Arrow's theorem shows that any social decision that is adopted must violate at least one of six self-evident axioms of normative political theory, commonly described by the following terms: range, universal domain, unanimity, non-dictatorship, independence of irrelevant alternatives, and rationality. Arrow's negative conclusion and its various corollaries pose a dramatic threat to the legitimacy of political decisions. The observation that the likelihood of cycling majorities decreases in situations where the number of decision-makers is much greater than the number of choices does not affect the practical relevance of Arrow's analysis applied to the political process, where the large number of decision-makers is actually concentrated into a restricted number of interest groups with "group" votes.

The heart of Arrow's theorem states that there are no non-dictatorial rules or procedures for collective decision-making that reflect the combined preferences of voters to a consistent collective outcome (Arrow, 1951). The implications of Arrow's theorem concern the existence of cyclical majorities which are capable of repealing any resolution that has been adopted previously. Parisi (1998) suggests that, if all voters are allowed to enter into binding agreements over the policy outcome to be adopted by the majority coalition, collective preferences in a multi-dimensional policy space will be transitive as long as individual preferences are single-peaked.

This intuition runs contrary to the common thought in public and social choice theory (Bernholz, 1973; Miller, 1977; Schwartz, 1977). Most of the literature on the stability implications of log-rolling considers log-rolling in the context of bargaining for the formation of coalitions where side-payments are only instruments for entering the majority coalition, and no side-payments are made by those who are not part of the majority.

The political reality is often different from that contemplated by these scholars. Bargaining is certainly permitted even between minority and majority voters, with exchanges taking place among all coalitions.

As shown by Parisi (1998), if we allow for a broader role for bargaining and side-payments and contemplate binding and enforceable political bargains across different coalitions, the results would be quite different.

One Man One Vote, and the Market for Votes

In situations in which no strong political consensus is reached on a given issue, intransitivity may result. Intransitivity implies that a different order in the decision making process may affect the outcome and that any winning coalition may be undermined by the reintroduction of an alternative it previously defeated. The structure of the voting process does not allow the cycle to be broken by looking at the intensity of voters' preferences. The outcome is arbitrarily determined by the order of motions, with no guarantee that the ultimate result will yield a higher level

of social welfare than that potentially afforded by any other defeated policy alternative. The inability of the democratic process to capture the intensity of the voters' preferences is a by-product of the generally espoused principle that every individual is entitled to one—and only one—vote. The "one man, one vote" rule is further explained by the fact that individual voters do not face the opportunity cost of casting their vote. Whether their preference is strong or weak, voters will cast their vote in favor of their favored option. Even if specifically designed to allow voters to indicate the intensity of voters' preferences, the voting ballot could not possibly capture such intensity. Absent a mechanism to extract the true intensity of their preferences, individual voters would tend to overstate their preference in order to maximize the impact of their votes.

Democracy gives equal weight to all votes when they are counted, regardless of how strongly the voters feel about the issue. In this way, numerically equal groups have equal political say in the process. However, if the distribution of sentiments on an issue is not symmetrical, and the minority holds strong preferences, the outcome would be inefficient. By introducing the possibility of bargaining and vote-trading in the process, the intensity of preferences will be reflected in the decision-making process. With bargaining and side-payments, the "one man, one vote" rule would provide the initial entitlement for each voter-trader. The exchange mechanism would then reveal the relative strength of individual preferences.[2]

Political bargaining may provide a solution to the intensity problem, and at the same time correct for the cyclicality problem. Politicians know well that under certain conditions the outcome may depend on the sequence of decisions and therefore on agenda-setting. For example, in a situation with intransitive preferences, the agenda-setter may influence the process in favor of his preferred policy by determining the sequence of decisions and introducing his preferred policy in the last motion. This point is well known among public choice theorists and legal practitioners. Judge Easterbrook (1983) has noted that, "someone with control of the agenda can manipulate the choice so that the legislature adopts proposals that only a minority support." (See also Levine and Plott, 1977; Long and Rose-Ackerman, 1982.)

Agenda-setting increases the internal predictability of the outcome for those who are involved in the process and have full information about it. Legislators sharing similar information on their respective prospects will have an opportunity to bargain under conditions of symmetric information, trading votes for issues on which they hold weak preferences in exchange for votes on issues which have more value for them. Economic theory teaches us that bargaining between politicians will continue until the marginal utility of gaining one vote on a certain issue equals the marginal cost of giving up one vote for another issue. We should further consider whether the outcome selected by majorities in such an environment of costless and enforceable political bargaining maximizes the combined welfare of the platforms. Parisi (1998) suggests that both stability and efficiency will be obtained through

---

[2] From an efficiency perspective, in fact, weight should be given to intensive preferences.

bargaining, as long as the exchanges are enforceable and relatively costless to carry out. The implications are very far-reaching and can be articulated in the following two propositions:

(a) If the conditions for the Coase theorem are present for all voters (i.e., if political agents can enter into coalition contracts with other agents and such contracts are enforceable as stipulated, unless mutually dissolved by all parties), the composition of the initial majority coalition is irrelevant for the policy outcome.
(b) If the Coase theorem holds, voters' preferences are strictly concave, and vote-exchange agreements are enforceable; cycling in a multi-dimensional policy space is excluded.

Thus, if political bargains are possible at no cost and political agreements are enforceable, the resulting political equilibrium will be unique and will occur at a point of social maximum. Any point other than the global maximum will be unstable, since there will always be enough surplus to allow for side payments to voters in exchange for policy concessions. Once the socially optimal point is reached, there will be no opportunity to destabilize the policy arrangement.

Enforcing Political Bargains

The above conclusions rest on a quite formidable assumption. Political agreements are assumed to be enforceable, just like ordinary contracts in a private law setting. This implies that any attempt to modify the bargained-for policy choice would have to be accepted by all parties—contracts can be resolved only with the consent of the contracting parties.

More generally, the Coasian bargaining assumption implies that all political promises are enforceable. In this setting, minority voters can join the coalition and have a marginal effect on the policy outcome by out-bidding or "bribing" all members of the pre-existing majority. With enforceable contracts members of a majority coalition cannot cheat on each other. Collectively, they will entertain offers made by minority voters who will influence the status quo with their side payments, but they will not be able to break away from an existing coalition, since such coalition agreements can be modified only with the consent of all parties. Finally, as well-known in the collective action literature, groups with lower collective action costs can be more effective in gathering the most effective bribe, as public choice theory has extensively shown in the various rent-seeking contexts (Olson, 1965; Kahn, 1990).

As pointed out by Cooter (2000), in real politics bargaining is afflicted by a special problem that is usually absent in private contracts. Political agents are limited in the extent to which they can enter into enforceable political bargains. For example, coalitions' agreements are only good until a new coalition is formed. Likewise, there is no way to bind future voting decisions in a logrolling context, or to constrain the choices of future office-holders.

In a traditional contract setting, a contractual agreement can be undone only with the consent of all original contracting parties. Conversely, in informal political agreements, any political agent can betray the original agreement and destabilize the original coalition. There are no direct legal remedies to render such agreements enforceable.

In general, no agreement between current members of Congress regarding future voting is enforceable under the law. For example, majority deliberations cannot be perpetuated prohibiting future amendments or requiring that such amendments be carried out with a super-majority vote.

Legislators sometimes have to be creative to make contracts enforceable in the real-world market for votes. In several occasions political actors attempt to signal the enforceability of their bargains (and ensure its influence against the status quo) in a future vote by publicly stating that they would not "go back and undo the things that they pledged that they would do." In other situations, the repeat interaction among politicians may induce the fulfillment of some political bargains, thus facilitating political cooperation.

However, the general non enforceability of political bargains limits the deals that can be struck among political representatives and among branches of government.

Limits of the Politics-Like-Markets Analogues

In real politics, legislative and political bodies seldom work like markets. Cooter (2000) points out three main challenges to the politics-like-market analogy. The first reason why political markets do not work like ordinary markets is that the value of a legislator's vote often depends upon how the other legislators' vote. There are pervasive externalities and resulting free riding incentives in political action. The second reason is that real life politics has too many political actors for each one to bargain with everyone else. Unlike the atomistic marketplace of traditional economics, bilateral negotiations would be prohibitively expensive in real life politics. Third, Cooter points out the diffuse hostility to a rationalization of politics as a market for consensus. Ordinary citizens with little information about legislative bargains would resist any institutionalization of political bargaining, objecting to their representatives participating in open logrolling.

Indeed, a full analysis of the politics-like-market analogy cannot be accomplished in a vacuum, but rather must be exposed to the reality of democratic politics. The following corrollaries are discussed by Parisi (1998) and are illustrative in this regard: (1) on issue bundling; (2) on free riding and bargaining failures; and (3) on agency problems and the political dilemma.

Issue Bundling

In the real world of politics, transaction costs are present. As a way to minimize the effect of transaction costs, policy "packages" are traded and voted upon in the usual course of dealing. Political deals are indeed characterized by a bundling of

different issues. Congressional voting normally requires a binomial vote on legislation supplying a bundle of bargained-for provisions. And House and Senate rules do not prevent amendments that are unrelated to the subject matter of the bill at issue (Dixon, 1985; Riggs, 1973). For example, when Congress sent President Clinton the 1997 appropriations bill that funds White House operations, it included legislative riders ranging from the repeal of a law allowing states to share in federal price discounts from the pharmaceutical industry, to a provision to clarify that imports manufactured by indentured child labor are prohibited (Rogers, 1997). Although an item veto might have enabled President Clinton to remove particular items from such bundles, that power was ruled to be unconstitutional by the Supreme Court of the United States (Penny, 1997).

From an efficiency perspective, bundling—just like tying in a commodity market—may generate suboptimal outcomes. In order for a vote exchange process to work at its best, all dimensions of the policy space should be the potential object of bargaining and trade. Bundling reduces the dimensions of the bargaining space. At the limit, all policy dimensions may collapse down to a two-dimensional policy space, limiting the domain of the optimization process.

In an ideal world with no transaction costs, no bundling should exist, in order to maximize the beneficial functioning of the political market. In a real world with positive transaction costs, a positive amount of bundling is to be expected and is part of the global optimization process. Elhauge (1991: 31) has noted that where there is issue bundling, "diffuse interests can be systematically under-represented even if voters face no collective action problem." But the market will adjust to reach the optimal tradeoffs between the savings on transaction costs and the inefficiencies of tying.

Free Riding and Bargaining Failures

An important assumption of the Coase theorem is the absence of transaction costs. A costless transaction requires the absence of strategic behavior in the bargaining process. This condition is highly problematic in the context of multi-party voting. The opportunity for individual strategic behavior is elevated where two polar groups seek compromise. In the real-world market for votes, the term "triangulation" has been to describe the result of efforts to legislate in the middle ground between ideological extremes, where vote-trading transaction costs are high (Broder (1997), attributing the "triangulation" concept to former Clinton-advisor Dick Morris).

All cyclicality problems require the presence of at least three voters. Bargaining among three voters in a two dimensional space is highly sensitive to free riding and other forms of strategic preference revelation.

If we think of this triangular situation in a spatial voting setting, we can realize that any movement in the policy space will generate benefits or losses for at least two parties. In the great majority of cases, all three parties will be affected by a

potential policy change. Under such conditions, any bargaining carried out by one voter has the potential of creating side benefits for another voter. Any policy change "purchased" by one voter is potentially a free good (or a free bad) for another voter. In a three-party bargaining, voters are thus faced with a collective action problem. The problem is exacerbated by an increase in the number of voters. In a multi-voter setting, strategic behavior may indeed plague the bargaining process.

The collective action problem described above is not different from any other free riding problem in a Coasian setting. Olson (1997) has discussed the collective action problem in the context of a Coasian bargaining, questioning the practical validity of the Coasian proposition in a multi-party context. If the object of one individual's bargaining generates a benefit to other individuals who are not involved in the bargain, what is obtained through the bargaining of one individual creates a positive externality to other individuals. Thus the incentives to undertake the bargaining may be seriously undermined. Every individual wishes to be the free rider, having somebody else pay the price of the common good. Thus, similar to any public good situation, there will be a sub-optimal level of bargaining for the common interest.

Agency Problems and the Political Dilemma

The analysis of the hypothetical market for votes considered in this article takes the will of the voters as a given. Further analysis should consider the effect of agency problems in the bargaining mechanism. In the real world of politics, most collective decisions are carried out by political representatives, who undertake collective decisions as agents of the represented individuals. Political representation is often undermined by serious agency problems. Public choice theory provides ample analysis of the factors of such incentive misalignment, including (a) rational abstention; (b) rational ignorance; and (c) regulatory capture, resulting in special interest legislation. Such discrepancies are most visible when an agency problem in political representation occurs at the margin of a crucial vote.

If bargaining is carried out in the absence of agency problems, the bargaining result maximizes the voters' utility, as illustrated above. But where the bargaining is carried out by interested representatives, the opportunity is present for departures from the optimality outcome described above.

In general terms, if market mechanisms are allowed to operate in political contexts, the collective decision-making mechanism is lubricated. In the absence of representation failures, the collective outcome will approximate the allocative outcome of a competitive market. If bargaining is carried out by agents whose underlying incentives differ from those of their principals, the market mechanism may generate greater discrepancies between the ideal and the real political outcomes, including the fact that agents may be induced to abandon their principals' core values.

The Cost of Legislation

The absence of legal enforcement mechanisms in political contracts increases transaction costs and often represents an insurmountable obstacle to political cooperation. According to Cooter (2000), the institutional design of lawmaking should promote institutional arrangements that minimize the transaction costs of political bargaining.

With respect to legislation as a source of law, the previous sections have shown that the politics-like-markets analogue risk overlooking the difficulties of correcting political failures through political bargaining. The existence of effective exchange mechanisms within politics accentuates the features of the underlying political system. In a world of good politics, it allows for better outcomes. In a world of political failures, it may exacerbate the existing problems.

In a world where political bargaining exists, however, the existence of enforcement mechanisms within politics will promote stability and reduce costly intransitivity of collective outcomes.

As discussed above, stability cannot be used as a proxy for efficiency. It is indeed well-known in the social and public choice literature that a "Condorcet winner" can at times be inefficient, but at least it can always be trusted to satisfy the preferences of the majority of voting individuals. Absent mechanisms to induce voters to reveal the true intensity of their preferences, democratic legislative systems cannot improve on Condorcet winners and should maintain rules that allow such alternatives to prevail when they exist.

If Condorcet winners do not exist, the method and sequence of voting (i.e., agenda setting, etc.) determines the political outcome. In these cases, as Cooter (2000) aptly puts it "democratic politics becomes a contest, not to satisfy the preferences of a unique majority, but to determine which majority's preferences will be satisfied". In these situations, institutions should be designed in order to minimize the welfare costs of voting intransitivity and instability. The existence of enforceable contractual mechanisms for political exchange may be a valuable instrument of stability.

These results confirm Buchanan and Tullock's (1962: 153) important observation that "with all side payments prohibited, there is no assurance that collective action will be taken in the most productive way." Likewise, they provide a conjectural solution to Tullock's (1981) puzzle as to why there is so much stability in the political process.

Common Law and the Economics of Judicial Lawmaking

Judge-made law and doctrines of *stare decisis* have varying degrees of importance in the various legal systems of the world. As well known, there is a substantial historical difference between the role played by precedents in the common law and

civil law traditions. In early legal systems, written legislation was utilized with great parsimony and great weight was given to the customary sources of law. Occasionally, sources of customary law were unable to provide solutions to emerging legal issues and to satisfy the changing needs of society. In these cases, precedents were recognized and followed as a matter of outright necessity.

With the gradual expansion of statutory law, the recognition of precedents as sources of law was no longer a practical necessity. In these settings, contemporary legal systems have developed a variety of doctrines to determine the effective role of judicial decisions in the presence of legislation and to guarantee an effective separation between these two branches of government. Principles of separation of powers provide the Constitutional foundations for balancing the institutional roles played by courts and legislators.

Separation of Powers and the Independent Judiciary

One key feature of most Constitutional systems of the Western legal tradition is the principle of separation of powers, with particular importance placed on an independent judiciary to ensure the fair adjudication of law. The principle of separation of powers implies that, unlike the legislative and executive branches, most judges are (or should be) systematically shielded from political or economic influence.

As a matter of institutional design, the independence of judges can be achieved by either turning the judiciary into a bureaucracy-type institution, where judges are selected and promoted according to pre-established standards of performance on the bench, or through political appointment with life-tenure, with the consequent elimination of any ties with the appointing political body (Cooter, 2000). The first approach is generally followed by most Civil law jurisdictions, while the second approach finds its typical incarnation in the Federal judiciary of the United States.

Landes and Posner (1975) examine the effect of the independent judiciary on lobbying, the de jure system of interest group purchase of legislative policy. Economic analysis of the role of the courts shows how an independent judiciary can make viable a governmental process that emphasizes interest group participation in policy formation. By enforcing laws validly passed, even in a previous legislative session, the judiciary ensures integrity in the Constitutional process by imposing prohibitive costs on public interest purchase of judicial decisions.

Landes and Posner work from the perspective of interest group analysis, pointing out that interest groups will not purchase policy programs if they cannot assume that the desired policy will last. In the absence of an enforceable contract, some other power must provide that guarantee. In the first instance, the high transaction costs are associated with the cumbersome process of enacting legislation supply stability. Accordingly, if courts, which must enforce legislation, were agents of the Congress in session, the legislature could cheaply arrange a de facto repeal by asking its courts to rewrite legislation by taking advantage of interpretive leeway. If, on the other hand, the judiciary is independent and interprets legislation in

accordance with the enacting Congress' intent, it then supports, rather than interferes with, purchase of legislation by interest groups. However, the independent judiciary may also impose costs by declaring the law unconstitutional or interpreting it in a way that reduces gains to the group that paid for the law.

Some questions have been raised in the literature regarding the actual level of independence of the judiciary. After all, in the U.S. legal system Congress does have powers, such as appropriations of funds, creation of new judgeships, and rewriting jurisdiction by which they might compel judicial acquiescence. However, self-interested judges can increase their independence by rendering predictable decisions in accord with the original meaning of the statute. According to Landes and Posner (1975) this increases the value of the judiciary to the current legislature because its members know that the courts will enforce the contracts they make. According to the authors, the structure of the judiciary—life tenure, rules against ex parte contact, and impeachment for accepting bribes—also prevents interest groups from influencing judges directly.

Landes and Posner (1975) further explore the positive implications of the economic theory of the judiciary. First, they consider the case of 'dependent' judiciaries, such as those established in specialized agencies, making a consistent finding that such entities are established when the chance of judicial nullification of political and legislative bargains is high. Mild judicial review allows the agencies to keep the terms of a particular legislative deal, but since that review is not wholly effective, administrative adjudication becomes far less consistent over time, as would be expected from a dependent judiciary that is not protected from shifts in political emphasis. The authors further consider the effect the economic system of legislation coupled with an independent judiciary, has on the form of interest group legislation. Building upon public choice models of rent-seeking, the authors suggest that interest groups purchase legislation that does not require substantial annual appropriations. Legislative rents that require yearly Congressional funding are quickly dissipated, as it would be necessary to lobby each new Congress to support the program, the costs of which eat into the net present value of the legislation for its intended beneficiaries. Since the judiciary cannot help to enforce new annual appropriations, interest groups tend not to purchase such legislation.

The law and economics literature also considers the role of the independent judiciary in enforcing the Constitution. According to Landes and Posner (1975) judicial independence has two purposes in this context. First, it establishes ground rules for a system of interest group politics enforced by the independent judiciary. Second, the Constitution confers specific protective legislation on powerful interests groups willing to purchase such a provision in their favor. For example, broad interpretation of the First Amendment is a form of protective legislation purchased by publishers as an interest group. The Constitution's purpose, supported by the independent judiciary, is to protect groups powerful enough to obtain a constitutional provision or a special interest legislation in their favor.

The conclusions reached by this literature stress that the independent judiciary is an essential element in the observed struggle among interest groups, which is a major component of political practice. Although the judiciary is a critical player in

this process, it, by itself, is not 'political,' but rather is above politics because it fulfills its role by enforcing the legislative deals of earlier legislatures, not because it has special wisdom, integrity, morality, or commitment to principle.

The Hypothesis that the Common Law is Efficient

To the extent to which judicial bodies are independent from political forces and shielded from interest group pressure, the process of judicial lawmaking can be considered immune from the collective decision making failures considered in the previous section.

In this setting, law and economics scholars formulated a conjecture, known as the efficiency of the common law hypothesis—according to which the common law (i.e., judge-made law) is the result of an effort—conscious or not—to induce efficient outcomes. This hypothesis was first intimated by Coase (1960) and was later systematized and greatly extended by Posner in numerous books and articles: common law rules attempt to allocate resources in either a Pareto or Kaldor-Hicks efficient manner. Further, common law rules are said to enjoy a comparative advantage over legislation in fulfilling this task because of the evolutionary selection of common law rules through adjudication. Several important contributions provide the foundations for this claim; the scholars who have advanced theories in support of the hypothesis are, however, often in disagreement as to its conceptual basis.

Rubin (1977) provides an important contribution to the emerging efficiency of the common law hypothesis. He maintains that the efficiency of the common law is best explained by an evolutionary model in which parties will more likely litigate inefficient rules than efficient ones. The pressure for the common law to evolve to efficiency, he argues, rests on the desire of parties to create precedent because they have an interest in similar cases in future. Rubin thus considers three basic situations: where both parties are interested in creating precedent, where only one party is interested in creating precedent, and where neither party has such an interest.

Where both parties have an interest in similar cases in future, and the current legal rule is inefficient, the party held liable will have an incentive to force litigation. Parties will continue to use the courts until the rule is changed. If the current rule is efficient, however, there is no incentive to change it, so it will remain in force. Where only one party has an interest in similar cases in future, the incentive to litigate will depend on the allocation of liability. If liability falls on a repeat player, litigation is likely to occur, whereas the other party would have no incentive to litigate. As a result, precedents will evolve in the interested party's favor, whether or not the rule is efficient. In the event when neither party is interested in precedent, the legal rule—whether efficient or not—will remain in force, and parties will settle out of court because they lack the incentive to change the current rule. Rubin thus concludes that the common law becomes efficient

through an evolutionary process based on the utility maximizing decisions of litigants, rather than on judges' desires to maximize efficiency.

Rubin's analysis was extended by Priest (1977), who articulated the idea that the common law tends to develop efficient rules independently of judicial bias in decision-making. Indeed, he asserts, efficient rules will develop even despite judicial hostility toward efficient outcomes. Priest parts with Rubin, however, on the source of the tendency toward efficiency, rejecting Rubin's conclusion that this tendency occurs only where both parties to a dispute have an interest in similar cases in future and therefore have an incentive to litigate. Instead, he asserts that litigation is driven by the costs of inefficient rules, rather than the desire for precedent.

According to Priest's analysis, inefficient rules impose greater costs on the parties subject to them than do efficient rules, thereby making the stakes in a dispute greater. Where the stakes are greater, litigation is more likely than settlement. Consequently, out of the complete set of legal rules, disputes arising under inefficient rules will tend to be litigated and re-litigated more often than disputes arising under efficient rules. This means that the rules not contested will tend to be efficient ones. Because they are less likely to be reviewed, including by judges hostile to efficient outcomes, these rules tend to remain in force. Further, as inefficient rules are reviewed, the process of review provides the chance that they will be discarded in favor of efficient rules which, in turn, are less likely to be reviewed. Thus, the selection of efficient legal rules will continue through the adjudication process.

Litigation as a Rule Selection Mechanism

An important component of the theories advanced by Rubin (1977) and Priest (1977) is the criteria for the selection of disputes for litigation. In fact, only a small fraction of disputes go to trial, and even fewer are appealed. Priest and Klein (1984) develop a model of the litigation process that explores the relationship between the set of disputes litigated and the set of disputes settled. According to their one-period model of dispute resolution, the proportion of plaintiff victories in any set of cases will be influenced by the shape of the distribution of disputes, the absolute magnitude of the judgment, litigation and settlement costs, and the relative stakes of the parties. Priest and Klein show that the set of disputes selected for litigation, rather than settlement, will therefore constitute neither a random nor a representative sample of the set of all disputes. They then derive a selection hypothesis: where both parties have equal stakes in the litigation, the individual maximizing decisions of the parties will create a strong bias toward a success rate for plaintiffs at trial (or appellants on appeal) of 50 percent, regardless of the substantive standard of law.

When the assumption that both parties have equal stakes in the dispute is relaxed (e.g., where one of the parties is a repeat player and has a stake in similar cases in

future), the rate of success in litigation begins to deviate from the hypothesized baseline, and the model predicts that the repeat player will prevail more frequently. Priest and Klein present a great deal of data, both derived from their own empirical investigations and from the major empirical studies of the legal system since the 1930s. While they caution, because of measurement problems, against the conclusion that these data confirm the selection hypothesis, the data are nonetheless encouraging.

Legal disputes are resolved at various stages of a sequential decision-making process in which parties have limited information and act in their own self-interest. An efficient resolution occurs when legal entitlements are assigned to the parties who value them the most, legal liabilities are allocated to the parties who can bear them at the lowest cost, and transaction costs are minimized. Following these premises, Cooter and Rubinfeld (1989) review economic models of legal dispute resolution, attempting to synthesize a model that provides a point of reference necessary to both an understanding of the courts, and deliberation over proposed changes in the legal rules. In the first stage of a legal dispute, the underlying event, efficiency requires balancing the cost of harm against the cost of harm avoidance. Because Coasian bargaining is typically not possible, the social costs of harm are externalized. Therefore, an initial allocation of entitlements is essential to creating incentives for efficient levels of activity and precaution. During the second stage, the harmed party decides whether or not to assert a legal claim. This requires the balancing of immediate costs, such as hiring an attorney, and the expected benefits from asserting a claim. In the third stage, once a legal claim is asserted, but before the trial commences, trial courts encourage parties to bargain together to reach a settlement. If the parties cannot privately settle their dispute, the court performs this function in the final stage of trial. Using their hybrid economic model of suit, settlement and trial, Cooter and Rubinfeld come to examine the incentives parties face as they proceed through the litigation process, and make predictions based on the decisions available to the parties, with a discussion of some of the concerns that arise from the pursuit of efficiency which pervades normative economic analysis.

Judicial Incentives and the Common Law

To understand judicial behavior, the first step is to analyze the incentives faced by judges in their judicial role. In the federal system, law and economics have had difficulty explaining judicial behavior in economic terms, in part because the federal judiciary is structured so as to shield judges from direct political or economic constraints. Posner (1994) articulates a positive economic theory of the behavior of federal appellate judges, using a model in which judicial utility is primarily a function of income, leisure, and judicial voting. He argues that appellate judges are ordinary, rational people whose behavior is somewhat analogous to that of managers of nonprofit enterprises, voters, and theatrical spectators.

In Posner's view appellate judges are like nonprofit managers in that it is difficult to determine the quality or value of the desired output (neutral "justice") from the full range of their services (rulemaking, private dispute resolution, and interposition between the government and its citizens). A rational public is reluctant to buy such services from a profit-making enterprise because a competitive market is not feasible, and they are reluctant to delegate such services to elected officials whose use of political criteria would not be easily monitored. The judiciary is called on to apply neutral justice with much discretionary power but without monetary or political compensation incentives.

The judiciary's nonprofit structure enables competent people to be attracted to judging at lower wages by not forcing judges to work as hard as comparable lawyers might in private practice. However, because most judges continue their judicial activity beyond the usual retirement age of their private sector counterparts, Posner postulates that judges must derive utility in judging beyond just money and leisure. Posner believes that an appellate judge's utility function additionally contains preferences for a good reputation, popularity, prestige, and avoiding reversal. He explicitly excludes from the judicial utility function a desire to promote the public interest because he says such preference cannot be assumed across the board for all individuals. While it might explain the decisions of a few judges, it is not a good standard overall.

Posner analogizes judicial decision making to political voting. There is pure utility in voting, as evidenced by participation in popular elections in which individuals incur a net cost in order to participate in the political process. This analogy suggests that voting on cases is one of the most important sources of judicial utility due to the deference judges' opinions receive by lawyers and the public. Judges further derive a consumption value in deciding for whom or what to vote. Judges balance this consumption against the opportunity cost of decision-making. Leisure-seeking by judges with weak preferences may result in "going along" voting: insistence that a particular decision is coerced by the law, joining opinions containing much dictum with which they disagree, or using procedural rules to avoid difficult or politically sensitive issues. Posner further suggests that this leisure-seeking explains why judges adhere to *stare decisis*, but not rigidly, given the partially offsetting utility of discretionary powers.

Posner's approach supports the theory that the conditions of judicial employment enable and induce judges to vote for their values (among which Posner believes efficiency to be particularly influential), and his hypothesis generates a number of testable economic predictions about judicial behavior which have engaged an entire generation of legal and economic scholars.

Rules, Standards and the Economics of Legal Rulemaking

Often a judge must choose between a rule that is precise and one that provides better results in most cases but has higher adjudication costs. Posner's foundational work on the economics of legal rulemaking is often associated with the general theory of

adjudication advanced in Ehrlich and Posner (1974). The choice of the degree of precision in the formulation of legal commands is largely based on the desire to minimize social costs. Specific legal rules and general legal standards lie at opposite ends of the precision spectrum. Ehrlich and Posner articulate the criteria for determining the optimal degree of specificity, given cost minimization as a dominant consideration. The authors begin with a static cost-benefit analysis of the optimum level of detail in legal rules.

Ehrlich and Posner (1974) discuss the benefits that precision brings to the legal system, including increased predictability and the consequential reduction in litigation expenditures, increased speed of dispute resolution, and reduced information costs associated with adjudication. The authors suggest that greater precision benefits the legal system. Such benefit results from increasing the marginal productivity of prosecuting the guilty relative to the innocent and reducing the marginal productivity of a guilty defendant's litigation expenditures relative to the innocent. Greater precision allocates scarce judicial resources more efficiently. It makes outcomes more predictable and thus encourages settlement. It decreases the number of legal issues and thus makes dispute resolution more speedy. It reduces information costs in dispute resolution by summarizing what has been learned in prior adjudications, and it facilitates monitoring of judicial agency costs by making incompetent or corrupt outcomes more detectable.

With these benefits, precision also brings costs: the costs of rule formulation (often substantial, given the high transaction costs of statutory decisions) and allocative inefficiency arising from both the over-inclusive and under-inclusive effects of rules. Greater specificity generates inefficiencies from imperfect fits between the coverage of a rule and the conduct it seeks to regulate. Greater precision imposes costs in obtaining and evaluating information and formulating rules (which increases with the number of decision-makers involved). Greater specificity increases the information barriers for laymen, who are more likely to understand general standards than specific rules which employ technical language. Additionally, it increases expenditures on legal counsel because most rules are not nearly as intuitive as standards. Lastly, specificity increases the rate of obsolescence of the rules under changing economic and social conditions.

Customary Law and the Economics of Decentralized Lawmaking

The hypothesis that legal rules evolve toward efficiency by a process similar to natural selection was originally formulated with reference to judge-made common law rules. While wealth-maximizing hypotheses of the common law have served as a baseline for the analysis of other sources of law, different theoretical frameworks are used to explain the economic structure of non-judge-made rules.

As part of this undertaking, scholars of law and economics have examined whether and how far the theory that law is an instrument for maximizing social wealth or efficiency can be extended to other decentralized processes of law formation, such as customary law and social norms.

Adjudicating Social Norms

According to the theory of spontaneous law, customary law has a comparative advantage over other institutional sources. The intellectual basis of this claim is related to the proposition that any social arrangement that is voluntarily entered upon by rationally self-interested parties is beneficial to society as a whole.

The inductive process which underlies spontaneous law builds upon the role of individuals giving direct effect to their revealed preferences, without the interface of third-party decision-makers. To the extent that social practices have emerged under competitive conditions (i.e., so long as there is an implicit cost for indulging in inefficient equilibria) without Pareto relevant externalities, we may be able to draw plausible conclusions regarding the desirability of emerging customs. It is in this latter sense that custom may reclaim full dignity as a source of law. The evolutionary and game-theoretical appraisals of the lawmaking process have indeed shed new light on the normative foundations of spontaneous law, but they require an appropriate analysis of the incentive structure in the originating social environment (Cooter, 1992).

Evolutionary theories of cooperation have indeed explained the ability of rationally self-interested individuals to cooperate for the sake of mutual gain. Evolutionarily stable cooperative strategies serve efficiency goals and may emerge as social norms recognized by the community to be obligatory. Once they have emerged, customary rules generate the expectations of the other members of society and those expectations in turn demand judicial enforcement. In some instances, peer pressure and spontaneous processes of norm internalization will support their enforcement.

The legal system may further this process by recognizing and enforcing welfare-maximizing social norms. In this regard, Cooter (1994) argues that legal recognition and enforcement should consequently be denied in the case of non-cooperative practices, under a test that amounts to a structural analysis of the social incentives that generated the norm. He further argues that in the process of common law adjudication, a distinction must necessarily be made between cooperative norms and non-cooperative practices. Courts are not specialized in the adjudication of most norms. They must therefore resort to a structural approach, first inquiring into the incentives underlying the social structure that generated the norms, rather than attempting to weigh their costs and benefits directly.

The Process of Customary Law Formation

A fundamental insight of the economic analysis of law is the notion that legal sanctions are "prices" set for given categories of legally relevant behaviour. This idea develops around the positive conception of law as a command backed by an enforceable sanction. Law and economics use the well-developed tools of price theory to predict the effect of changes in sanctions on individual behavior. One essential question, however, remains unanswered: How can a legal system set

efficient prices if there is no market process that generates them? In other words, how can legal rules reflect the level of social undesirability of the conduct being sanctioned?

Although the legal system sometimes borrows a price from the actual market (e.g., when the sanction is linked to the compensatory function of the rule of law), there is a wide range of situations in which legislative and judicial bodies set prices in the absence of a proper market mechanism. In a law and economics perspective, customary law can be viewed as a process for generating legal rules that is analogous to a price mechanism in a partial equilibrium framework.

Both the emergence of custom from repeated contractual practice and the role of custom as a non-contractual solution to game inefficiencies have been the object of study in both economic and philosophical literature. Law and economics have revisited this familiar theme, considering the spontaneous emergence of customary law, and, more recently, emphasizing the issue of legal and institutional change in an evolutionary setting (see, e.g., Cooter 1994; Parisi 1995a, b and 1998; Posner 1996a, b; Bernstein 1996). Further, Parisi (2000) has considered the public choice dimension of the process of customary law formation, considering the potential for norm manipulation and the desirability of an increased recognition and incorporation of customary norms by the legal system.

Customary Law in the Age of Legislation

In the "social contract" framework, customary rules can be regarded as an implied and often non-verbalized exercise of direct legislation by the members of a society. Those legal systems that grant direct legal force to customary rules regard custom as a primary, although not exclusive, source of law. In such legal traditions, courts enforce customary rules as if they had been enacted by the proper legislative authority. Custom thus amounts to a spontaneous norm, which is recognized by the legal system and granted enforcement as a proper legal rule.

Modern legal systems generally recognize customary rules that have emerged either within the confines of positive legislation (*consuetudo secundum legem*) or in areas that are not disciplined by positive law (*consuetudo praeter legem*). Where custom is in direct conflict with legislation (i.e., *custom contra legem*) the latter normally prevails. In some instances, however, a custom supersedes prior legislation (i.e., *abrogative custom*), and some arguments have been made in support of emerging practices that conflict with obsolete provisions of public international law (*desuetude*, or abrogative practice) (Kontou 1994).

Judicial recognition of customary practice amounts to a declaratory (rather than a constitutive) function that treats custom as a legal fact. The legal system "finds" the law by recognizing such practices, but does not "create" the law. The most notable illustration is the system of international law, where, in the absence of a central legislative authority, custom stands next to treaties as a primary source of law. (See, e.g., Article 38(1) of the *Statute of the International Court of Justice*; and Restatement 102 of the *Foreign Relations Law of the United States*.)

Whenever they are granted legitimate status in a legal system, customary rules are usually given the same effect as other primary sources of law. Although often subordinated to formal legislation, customary rules derive their force from the concurrence of a uniform practice and a subjective belief that adherence to them is obligatory (*opinio iuris*), without necessarily being formally incorporated into any written body of law. For this reason, they are usually classified as "immaterial" sources of law (H.L.A. Hart 1961: 246–247; Brownlie 1990). This notion implies that the custom remains the actual source of law even after its judicial recognition. In this setting, the judicial decisions that recognize a custom offer only persuasive evidence of its existence and do not themselves become sources of law. In turn, this prevents the principle of *stare decisis* from crystallizing into customary law.

The Anatomy of Customary Law

The theory of customary law defines custom as a practice that emerges outside of legal constraints, and which individuals and organizations spontaneously follow in the course of their interactions, out of a sense of legal obligation. Gradually, individual actors embrace norms that they view as requisite for their collective well-being. An enforceable custom emerges from two formative elements: (a) a quantitative element consisting of a general or emerging practice; and (b) a qualitative element reflected in the belief that the norm generates a desired social outcome.

The Quantitative Element

The quantitative requirements for the formation of customary law concern both the length of time and the universality of the emerging practice. Regarding the time element, there is generally no universally established minimum duration for the emergence of customary rules. Customary rules have evolved from both immemorial practice and a single act. Still, French jurisprudence has traditionally required the passage of forty years for the emergence of an international custom, while German doctrine generally requires thirty years (Tunkin 1961; Mateesco 1947.). Naturally, the longer the time required to form a valid practice, the less likely it is for custom to effectively anticipate the intervention of formal legislation, and to adapt to changing circumstances over time.

Regarding the condition of universality, international legal theory is ambivalent. Charney (1986) suggests that the system of international relations is analogous to a world of individuals in the state of nature, dismissing the idea that unanimous consent by all participants is required before binding customary law is formed. Rather than universality, recent restatements of international law refer to "consistency" and "generality." (D' Amato 1971.) Where it is impossible to identify a general practice because of fluctuations in behavior, the consistency requirement is

not met (*Asylum case* (1950), at 276–277; *Wimbledon case* (1923), Ser. A, no. 1.) Similarly, more recent cases in international law restate the universality requirement in terms of "increasing and widespread acceptance" (*Fisheries Jurisdiction case* (1974), at 23–26; *North Sea Continental Shelf cases* (1969), at 42), allowing special consideration for emerging general norms (or local clusters of spontaneous default rules) that are expected to become evolutionarily stable over time.

With regard to rules at the national or local level, the varying pace with which social norms are transformed suggests that no general time or consistency requirement can be established as an across-the-board condition for the validity of a custom. Some variance in individual observation of the practice should be expected because of the stochastic origin of social norms. A flexible time requirement is particularly necessary in situations of rapid flux, where exogenous changes are likely to affect the incentive structure of the underlying relationship.

The Qualitative Element

The second formative element of a customary rule is generally identified by the phrase *opinio iuris ac necessitates*, which describes a widespread belief in the desirability of the norm and the general conviction that the practice represents an essential norm of social conduct. This element is often defined in terms of necessary and obligatory convention. (Kelsen 1939, D'Amato, 1971; Walden 1977). The traditional formulation of *opinio iuris* is problematic because of its circularity. It is quite difficult to conceptualize that law can be born from a practice which is already believed to be required by law.

The practical significance of this requirement is that it narrows the range of enforceable customs: only those practices recognized as socially desirable or necessary will eventually ripen into enforceable customary law. Once there is a general consensus that members of a group ought to conform to a given rule of conduct, a legal custom can be said to have emerged when some level of spontaneous compliance with the rule is obtained. As a result, observable equilibria that are regarded by society as either undesirable (e.g., a prisoner's dilemma of uncooperative outcome) or unnecessary (e.g., a common practice of greeting neighbours cordially) will lack the subjective and qualitative element of legal obligation and, therefore, will not generate enforceable legal rules.

The concept of *opinio iuris* introduces a distinction between mere behavioral regularities and internalized obligations. This distinction may be related to the parties' awareness of the expected aggregate payoffs from the game, a distinction that is crucially important in the normative setting. Two categories of social rules are generally distinguished: (a) those that reflect mere behavioral patterns that are not essential to the legal order; and (b) those that reflect an internalized belief that the practice is necessary or socially desirable. A mere behavioral regularity, lacking the qualitative element of *opinio iuris*, does not generate a customary rule. In legal jargon, such behavior is a mere usage; in economic terms it simply represents an equilibrium convention. On the other hand, norms considered necessary for social

well-being are treated as proper legal customs and can enter the legal system as primary sources of law.

The Domain of Customary Law

The literature on social norms proceeds by considering the appropriate domain of customary law and studying the situations that are more easily governed by spontaneous law.

The earliest economic models of spontaneous norm formation consider the role of morality and internalized obligations as a means for inducing cooperation in conflict games (Gauthier 1986; Ullmann-Margalit 1977). Internalization of the norm is a source of spontaneous compliance. For example, individuals internalize obligations when they disapprove and sanction other individuals' deviations from the rule, or when they directly lose utility when the norm is violated. In this setting, Cooter (1994) suggests that a legal custom will successfully evolve when the *ex ante* individual incentives are aligned with the collective public interest. Cooter (1994: 224) calls this proposition the "alignment theorem."

The perfect alignment of individual interests rarely occurs in real life situations, however, so proxies for structural harmony (such as role reversibility and reciprocity) have been considered by the more recent literature.

Reciprocity and Incentive Alignment

Individuals choose among alternative rules of behavior by employing the same optimization logic they use for all economic choices. True preferences are unlikely to be revealed when individual interests are not aligned. Traditionally, strategic preference revelation is viewed as a hindrance to the spontaneous emergence of cooperation. Such a problem is likely to be minimized in situations of role reversibility or stochastic symmetry (Parisi 1995a). Role reversibility and stochastic symmetry induce each member to agree to a set of rules that benefits the entire group, thus maximizing her expected share of the wealth.

These conditions in fact occurred during the formative period of the medieval law merchant (*lex mercatoria*), when traveling merchants acted in the dual capacity of buyer and seller. If they articulated a rule of law which was favorable to them as sellers, it could have the opposite effect when they acted as buyers, and vice-versa. This role reversibility changed an otherwise conflicting set of incentives (buyer vs. seller) into one that converged toward symmetrical and mutually desirable rules.

The law merchant therefore illustrates a successful system of spontaneous and decentralized law (Benson 1989, 1990; Greif 1989). Fuller (1969: p. 24) observes that frequent role changes foster the emergence of mutually recognized and accepted duties "in a society of economic traders. By definition the members of such a society enter direct and voluntary relationships of exchange. . . . Finally, economic

traders frequently exchange roles, now selling, now buying. The duties that arise out of their exchanges are therefore reversible, not only in theory but in practice.''

Certainly, the emergence of consensus for a given rule does not exclude the possibility of subsequent opportunistic deviation by some individuals when roles are later reversed. This is a typical enforcement problem, however, and the possibility of strategic defection does not undermine the rule's qualitative features. The general acceptance of (or acquiescence to) a custom depends primarily on its anticipated effect on the group. Those strategies that maximize the expected payoff for each participant if reciprocally undertaken evolve into norms. This conception of spontaneous law is examined by Stearns (1994: 1243–1244), who observes that if the participants were unable to devise rules governing future interactions, and unforeseen circumstances placed them in a forced market relationship requiring post-contractual negotiations, courts and legislatures might have a comparative advantage over the participants in devising market facilitating rules. Unlike market participants, courts and legislatures choose from among alternative solutions as if the underlying events had not yet occurred, without attempting to maximize strategically the advantage caused by unforeseen circumstances (see also Shubik 1987).

Where rules are breached following role reversal, norms play a collateral yet crucial role in sanctioning case-by-case opportunism. A merchant who invokes a particular rule when buying yet refuses to abide by the same rule when selling would be regarded as violating a basic norm of business conduct, and would suffer reputational costs within the business community. Conditions of role reversibility, coupled with norms that generate disincentives to adopt opportunistic double standards, are therefore likely to generate optimal rules via spontaneous processes. The group's ability to impose a sanction obviously depends on an individual's accountability for his past behavior. Benson (1992: 5–7) explores the role of reputation in situations of repeated market interaction, observing that reputation serves as a source of collective knowledge regarding past actions.

Reciprocity Rules in Customary Legal Systems

When unilateral defection promises higher payoffs and there is no contract enforcement mechanism, players are tempted to depart from optimal strategies, often generating outcomes that are Pareto inferior for all (e.g., the well-known prisoner's dilemma game). Prisoner's dilemma-type games are plagued by the dominance of opportunistic behaviour because of the potential accessibility of off-diagonal, non-cooperative outcomes. Schotter (1981), Lewis (1969) and Leibenstein (1982) analyze the role of conventions in correcting prisoner's dilemma situations.

Among the devices capable of correcting prisoner's dilemma-type games, the players can bind their strategic choices to those of their opponents, drastically changing the equilibrium of the game. Ensuring automatic reciprocity by binding a player's strategy to that of his opponent eliminates access to off-diagonal outcomes and renders the reward for unilateral defection unobtainable. Just as no rational player will employ defection strategies in the hope of obtaining higher

payoffs from unilateral cheating; neither will a rational player be induced to select defection strategies as a merely defensive tactic. Automatic reciprocity mechanisms thus guarantee the destabilization of mutual defection strategies and the shift toward optimizing cooperation. (Axelrod 1981.)

Interestingly, where custom is recognized as a primary source of law, mechanisms of automatic reciprocity are generally regarded as meta-rules of the system. One may consider the following two illustrations, drawn respectively from ancient law and modern international law, which reveal substantial structural similarities.

Lawless environments are characterized by structural reciprocity. In such environments, rules of reciprocity emerge as fundamental customary norms. In the absence of an established legal system or commonly recognized rule of law, reciprocity implies that parties can do unto others what has been done to them, subject to the limits of their reciprocal strengths. Ancient customs of retaliation, based on conceptions of symmetry and punitive balance, provide an intriguing illustration of the principle of reciprocity at work. (See, e.g., Exodus 2123; and Code of Hammurabi Paragraphs 108 and 127.)

Similarly, in the so-called law of nations (the system that governs the relationships between states), the voluntary recognition of rules by sovereign states implies that in the absence of a commonly accepted standard of conduct, lawless freedom applies. Positions that are unilaterally taken by one state generate a standard which may be used against the articulating state in future occasions.

Thus, in both ancient law and modern international law, the principle of reciprocity serves as a crucial pillar for the process of law formation. Often, situations of post-contractual behavior capable of modifying states' obligations arise in the law and practice of international relations. The international law formation process provides states with numerous occasions for opportunistic behavior, including hold-out strategies and free riding. Left unconstrained, states' unilateral defection strategies would dominate in equilibrium. To cope with this reality, basic norms of reciprocity are generally recognized as rules of the game.

As a further illustration, one can consider Art. 21(1)b of the 1969 Vienna Convention, which articulates an established custom of reciprocity, creating a mirror-image mechanism in the case of unilateral reservations: "A reservation established with regard to another party ... modifies those provisions to the same extent for that other party in its relations with the reserving state." The effects of this automatic reciprocity mechanism are similar to a tit-for-tat strategy without the need for active retaliation by states: whenever a treaty is modified unilaterally in favor of one state, the result will be as if all the other states had introduced an identical reservation against the reserving state. As shown by Parisi (1995b), by imposing a symmetry constraint on the parties' choices, this rule offers a possible solution to the 'prisoner's dilemma' problems.

It should be noted that, while the principle of reciprocity solves conflict situations characterized by a prisoner's dilemma structure (in both symmetric and asymmetric cases) alone, it is incapable of correcting other strategic problems. For example, when a conflict occurs along the diagonal possibilities of the game (such that the obtainable equilibria are already characterized by symmetric

strategies), a reciprocity constraint will not eliminate the divergence of interests between the players and will not affect the results of the game (e.g., in a "Battle of the Sexes" game, reciprocity is ineffective). The same holds for pure conflict (i.e, zero-sum) situations.

Other Environmental Conditions that Foster Efficient Customs

Evolutionary models further examine the role of long-term relationships in the equilibrium of the game. In long-term human interactions, reciprocity and close-knittedness provide individuals with incentives to choose globally optimizing strategies. Introducing interdependent utility functions into the model, the horizons of individual maximization are extended to include payoffs from future interactions with a direct computation of the well-being of close members within the group. Such a theoretical framework obviously allows for a far more optimistic prediction of spontaneous order. This insight is consistent with the predictions of evolutionary models of social interaction, where low discount rates for future payoffs and the close-knittedness of the group are found to be positively correlated with the emergence of optimal social norms. Models based on interdependent utility and close-knittedness generate results that are qualitatively similar to those discussed for the case of role reversibility.

If the models are further modified to allow the intensity of sentiments of social approbation or disapprobation to vary with the relative frequencies of the two strategies in the population, the degree of spontaneous norm enforcement is likely to increase with a decrease in the proportion of defectors in society. Likewise, norms that are followed by a large majority of the population are more likely to be internalized by marginal individuals in the absence of coercion. Generally, if the measure of spontaneous enforcement and internalization of the norm depends on the proportion of the population that complies with the norm's precepts, the dynamic adjustment will become even more conspicuous. Along with the adjustments taking place in the initial time period, an additional "internalization effect" will occasion a dynamic adjustment of the equilibrium. An initial change in the players' level of norm internalization reproduces the conditions of instability occasioned by the initial emergence of the norm. In this setting, norms become self-reinforcing in that they are likely to occasion an increase in both spontaneous compliance and expected payoffs to the norm-abiding players, with a threshold level of compliance marking the "tilt point" for the survival of the norm.

The various models formulated in the literature suggest that iterated interactions with role reversibility, reciprocity constraints, and structural integration facilitate the emergence and recognition of customary law. The dynamic of the norm formation may unveil the existence of a "tilt point" beyond which emerging beliefs become stable and self-sustaining. In light of reciprocal constraints undertaken by other members of the community, individuals who frequently exchange roles in their social interactions have incentives to constrain their behaviour to conform to socially optimal norms of conduct. Buchanan (1975) insightfully anticipated this

result, suggesting that even stronger logic explains the emergence of cooperation in situations of induced reciprocity. In both cases, the non-idealistic and self-interested behaviour of human actors will generate optimal norms.

Articulation Theories of Customary Law

Notable scholars have considered the conditions under which principles of justice can emerge spontaneously through the voluntary interaction and exchange of individual members of a group. As in a contractarian setting, the reality of customary law formation relies on a voluntary process through which members of a community develop rules that govern their social interaction by voluntarily adhering to emerging behavioural standards. In this setting, Harsanyi (1955) suggests that optimal social norms are those that would emerge through the interaction of individual actors in a social setting with impersonal preferences. The impersonality requirement for individual preferences is satisfied if the decision makers have an equal chance of finding themselves in any one of the initial social positions and they rationally choose a set of rules to maximize their expected welfare. Rawls (1971) employs Harsanyi's model of stochastic ignorance in his theory of justice. However, the Rawlsian "veil of ignorance" introduces an element of risk aversion in the choice between alternative states of the world, thus altering the outcome achievable under Harsanyi's original model, with a bias toward equal distribution (i.e., with results that approximate the Nash criterion of social welfare). Further analysis of the spontaneous formation of norms and principles of morality can be found in Sen (1979), Ullmann-Margalit (1977), and Gauthier (1986).

Legal theorists and practitioners have addressed a similar issue when considering the requirements of *opinio iuris*. In attempting to solve one of the problems associated with the notion of *opinio iuris*, namely the troublesome problem of circularity, legal scholars (notably, D'Amato 1971) have considered the crucial issue of timing of belief and action in the formation of customary rules. The traditional approach emphasizes the awkward notion that individuals must believe that a practice is already law before it can become law. This approach basically requires the existence of a mistake for the emergence of a custom: the belief that an undertaken practice was required by law, when instead, it was not. Obviously, this approach has its flaws. Placing such reliance on systematic mistakes, the theory fails to explain how customary rules can emerge and evolve overtime in cases where individuals have full knowledge of the state of the law.

In this context, legal theorists have proposed to look past the notions of *opinio iuris* and usage concentrating on the qualitative element of "articulation." Articulation theories capture two important features of customary law: (a) customary law is voluntary in nature; and (b) customary law is dynamic. According to these theories, in the process of ascertaining the qualitative element of *opinio iuris*, relevance must be given to the statements and expressions of belief (i.e., articulations) of the various players. Individuals and states articulate desirable norms as

a way to signal that they intend to follow and be bound by such rules. In this way, articulation theories remove the guessing process from the identification of *opinio iuris*.

Consistent with the predicament of the economic models, articulation theories suggest that greater weight should be given to beliefs that have been expressed prior to the emergence of a conflict. Here, it is interesting to point out a strong similarity between the legal and the economic models. Articulations that are made prior to the unveiling of conflicting contingencies can be analogized to rules chosen under a Harsanyian veil of uncertainty.

States and individuals will have an incentive to articulate and endorse norms that maximize their expected welfare. Given some degree of uncertainty as to the future course of events, the emerging rules will be such as to maximize the expected welfare of the community at large. Conversely, rules that are articulated after an outburst of conflict may be strategically biased. Once the future is disclosed to them, parties will tend to articulate rules that maximize their actual welfare, rather than the expected welfare to be derived from an uncertain future. Thus, *ex ante* norms should be given greater weight in the adjudication process.

This predicament seems to be contradicted, however, by the empirical and anecdotal evidence on commercial customary law. Bernstein (1996) examines customary rules that have developed in various modern commercial trades. Her findings seem to indicate that in the adjudication of business disputes, commercial tribunals tend to enforce customary rules that are quite different from the business norms spontaneously followed by the parties in the course of their relationship. Rather, customary rules develop around practices developed during the phase of conflict in a relationship. In this setting, Bernstein distinguishes between relationship norms and end-of-the-game norms. When adjudicating a case, courts are faced with parties who have reached the end point in their relationship. The end-of-the-game norms of the phase of conflict thus tend to be enforced, while the cooperative norms developed in the course of their relationship remain outside the domain of adjudication.

The Limits of Customary Law

Customary rules are generally accepted by the community, with a larger share of rules followed spontaneously by the community and a consequent reduction in law enforcement costs. In the decentralized dynamic of spontaneous law, individual decision-makers directly perceive the costs and benefits of alternative rules, and reveal their preferences by supporting or opposing their formation. The formative process of customary law proceeds through a purely inductive accounting of subjective preferences. Through his own action, each individual contributes to the creation of law. The emerging rule thus embodies the aggregate effects of independent choices by various individuals who participate in its formation. This inductive process allows individuals to reveal their preferences through their own action, without the interface of third-party decision-makers.

The analogy between customary rules and spontaneous market equilibria, however, calls for an assessment of the potential insufficiencies of the spontaneous processes of law formation. The literature in this area is relatively thin and much work still needs to be done to develop a coherent understanding of the limits of spontaneous sources of law.

Path Dependence and the Idiosyncrasies of Customary Law

Norms and conventions vary from place to place. Any theory about the efficiency of spontaneous law should explain the diversity of norms and conventions across time and space. In my view, there are two primary ways to provide such an explanation.

The first is to look for idiosyncratic environmental or institutional factors which might attribute to the diversity of observed rules. If the underlying social, economic, or historical realities are found to be different from one another, different norms or conventions should be expected. Rules, norms and conventions develop in response to exogenous shocks through a natural process of selection and evolution. This "survival of the fittest" explanation would suggest that whatever exists in equilibrium is efficient, given the current state of affairs. This belief, borrowed from Darwinian evolutionism, is pervasive in the literature of law and economics, and, when applied to spontaneous law, risks becoming a tautological profession of faith. Ironically, we should note that the originators of such a claim, socio-biologists, have now widely refuted its validity.

The second way to reconcile the efficiency claim to the observed diversity of spontaneous rules is to consider the role of path dependence in the evolution of norms and conventions. Evolution toward efficiency takes place with some random component. Random historical and natural events (the random element of chaos theory) determine the choice of the initial path. This may be the case particularly where initial choices are made under imperfect information. Evolution then continues towards efficiency along different paths, with results that are influenced and constrained by the initial random conditions.

If we agree that path dependence has something to do with the emergence and evolution of customary law, we should follow this logic to its conclusion, revisiting the very foundations of the efficiency claim. The main question is whether path dependence could ever lead to inefficient results. According to current research (Roe, 1996), path dependence may lead to inefficient equilibria. Once a community has developed its norms and conventions, the costs of changing them may outweigh the benefits. Less efficient rules may persist if the transition to more efficient alternatives is costly. Thus, if one allows for some randomness and path dependence, norms and conventions, although driven by an evolution-towards-efficiency dynamic, may stabilize around points of local, rather than global, maximization. Our history, in this sense, constrains our present choices. We may wish we had developed more efficient customs and institutions, but it would be foolish now to attempt to change them. The claim of efficiency of spontaneous law thus becomes a relative one vis-à-vis the other sources of law. The point then becomes that of

weighing the relative advantages of spontaneous law-making against the attributes of engineered legislation, taking full account of the pervasive public choice and information problems underlying such alternatives.

Rational Abstention and Norm Manipulation

A public choice analysis of customary law should further consider the vulnerability of norms and customs to the pressure of special interest groups. This line of analysis— relatively undeveloped in the current literature—should search for parallels between the legislative process and the dynamics of norm formation. In that setting, the opportunity for collective beliefs and customs to be manipulated by special interest groups should be analyzed. Any claim that customary sources are superior to proper legislation will have to rest on a solid understanding of the relative sensitivity of each source to possible political failures.

The application of a well known theorem of public choice to the study of customary law generates very interesting results. Unlike legislation in a representative democracy, customary law rests on the widespread consensus of all individuals affected by the rule. If 'principal-agent' problems are likely to arise in a political world characterized by rational ignorance and rational abstention of voters, no such problems appear to affect customary sources. Individuals are bound by a customary rule only to the extent that they concurred—actively or through voluntary acquiescence—in the formation of the emerging practice.

Imperfect information, however, may induce voluntary acquiescence—or even active concurrence—to an undesirable practice. Economic models of cascade or bandwagon behavior have shown how inferior paths can be followed by individuals who rely on previous choices undertaken by other subjects, and value such observed choices as signals of revealed preference. Economic models have shown that, when information is incomplete, excessive weight can be attached to the signal generated by others. The choices of others may be followed even when the agent's own perception conflicts with the content of the observed signal. In this way, a biased or mistaken first-mover can generate a cascade of wrong decisions by all his followers, with a result that may prove relatively persistent under a wide array of conditions.

Cascade arguments may also unveil the relative fragility of spontaneous sources of law in light of the possible manipulation of collective beliefs through biased leadership. If information is imperfect, the input of politically biased first-movers may generate undesirable norms. These norms may persist because of the weight attached to the choices of our predecessors. Thus, once generated, wrong beliefs may become stable and widespread in any community of imperfect decision makers.

Collective Action Problems in Customary Legal Regimes

Another potential weakness of customary law is revealed by the application of a collective action framework to the study of the formation and enforcement of customary rules. We can start the analysis by observing that legal rules and law

enforcement are for public good. In the case of customary rules, collective action problems may thus arise at two distinct stages: first, in the formative process of customary rules; and second, in the enforcement of the customs that have emerged.

The process of a custom formation relies on the spontaneous and widespread acceptance of a given rule by the members of a group. Individuals often face a private cost when complying with the precepts of the rule, and they generally derive a benefit because of the compliance of others with existing rules. Thus, the formation of customary law can be affected by a 'public good' problem. When discussing the conditions under which customary rules can effectively develop, I illustrated the analysis with a game-theoretical framework. The 'public good' problem considered here is in many respects similar to the strategic tension that we have examined in the context of customary law formation. If individuals face a private cost and generate a public benefit through norm creation, there will be a suboptimal amount of norms created through spontaneous processes. Any individual would like others to observe a higher level of norm compliance than he or she observes. The resulting level of norm compliance would thus be suboptimal. Collective action problems in the formation of customary rules have traditionally been corrected by norms which sanctioned opportunistic double standards, and by meta-rules imposing reciprocity constraints on the parties.

More serious collective action problems emerge in the enforcement of spontaneous norms. If the enforcement of norms is left to the private initiative of individual members of the group, a large number of cases will be characterized by a suboptimal level of enforcement. Punishing violators of a norm creates a 'public good' because of the special and general deterrent effect of the penalty. Yet imposition of the penalty is left to private initiative; punishers would be willing to enforce norms only to the point to which the private marginal cost of enforcement equals its private marginal benefit. This equilibrium obviously diverges from the social optimum, where enforcement would be carried out until the marginal cost equals the social, rather than private, marginal benefit.

This consideration explains why the customs of ancient societies recognized and sanctioned only a limited category of wrongs. Generally speaking, only those wrongs that had a well-identified victim were likely to be addressed through a system of private law enforcement. For the system of private law enforcement to function properly, it was necessary for the victim or his clan to have a strong interest in carrying out the punishment. This also explains why other categories of wrong with a broader class of victims tend to emerge during more advanced stages of legal development, when law enforcement is delegated to a central authority.

In sum, problems of collective action may be pervasive in the enforcement of customary rules, with a consequential risk that enforcement will be suboptimal. This conclusion suggests that the decentralized process of law formation may be successfully coupled with a centralized mechanism of law enforcement. In this way, the advantages that customary sources have in gathering diffuse information will be available, free from the problems of collective action that typically affect the decentralized processes of law enforcement.

Conclusion: Public Choice and Functional Law and Economics

Scholarship in law and economics has traditionally been labeled as normative or positive. In both such versions, law and economics focus on the role of law as an instrument of behavioral control, and treats the political process and institutions as exogenous.

A new generation of literature—developed at the interface of law, economics and public choice theory—pushes the boundaries of economic analysis of law, studying the origins and formative mechanisms of legal rules. The resulting approach, which can be labeled a "functional" approach (Posner and Parisi, 1998) is quite skeptical of both the normative and the positive alternatives.

The functional approach is wary of the generalized efficiency hypotheses espoused by the positive school. In this respect, the functionalists are aligned with the normative school. Nothing supports a generalized trust in the efficiency of the law in all areas of the law. Even more vocally, the functional school of law and economics is skeptical of a general efficiency hypothesis when applied to sources of the law other than common law.

The functional approach is also critical of the normative extensions and ad hoc corrective policies, which are often advocated by the normative schools. Economic models are a simplified depiction of reality and, according to the functionalists, it is often dangerous to utilize such tools to design corrective or interventionist policies on the basis of such imperfect assumptions. In this respect, the functionalists are aligned with the positive school in their criticism of the normative approach. According to both the positivists and the functionalists, normative economic analysis often risks overlooking the many unintended consequences of legal intervention.

Public choice theory provides strong methodological foundations for the functional school of law and economics. Public choice theory indeed provides the tools for the appraisal of the traditional assumptions of law and economics. The findings of public choice theory, while supporting much of the traditional wisdom, pose several challenges to the theoretical foundations of the neoclassical law and economics approach.

In the above pages we have revisited the important questions concerning the institutional design of lawmaking through the lenses of public choice theory. Alternative sources of law are evaluated considering their respective advantages in the production of legal rules. The functionalist approach to legal analysis sheds new light on the process of law formation suggesting that the comparative evaluation of alternative sources of law requires an appropriate analysis of the incentive structure in the originating environment.

In spite of the sophisticated mathematical techniques of economic analysis, scholars, judges and policymakers in many situations still lack the expertise and methods for evaluating the efficiency of alternative legal rules. Courts and policymakers can thus resort to a functional approach, first inquiring into the incentives underlying the legal or social structure that generated the legal rule, rather than attempting to weigh the costs and benefits of individual rules directly. In this way,

the functionalist approach to law and economics, building upon the solid grounds of public choice theory, can extend the domain of traditional law and economics inquiry to include both the study of the influence of market and non-market institutions (other than politics) on legal regimes, and the study of the comparative advantages of alternative sources of centralized or decentralized lawmaking in supplying efficient rules.

Undoubtedly, the field is still far from a point of maturity. The relationship between competing sources of social and legal order remains for the great part still to be evaluated in light of the important criteria that should govern the institutional design of lawmaking.

References

Axelrod, R.M. (1981). "The emergence of cooperation among egoists." *American Political Science Review*, 75, 306–318.

Benson, B.L. (1989). "The spontaneous evolution of commercial law." *Southern Economic Journal*, 55, 644–661.

Benson, B.L. (1990). *The Enterprise of Law: Justice Without the State*. San Francisco: Pacific Research Institute.

Benson, B.L. (1992). "Customary law as a social contract: International Commercial Law." *Constitutional Political Economy*, 3, 1.

Bernholz, P. (1973). "Logrolling, arrow-paradox and cyclical majorities." *Public Choice*, 15, 87–95.

Bernstein, L. (1996). "Merchant law in a merchant court: Rethinking the code's search for immanent business norms." *University of Pennsylvania Law Review*, 144, 1765.

Broder, D.S. (1997). "Catatonic Politics." *Washington Post*, November 11, A19.

Brownlie, I. (1990). *Principles of Public International Law*. Oxford: Clarendon Press.

Buchanan, J.M. (1975). *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent. Logical Foundations of Constitutional Democracy*. Ann Arbor, MI: University of Michigan Press.

Charney, J.I. (1986). "The persistent objector rule and the development of customary international law." *British Yearbook of International Law*, 56, 1.

Coase, R.H. (1960). "The problem of social cost." *Journal of Law and Economics*, 3, 1–44.

Cooter, R.D. (1992). "Against legal centrism." *California Law Review*, 81, 425.

Cooter, R.D. (1994). "Structural adjudication and the new law merchant: A model of decentralized law." *International Review of Law and Economics*, 14, 215–231.

Cooter, R.D. (2000). *The Strategic Constitution*. Princeton, NJ: Princeton University Press.

Cooter, R.D. and Rubinfeld, D.L. (1989). "Economic analysis of legal disputes and their resolution", *Journal of Economic Literature*, 27, 1067–97.

D'Amato, A. (1971). *The Concept of Custom in International Law*. Ithaca, NY: Cornell University Press.

Dixit, A. and Olson, M., Jr. (1997). "Does voluntary participation undermine the Coase theorem"? (unpublished manuscript on file with the *George Mason Law Review*).

Dixon, A.J. (1985). "Line-item veto controversy." *Congressional Digest*, 64, 259–282.

Easterbrook, F.H. (1983). "Statutes' Domains." *University of Chicago Law Review*, 50, 533.

Ehrlich, I. and Posner, R.A. (1974). "An economic analysis of legal rulemaking." *Journal of Legal Studies*, 3, 257–286.

Elhauge, E.R. (1991). "Does interest group theory justify more intrusive judicial review?" *Law Journal*, 101, 31.

Fuller, L.L. (1969). *The Morality of Law*, Revised Edition. New Haven, CT: Yale University Press.

Gauthier, D. (1986). *Morals by Agreement*. Oxford: Clarendon Press.

Greif, A. (1989). "Reputation and coalitions in medieval trade: Evidence on the maghribi traders." *Journal of Economic History*, 49, 857.

Harsanyi, J.C. (1955). "Cardinal welfare individualistic ethics, and interpersonal comparisons of utility." *Journal of Political Economy*, 63, 315.

Hart, H.L.A. (1961). *The Concept of Law*. Oxford: Clarendon Press.

Kahn, P.L. (1990). "The politics of unregulation: Public choice and limits on government." *Cornell Law Review*, 75(101), 280–312.

Kelsen, H. (1939). "Théorie du Droit international Coutumier." *Revue Internationale de la Théorie du Droit*, 1(New Series), 263.

Kontou, N. (1994). *The Termination and Revision of Treaties in the Light of New Customary International Law*. Oxford: Clarendon Press.

Landes, W.M. and Posner, R.A. (1975). "The independent judiciary in an interest-group perspective." *Journal of Law and Economics*, 18, 875–901.

Leibenstein, H. (1982). "The prisoners' dilemma in the invisible hand: An analysis of intrafirm productivity." *American Economic Review Papers and Proceedings*, 72, 92.

Levine, M.E. and Plott, C.R. (1977). "Agenda influence and its implications." *Virginia Law Review*, 63, 561.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Long, C. and Rose-Ackerman, S(1982). "Winning the contest by agenda manipulation." *Journal of Policy Analysis and Management*, 2, 123.

Mateesco, N.M. (1947). *La Coutume dans les Cycles Juridiques Internationaux*. Paris.

Miller, N.R. (1977). "Logrolling, vote trading and the paradox of voting: A game theoretical overview." *Public Choice*, 30, 51.

Olson, M. (1965). *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Olson, M. (1997). "*The coase theorem is false?*" Paper presented at American Law and Economics Association annual conference held in May 1997, Toronto, Canada.

Parisi, F. (1995a). "Private property and social costs." *European Journal of Law and Economics*, 2, 149–173.

Parisi, F. (1995b). "Toward a theory of spontaneous law." *Constitutional Political Economy*, 6, 211–231.

Parisi, F. (1997a). "The cost of the game: A topology of social interactions." *European Journal of Law and Economics*.

Parisi, F. (1997b). "The economics of customary law: Lessons from evolutionary socio-biology," in B. Bouckaert, and G. De Geest (eds.) *The Production of Law: Essays in Law and Economics*. Antwerpen: Maklu Publisher.

Parisi, F. (2000). 'Spontaneous emergence of law: customary law' Encyclopedia of Law and Economics Edited by R. Posner and F. Parisi, Aldershot: Edward Elgar Publishing, 603–630.

Penny, T.J. (1997). "Pork is safe from this president's line-item vetoes." *Wall Street Journal*, November 13, A22.

Posner, E. (1996a). "Law, economics, and inefficient norms." *University of Pennsylvania Law Review*, 144, 1697.

Posner, E. (1996b). "Norms, formalities, and the statute of frauds: A comment." *University of Pennsylvania Law Review*, 144, 1971.

Posner, R. (1994). "What do judges and justices maximize? (The same thing everybody else does)." *Supreme Court Economic Review*, 3, 1.

Posner, R. and Parisi, F. (1998). "Scuolee Tendenze nella Analisi Economica del diritto." *Biblioteca della Liberta*, 147, 3–19.

Priest, G.L (1977). "The common law process and the selection of efficient rules." *Journal of Legal Studies*, 6, 65–82.

Priest, G.L. and Klein, B. (1984). "The selection of disputes for litigation." *Journal of Legal Studies*, 13, 1–55.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Riggs, R.A. (1973). "Riggs, separation of powers: Congressional riders and the veto power." *University of Michigan Journal of Law Reform*, 6, 735, 743–745.

Roe, M. (1996). "Chaos and evolution in law and economics." *Harvard Law Review*, 109, 641–668.

Rogers, D. (1997). "Clinton warns he's prepared to veto spending bills if changes aren't made." *Wall Street Journal*, October 2.

Rubin, P.H. (1977). "Why is the common law efficient?" *Journal of Legal Studies*, 6, 51–63.

Schotter, A. (1981). *Economic Theory of Social Institutions*. Cambridge, MA: Harvard University Press.

Schwartz, Th. (1977). "Collective choice, separation of issues and vote trading." *American Political Science Review*, 72.

Sen, A.K. (1979). "Rational fools: A critique of the behavioral foundations of economic theory," in F. Hahn, and M. Hollis (eds.) *Philosophy and Economic Theory*. Oxford: Clarendon Press.

Shubik, M. (1987). *Game Theory in the Social Sciences: Concepts and Solutions*, Fourth Edition. Boston: M.I.T. Press.

Stearns, M.L. (1994). "The misguided renaissance of social choice." *Yale Law Journal*, 103, 1219.

Tullock, G. (1981). "Why so much stability." *Public Choice*, 37(2), 189–202.

Tunkin, G.I. (1961). "Remarks on the juridical nature of customary norms in international law." *California Law Review*, 49, 419.

Ullmann-Margalit, E. (1977). *The Emergence of Norms*. Oxford: Clarendon Press.

Walden, R.M. (1977). "The subjective element in the formation of customary international law." *Israel Law Review*, 12, 344.

Chapter 16
The Perspective of Sociology[1]

Viktor Vanberg

Introduction

One of the most significant developments in modern social science is, without doubt, the expansion of economic analysis beyond the customary boundaries of economics into the domains of other disciplinary fields such as law, history, sociology, and political science, a development often referred to as "economic imperialism" (Tullock, 1972; Radnitzky and Bernholz, 1987; Swedberg, 1990, p. 14; Frey, 1999, p. viii). *Public Choice* or, as it has also been called, the *New Political Economy* or *Non-Market-Economics* has played a prominent role in this development, which has significantly changed the relationship between economics and its scientific neighbors. In contrast to the exclusive focus on the mechanics of market forces and the pronounced tendency towards disciplinary isolation that has characterized neoclassical, mainstream economics, the new political economy has systematically extended the "economic perspective" into areas of inquiry that have traditionally been regarded as the domain of other social sciences.

Public choice theory has had its most visible influence in political science, whereas its impact in sociology has been much weaker. Yet, sociology is at the same time the social science that feels most fundamentally challenged by the new, generalized economics. In sociology, more than in any other social science, "economic imperialism" is perceived as a threat to the field's disciplinary identity. Why this is so can be better understood if one takes a look at the history of the relation between economics and sociology, the two neighboring social sciences that "have been estranged from each other far too long" (Swedberg, 1990: p. 3).

---

[1] This chapter is a reprint of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 244–251.

On the History of the Relation between Economics and Sociology

As has often been noted, public choice theory, in particular in its Virginia tradition (Buchanan, 1986, p. 10ff.), is not bringing an entirely new perspective to economics. It is in essence a revival and systematic extension of a research program that was very much inherent in the classical beginnings of economics as a scientific enterprise, in particular so in the political economy of Adam Smith (Buchanan, ibid. p. 10; Buchanan, 1978, p.18, 1987a, p. 585; West, 1990). It is only because this research program had largely been forgotten during the neoclassical period in economics that the new political economy can be said to be "new." Adam Smith held a chair in "moral philosophy," a field that in modern terminology may be most adequately described as *social science* (Lindenberg, 1986: p. 21). Smith and others who, in the history of ideas, are commonly referred to as *Scottish Moral Philosophers*, including, in particular, David Hume and Adam Ferguson, developed in the second half of the eighteenth century a research program that was not at all confined to economic issues in a technical sense but constituted a general paradigm in social theory, integrating economic, legal, political, and social analysis (Rowley, 1998: p. 474ff.; Vanberg, 1975: p. 5ff.). It is a paradigm based on methodological individualism and centered around the idea that social aggregate phenomena should be explained as the largely unintended outcomes "of individual actions directed toward other people and guided by their expected behavior" (Hayek, 1948, p. 6).

The advantages of specialization about which Adam Smith wrote have, of course, shaped the development of academia no less than that of other areas of human activity. Since Smith's time "moral philosophy" has become subdivided into a growing number of specialized social sciences, just as its counterpart, "natural philosophy" has split up into the various specialized natural sciences. Yet, while among the latter the subdivision into specialized fields was largely a matter of a pragmatic division of labor, the situation in the social sciences was different. In particular, the separation between sociology and economics turned into a paradigmatic divide, leading to two fundamentally different theoretical traditions.

In the case of sociology, the manner in which the French sociologist Emile Durkheim (1855–1917) defined the field was critical in its early emergence as an academic discipline. It is largely due to his influence that sociology came to found its claims for disciplinary identity on the assertion that its own theoretical perspective is categorically different from the individualistic, utilitarian perspective of economics, and that the latter, for inherent reasons, is incapable of accounting for important aspects of social reality, in particular for its normative and institutional dimensions. According to Durkheim, the science of sociology has its origins in eighteenth century French social philosophy, the polar counterpart to the individualist paradigm of the Scottish Moralists (Vanberg, 1975, p. 134ff.), and in particular in the social philosophy of Auguste Comte, who was the first to use the term "sociologie" in his *Cours de Philosophie Positive*, published between 1830 and 1842. Though Durkheim rejected Comte's historicist claims, i.e., his concern with

laws of history that supposedly govern the development of human society, he fully endorsed Comte's anti-individualist premises and his claim that society must be looked at as a specific kind of reality apart from its individual constituents and governed by its own laws (Durkheim [1900] 1964a, [1915] 1964b).

In his ambition to secure for sociology a distinct place in an academic environment, in which disciplines like psychology and economics had already established themselves, Durkheim defined sociology in a twofold manner. On one hand he defined it in terms of its subject matter, namely "as the science of institutions, of their genesis and of their functioning" (Durkheim [1895] 1938: lvi), while, on the other, he defined it in methodological terms, namely as a science that has its own, non-individualist theory. It was his methodological commitment that implied that "sociology cannot be based upon a theory that treats the individual as the *starting point* of analysis" (Giddens, 1971, p. 211) and that, therefore, it had to be based on fundamentally different theoretical principles than economics. For economists, he censured, "there is nothing real in society but the individual" (Durkheim, 1978, p. 49). Even worse, their concept of the individual is empirically inadequate, ignoring "all circumstances of time, place, and country in order to conceive of man's abstract type in general" (ibid.), "the sad portrait of an isolated egoist" (ibid.).

Even though later generations of sociologists were not quite as explicit in their programmatic anti-individualism as Durkheim, and even though there were prominent exceptions, such as Max Weber, in its mainstream, sociology has been dominated ever since by the silent premise that the "utilitarian tradition" on which economics is based can, for reasons of principle, not be an adequate foundation for sociological theory (Camic, 1979). In modern sociology, the Durkheim program has most forcefully been restated by Talcott Parsons who proclaimed that "anything like a satisfactory sociological theory could not have been developed at all within the utilitarian framework," and that only "the break with utilitarian premises" allowed for the "emergence of sociological theory" (Parsons, 1968, p. 234).

While sociology, in the manner described above, committed itself to excluding a priori as sociologically inadequate any explanation of social phenomena that starts from assumptions about individual human behavior, post-Smithian economics retained the individualist paradigm of its classical origins, yet it increasingly focused its analytical interest on an ever more narrowly defined aspect of social reality, namely the properties of a highly stylized market, described in terms of a highly stylized model of man. Compared to the much broader outlook of Adam Smith's political economy, the writings of David Ricardo marked the beginnings of a shift in emphasis about which Harold Demsetz (1982, p. 6f.) has said: "Markets became empirically empty conceptualizations of the forums in which exchange costlessly took place. The legal system and the government were relegated to the distant background by the simple device of stating, without clarification, that resources where 'privately owned.'" This shift in the theoretical orientation of economics found its most influential statement in Leon Walras's *Éléments D'Économie Politique Pure* of 1874 which defined the neoclassical research program of modern economics. It was Walras's ambition to develop a "science of pure

economics" in the same spirit as "a physico-mathematical science like mechanics or hydrodynamics" (Walras [1874] 1954, p. 71). His pure economics was to be concerned with "how prices result from a hypothetical régime of absolutely free competition" (ibid., p. 256), supposing "that the market is perfectly competitive, just as in pure mechanics we suppose, to start with, that machines are perfectly frictionless" (ibid., p. 84). Phenomena which he "classified under the heading of institutions" (ibid., p. 63) Walras explicitly excluded from the domain of pure economics, without denying, though, that they can be a proper subject of economics more broadly understood. Even though the study of institutional phenomena fell, in his view, outside of "economics as an exact science" (ibid., p. 47), he considered it the appropriate subject of what he called "social economics" (ibid., p. 79). However, the part that his "social economics" would have had to play in a more broadly conceived economics was never developed, not by Walras himself nor in what has come to be known as the Walrasian tradition. Neoclassical mainstream economics remained occupied with advancing and formalizing in ever more refined ways Walrasés program for "a scientific theory of the determination of prices" (ibid., p. 40), and left unattended the institutional issues that Walras had assigned to "social economics."

Convergences between Economics and Sociology

A sociology that would have concentrated on the study of institutional phenomena, without pre-committing to a non-individualist theoretical perspective, and an economics that would have primarily concerned itself with the study of market processes, without excluding categorically the institutional dimension from its explanatory enterprise, such a sociology and such an economics could have productively co-evolved as two disciplines that, in a pragmatic division of labor, focus on different kinds of issues, but share in the same theoretical foundation and they could have easily supplemented each others explanatory contributions wherever the task at hand required it. Yet, the diverging developments described above created a paradigmatic divide between a non-individualist sociology claiming institutional phenomena as its own domain and an individualist economics studying the mechanics of markets, explicitly disregarding the institutional dimension. As a consequence, these developments left a "'no-man's land' between economics and sociology" (Swedberg, 1990, p. 316), namely the systematic study of institutional phenomena and, more generally, non-market phenomena from an individualist perspective. The significance of "Public Choice" and related theoretical approaches that emerged within economics during the second half of the twentieth century lies in the very fact that they have embarked on a systematic exploration of this "no-man's land."

James M. Buchanan has explicitly characterized public choice theory as "an attempt to close up the analysis of social interaction systems" (Buchanan, 1972, p. 11). It is an effort to pursue the development "of an internally consistent social

science" (ibid., p. 23) by systematically extending the individualist economic approach beyond its traditional domain as "a highly developed theory of market behavior" (ibid., p. 11) to the "'public choices' that define the constraints within which market behavior is allowed to take place" (ibid.) and, more generally, to non-market behavior (ibid., p. 23). Along with other approaches to a "new institutional economics," public choice is about the extension of the "*homo oeconomicus* postulate from market to collective institutional settings" (Buchanan, 1983, p. 12). It is an effort to analyze institutional and non-market phenomena within the same general paradigm as market phenomena, i.e., "with individual decision-makers as the basic units" (ibid., p. 9). Characterizing the various related approaches that seek to go beyond the boundaries of traditional mainstream economics, Ronald Coase (1994, p. 36) notes that "economists are extending the range of their studies to include all of the social sciences." Hans Albert (1979, p. 8) speaks of these approaches as a revival of the "general sociological research program" that was present at the classical origins of economics, a research program that constitutes a principal alternative to theoretical perspectives prevalent in sociology, and that can well be developed into "a general paradigm for social science" (ibid., p. 23). And Bruno S. Frey describes the generalized economic perspective that is "known under such terms as 'Non-Market Economics,' 'New Political Economy,' or 'New Institutionalism'" (Frey, 1999: viii) as "a *new paradigm* for the *social sciences*" (ibid.,p. vii), as an outlook that "applies the same theoretical approach to many different areas," thus advancing the "unity of the social sciences" (ibid., p. 15).

Just as within economics the discontent with the narrow focus of the Walrasian research program led to efforts to revive the general social science perspective of classical political economy, in sociology the discontent with the a priori methodological commitment of the Durkheim program led to attempts to advance an individualist sociological theory, expressly compatible with the behavioral foundations of economics and, in fact, with direct connections to public choice theory (Vanberg, 1983). Even though these approaches have hardly begun to grow out of their somewhat marginal existence within the sociological profession, they deserve attention in the present context. It was the Harvard sociologist George C. Homans who initiated the modern emergence of an individualist sociology with his 1958 article "Social Behavior as Exchange," in which he argued that all human interaction can be looked at as exchange, in much the same way that economists look at market behavior. In his 1964 presidential address to the American Sociological Association, Homans challenged his colleagues with his plea to "bring man back" into sociology, and to acknowledge the fact that beneath their programmatic anti-individualism sociologists have actually no other operable theory to work with than conjectures about human behavior. Calling for an end to what he chastised as "intellectual hypocrisy," Homans (1964, p. 818) noted: "It would unite us with the other social sciences, whose actual theories are much like our actual ones, and so strengthen us all." And he invited his fellow sociologists to see as their task to explain how relatively enduring social structures are "created and maintained by the actions of individuals, actions of course taken under the influence and constraint of the actions of other individuals" (Homans, 1975, p. 64).

The fact that Homans' own work remained very much focused on the analysis of "elementary social behavior" (Homans, 1974) made it easy for mainstream sociologists to discount his challenge as not really demonstrating that social phenomena at the structural and institutional level can be explained satisfactorily on the basis of assumptions about individual human behavior, without recourse to a "genuine sociological theory." A more direct challenge to the traditional doctrine arose in the work of James S. Coleman (1990), who said about himself that he had been converted by Homans' arguments from a "Durkheimian" to an advocate of methodological individualism (Coleman in Swedberg, 1990, p. 49; Coleman, 1986, p. 2). Coleman applied the individualist paradigm explicitly to the institutional and organizational level, notably with his theory of corporate action, a theory that seeks to answer, on the basis of assumptions about individual behavior, the question of how men act collectively through corporate units such as households, firms, organizations, political parties, nation-states, etc. (Vanberg, 1982, p. 8ff.). Distancing himself from the standard self-image of his fellow sociologists, he called for a sociological approach that "does not afford itself the luxury of beginning with already formed units of social organization. Instead, it must begin with persons, and move up from there, or if, in an application, it begins at a level above persons, it must be ultimately analyzable into relations among persons" (Coleman, 1975, p. 85f.).

Coleman has been one of the early members of the Public Choice Society, and he has explicitly argued that sociology can build on the same theoretical foundation as microeconomics, namely the "purposive actor model" (Coleman, 1975, p. 88). He has added, though, that borrowing their basic behavioral model from economics does not dispense sociologists from their task to build sociological theory (ibid., p. 93). In the introduction to a collection of his essays, entitled *Individual Interest and Collective Action* he describes the articles included as "attempts to investigate some of the most important problems of sociology. . . beginning with a paradigm of rational action borrowed from economics (and slightly elaborated)" (Coleman, 1986, p. 10f.).

Coleman's work has become the focal point of an international group of like-minded sociologists that includes, among many others, such scholars as Raymond Boudon in France, Michael Hechter and Douglas D. Heckathorn in the United States, Siegwart Lindenberg and Werner Raub in the Netherlands, Karl-Dieter Opp and Thomas Voss in Germany, or Toshio Yamagishi in Japan. This group of individualist sociologists is often subsumed under the umbrella-name of "rational choice sociology," even though not all of them are firmly wedded to the rational choice model, and one of their main outlets for their publications is the journal *Rationality and Society* that Coleman founded in 1989. In his "Editorial Introduction" to the journal's inaugural issue, Coleman (1989, p. 6) notes that *Rationality and Society* "explicitly espouses methodological individualism" and that its focus is "on the paradigm of rational action," a paradigm on which he comments: It is a "paradigm in social science that offers the promise of bringing greater theoretical unity among disciplines than has existed until now. . . . It is the paradigm on which economic theory rests. It is the basis for the expanding domain of public choice

within political science. It is the paradigm of the burgeoning field of law and economics. . . . Social exchange theory is one of the manifestations of this paradigm in sociology" (ibid., p. 5).

Sociology as a "Multi-Paradigm Science"

From the perspective of rational choice sociology there is obviously no reason to perceive public choice theory along with the other branches of the new generalized economics as an "imperialistic" threat. Instead, both sides can be seen as perfectly compatible research programs that build on the same theoretical foundations, even though, due to the different substantive interests of their respective home-disciplines, they typically apply their shared basic paradigm to different kinds of explanatory issues. Yet, as noted before, the group of sociologists who explicitly subscribe to methodological individualism represents only a minority fraction within the sociological profession at large. As it has been since its origins as a separate academic discipline, sociology continues to be dominated by a methodological commitment to a non-individualist, anti-utilitarian outlook, in spite of the fact that this commitment, quite apparently, even after more than a century since its inception has not come to fruition in the sense of producing a theoretical paradigm that would unite the field. Instead, as soon as one looks beyond the surface of a generally—explicitly or tacitly—shared anti-individualist outlook, modern sociology presents itself as a theoretically fragmented discipline, lacking even a consensus on what criteria one could possibly employ to judge what may count as a "theory," let alone as a "good theory." In his introduction to a textbook on contemporary sociological theories one author describes the state of the field in these terms: "Beneath the surface of professional association membership are numerous disagreements, tensions, and disputes that threaten to break up even the formal unity of sociologists. . . . Such disputes reveal that sociologists are, as yet, unsure of the foundation of their discipline and to some, the foundations have yet to be laid" (Wells, 1978, p. 1).

Euphemistically, sociology may be described as a "multi-paradigm science" (Ritzer, 1975), reflecting the variety of distinct theoretical perspectives that one typically finds discussed in surveys on "sociological theory," perspectives such as functionalism, structuralism, conflict theory, critical theory, Marxism, symbolic interactionism, ethno-methodology, post-modernism, and others. In a more critical assessment of the current state of sociological theory one might suspect that its a priori commitment to a non-individualist outlook may have led sociology into a blind alley. As George C. Homans has put it, misled by its self-imposed programmatic restrictions sociology has looked "for its fundamental principles in the wrong places and hence without success" (Homans, 1967, p. 73).

A brief comment should be added here on Max Weber (1864–1920), who was mentioned earlier as a prominent exception to the dominant programmatic commitment in sociology. Trained as an economist in the tradition of the German

Historical School, Weber was sympathetic to the analytical economics of the Austrian tradition, and one of his aims in developing his research program of social economics (*Sozialökonomik*) was to reconcile the conflicting views in the "Methodenstreit" that had erupted with the controversy between Carl Menger and Gustav Schmoller. As a sociologist, Weber clearly considered himself a "methodological individualist", even though he did not use the term that was not yet common in his time, defining sociology as a science that seeks to explain social and economic phenomena through "understanding" the human actions that, collectively and interactively, bring them about. He once explicitly stated (in a letter, dated March 9, 1920, to his colleague Robert Liefmann) that when he moved from economics into sociology he made it his mission to fight the widespread misuse of collectivist notions and to insist that the individualist approach is the only adequate method in sociology (Vanberg, 1975, p. 103f.).

Max Weber's theoretical outlook could have well become the focal point of an individualist tradition in sociology, yet this is not what happened. Even though, since decades, he has been and continues to be one of the most often cited "authorities" in matters of sociological theory, it is not his methodological individualism but rather other aspects of his work that gained him popularity, such as his concept of "ideal types" and his emphasis on the method of "Verstehen," or his historical studies on the rise of capitalism and other subjects. In fact, even those sociologists who explicitly regard themselves as working in the Weberian tradition are typically no less convinced than their fellow sociologists that the individualist-utilitarian approach of economics is entirely inadequate as a foundation of sociological theory. Very few of them are likely to form an alliance with rational choice sociologists.

Economic Imperialism?

In light of the history, briefly traced above, of the relation between economics and sociology it is both unfortunate and misleading that the label *economic imperialism* has come to be used to describe the efforts within public choice and related approaches to reconstitute economics as a general *social science*. The use of this label is unfortunate because with "imperialism" one tends to associate the notions of invasion and conquest, notions that make it appear as if what is at stake is a hostile takeover of other social sciences by economics. Such framing of the issue is, for obvious reasons, unlikely to invite an open discourse on the relative merits of alternative theoretical approaches. Instead, it is bound to provoke little more than defensive reactions on part of the prospective victims of economics' expansionist ambitions. The label *economic imperialism* is, however, also misleading, and this in two ways. First, it distracts from the fact that the real issue is not about the relation between different disciplines within the social sciences but about the explanatory potential of alternative theoretical paradigms *for the social sciences*. And second, it is misleading because it distracts attention from the fact that the theoretical

foundations of "the economic approach" cannot remain unaffected in the process, but that its generalized application makes apparent the need to reform the "economic model of man" in certain ways. Both aspects are briefly discussed below.

The true significance of what is misleadingly labeled as "economic imperialism" lies not in the fact that economics is expanding its domain at the expense of other social sciences, or that economists were about to claim a general competence in the various areas traditionally covered by other disciplines. Economists should be the last to ignore that division of labor is of no lesser importance in academia than in other realms of life. What is at stake is the *theoretical unity of the social sciences*, not the ambition to turn the various social sciences into branches of economics. As was described above, it was the paradigmatic split in the social sciences, in particular the split between sociology and economics, that made a thoroughgoing methodological individualism the trademark of the "economic approach." But, apart from the fact that economics has been the only social science that has been predominantly and consistently committed to methodological individualism, there is nothing specific "economic" about an individualist approach. Therefore, to show how an individualist methodology can be successfully used to solve explanatory problems traditionally studied by other social sciences is not about exporting a uniquely "economic approach." It means only to show that a theoretical approach that has been used in economics, and largely remained confined to economics, has a much broader explanatory potential than previously recognized. In this sense the reorientation in the social sciences that public choice and related approaches in modern economics have initiated is, in essence, about the consistent application of an individualist methodology throughout the social sciences. It does not put in question that there can be a meaningful division of labor between various specialized disciplines. What it does put in question is the theoretical or paradigmatic divisions that have fragmented the social sciences in the past and that have robbed them of the opportunity to communicate effectively among each other.

As noted, there is a second sense in which the label "economic imperialism" is misleading. It suggests that the need to change and to adapt their theoretical orientation is exclusively on the side of the "invaded" disciplines, while the invading "economic paradigm" remains essentially unaffected in the process. To be sure, that the new generalized economics departs from the neoclassical tradition in that it explicitly seeks to account for institutional aspects is, as has been noted above, well recognized. What is not equally well recognized is the fact that the explanatory potential of neoclassical theory is not only limited by its focus on highly stylized markets, but also by the fact that it employs a highly stylized model of man, of *homo economicus*, the perfectly rational, fully informed maximizer of his own welfare. Reconstituting economics as a general social science does, however, not only require one to rectify the institutional deficiency of the neoclassical tradition, it also requires one to modify its problematic behavioral assumptions. Someone who in his thorough critique of the Walrasian tradition has always emphasized both aspects is Hans Albert. While maintaining that the research program of classical political economy represents a general theoretical approach in social science, he blames the "model-Platonism" of neoclassical economics for

its disregard of institutional aspects as well as for its behavioral deficiency, i.e., for the fact that it turned the broadly utilitarian psychology of the classics into a purely formal "*decision logic* or logic of choice" (Albert, 1979, p. 9), instead of developing it into an empirically content-full theory of human behavior. And while he acknowledges that the new institutionalism in economics "has undoubtedly rehabilitated one of the important ideas of the economic research program" (ibid., p. 20), he leaves no doubt that in his view serious problems remain with the behavioral foundations of the economic approach.

The Behavioral Foundations of Economics as a Social Science

One may well agree with Ronald Coase's (1994, p. 45) supposition that "the success of economists in moving into other social sciences is a sign that they possess certain advantages in handling the problems of those disciplines." The crucial question, however, is which of the specific attributes that may define the economic approach confer to it the advantages that account for its success. There are good reasons to presume that methodological individualism and the general notion of self-interested human behavior are the crucial contributing factors here, not, however, the particular model of "rational choice" in its standard interpretation. Doubts about the explanatory power of the latter are, in fact, voiced by Coase (ibid., p. 43) when he notes: "To say that people maximize utility... leaves us without any insight into why people do what they do." James Coleman who, as noted above, expressly calls for a sociology based on rational choice theory, also voices some caution when he argues: "For the moment, it is the only well developed conception of rational action that we have; and though it may well be replaced when cognitive psychology is more fully developed, there is nothing to replace it now" (Coleman 1975. p. 81). Such a waiting attitude is not what Dennis C. Mueller (1986) has recommended to his colleagues. In his presidential address to the 1986 Public Choice Society Meeting, he made a case for replacing what he calls the "rational egoism postulate" of economics by a behavioral theory that, while maintaining the assumption of self-interested, payoff-oriented behavior, puts less emphasis on *rational choice* than on *adaptive learning*. Specifically, Mueller advocates "starting with behaviorist psychology" (ibid., p. 15), which, incidentally, is the same choice that George C. Homans had made for his individualist sociology, noting: "We believe that the propositions of behavioral psychology are the general explanatory propositions of all the social sciences. Accordingly, they are the general propositions of economics" (Homans, 1974, p. 74). In a similar spirit Mueller notes as an advantage of "starting with behaviorist psychology" that "it allows us to begin with a unified view of human behavior" and that it "is less of a methodological leap for a social scientist who works with rational egoist models than going to some competing sociological-psychological theories" (Mueller, 1986, p. 15).

Whether a more behaviorist approach, as suggested by Mueller, or a model closer to cognitive psychology, as suspected by Coleman, is a more likely candidate for replacing the traditional rational choice model, remains to be seen. It should be apparent, though, that economists who—be it in public choice or in other branches of the new generalized economics—seek to re-establish, in the spirit of classical political economy, the economic approach as a "general paradigm for social science" (Frey) cannot evade the task of rectifying not only its *institutional* but also its *behavioral* deficiency. In fact, it is more and more recognized that, ultimately, institutional phenomena cannot be consistently explained without substituting the standard rational choice model with a behavioral theory that accounts for the element of *rule-following* in human conduct. The fact that rational choice theory with its focus on the incentive-contingencies of *single choices* tends to emphasize the role of situational calculation in human behavior and to ignore the extent to which such behavior is a matter of habits, routines and rule-guided conduct, has been a principal obstacle in the relation between economics and sociology. Durkheim was surely right when he argued that "mores and the prescriptions of law and morality would be impossible if man were not capable of acquiring habits" (Durkheim, 1978, p. 51), and he had reason to conclude that an economics that models man as a case-by-case maximizer is unable to systematically explain the role of habits.

While Durkheim erred when he diagnosed that a non-individualist alternative to the approach of economics is needed to account for the role of norms and institutions in social life, his challenge to the economic model of man remains to be answered. Habits, routines, and other forms of individual rule-guided behavior are, indeed, the building blocks of what we call institutions (Nelson, 2002, p. 21ff.), and a model of man on which explanations of institutional phenomena are to be founded must be able to provide a systematic explanation for the obvious rule-following element in human behavior. The reluctance of economists to part with their traditional model of man may be somewhat softened by recent theoretical developments in cognitive psychology and evolutionary theory that converge towards a model of rule-based or program-based behavior. This model maintains much of what has made rational choice theory so attractive to economists but provides at the same time a systematic bridge between the notion of pay-off governed individual behavior on the one hand and institutional phenomena on the other (Vanberg, 2002).

Conclusion

As has been described above, sociology was established as an academic discipline in explicit opposition to the individualist-utilitarian approach of economics, and the conviction that this approach cannot provide an adequate foundation for sociological theory still very much dominates—if only as a tacit premise—the profession. From the viewpoint of the majority of the profession, public choice is, therefore,

perceived as "economic imperialism" that threatens the very identity of sociology. The apparent failure of the Durkheim program to produce a non-individualist, genuine "sociological" theory that would be able, through its explanatory success, to unite the field, has led, however, to the emergence of individualist approaches in sociology, in particular *rational choice sociology*, with close affinities to public choice and other branches of the new generalized economics. Viewed from the perspective of rational choice sociology, public choice is not perceived as an imperialist threat at all but as a promising development towards a theoretically unified social science. Rational choice sociologists would readily agree with James Buchanan's (1987b, p. 234) diagnosis that public choice and related approaches in modern economics "point toward a fundamental revision of existing orthodoxy, and an emerging consensus on what may be called a general theory of social structures, which will surely include political organization as only one among an array of forms. These developments should help to break down the barriers among the disciplinary specializations in the social sciences, barriers which have been, at best, arbitrarily erected and maintained."

References

Albert, H. (1979). "The economic tradition: Economics as a research program for theoretical social science," in K. Brunner (ed.) *Economics and Social Institutions*. Boston: Martinus Nijhoff, pp. 1–27.

Buchanan, J. M. (1972). "Toward analysis of closed behavioral systems," in J.M. Buchanan and R.D. Tollison (eds.) *Theory of Public Choice — Political Applications of Economics*. Ann Arbor: The University of Michigan Press, pp. 11–23.

Buchanan, J. M. (1978). "From private preferences to public philosophy: The development of public choice," in *The Economics of Politics*. London: The Institute of Economic Affairs, pp. 1–20.

Buchanan, J. M. (1983). "The public choice perspective." *Economia Delle Scelte Pubbliche* 1, 7–15, reprinted in Buchanan 1986, pp. 19–27.

Buchanan, J. M. (1986). *Liberty, Market and State: Political Economy in the 1980s*. New York: New York University Press.

Buchanan, J. M. (1987a). "Constitutional economics." *The New Palgrave: A Dictionary of Economics*, Volume 1. London: Macmillan, pp. 585–588.

Buchanan, J. M. (1987b). "An individualistic theory of political process," in *Economics: Between Predictive Science and Moral Philosophy*. College Station: Texas A&M University Press, 223–235 (first published in D. Easton, ed., *Varieties of Political Theory*, Englewood Cliffs: Prentice-Hall, 1966).

Camic, C. (1979). "The utilitarians revisited." *The American Journal of Sociology*, 85, 516–550.

Coase, R. H. (1994). "Economics and contiguous disciplines," in R.H. Coase (ed.) *On Economics and Economists*. Chicago and London: The University of Chicago Press, 34–46.

Coleman, J. S. (1975). "Social structure and a theory of action," in P.M. Blau (ed.) *Approaches to the Study of Social Structure*. New York: The Free Press, pp. 76–93.

Coleman, J. S. (1986). *Individual Interests and Collective Action: Selected Essays*. Cambridge: Cambridge University Press.

Coleman, J. S. (1989). "Editor's introduction: Rationality and society." *Rationality and Society*, 1, 5–9.

Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA.: The Belknap Press of Harvard University Press.

Demsetz, H. (1982). *Economic, Legal, and Political Dimensions of Competition*. Amsterdam: North-Holland.

Durkheim, E. (1938). *The Rules of Sociological Method*. Glencoe, IL: The Free Press (originally published in 1895).

Durkheim, E. (1964a). "Sociology and its scientific field," in K.H. Wolff (ed.) *Essays on Sociology and Philosophy by Emile Durkheim*. New York: Harper & Row, pp. 354–375 (originally published in 1900).

Durkheim, E. (1964b). "Sociology," in K.H. Wolff (ed.) *Essays on Sociology and Philosophy by Emile Durkheim*. New York: Harper & Row, pp. 376–385 (originally published in 1915).

Durkheim, E. (1978). *Emile Durkheim on Institutional Analysis* (edited, translated and with an introduction by Mark Traugott). Chicago and London: University of Chicago Press.

Frey, B. S. (1999). *Economics as a Science of Human Behavior: Towards a New Social Science Paradigm*, extended Second Edition. Boston: Kluwer Academic Publishers.

Giddens, A. (1971). "The 'individual' in the writings of Emile Durkheim." *Archives européennes de sociologie*, 12, 210–228.

Hayek, F. A. (1948). *Individualism and Economic Order*. Chicago: The University of Chicago Press.

Homans, G. C. (1958). "Social behavior as exchange." *The American Journal of Sociology*, 62, 597–606.

Homans, G. C. (1964). "Bringing men back in." *American Sociological Review*, 29, 809–818.

Homans, G. C. (1967). *The Nature of Social Science*. New York: Harcourt.

Homans, G. C. (1974). *Social Behavior: Its Elementary Forms*, Revised Edition. New York: Harcourt.

Homans, G. C. (1975). "What do we mean by social 'structure'?" in P.M. Blau (ed.) *Approaches to the Study of Social Structure*. New York: The Free Press, pp. 53–65.

Homans, G. C. (1982). "The present state of sociological theory." *The Sociological Quarterly*, 23, 285–299.

Lindenberg, S. (1986). "How sociological theory lost its central issue and what can be done about it," in S. Lindenberg, J.S. Coleman, S. Nowak (eds.) *Approaches to Social Theory*, New York: Russel Sage Foundation, pp. 19–24.

Mueller, D. C. (1986). "Rational egoism vs. adaptive egoism as fundamental postulate for a descriptive theory of human behavior." *Public Choice*, 51, 3–23.

Nelson, R. R. (2002). "Bringing institutions into evolutionary growth theory." *Journal of Evolutionary Economics*, 12, 17–28.

Parsons, T. (1964). *Essays in Sociological Theory*, Revised Edition. New York: The Free Press.

Parsons, T. (1968). "Utilitarianism: sociological thought." *International Encyclopedia of the Social Sciences*, Volume 16. New York: Macmillan, pp. 229–236.

Radnitzky, G. and Peter, B. (eds.) (1987). *Economic Imperialism. The Economic Method Applied Outside the Field of Economics*. New York: Paragon.

Ritzer, G. (1975). *Sociology: A Multiple Paradigm Science*. Boston: Allyn and Bacon.

Rowley, C. K. (1998). "Law-and-economics from the perspective of economics." *The New Palgrave Dictionary of Economics and the Law*. London: Macmillan, pp. 474–486.

Swedberg, R. (1990). *Economics and Sociology — Redefining their Boundaries: Conversations with Economists and Sociologists*. Princeton, NJ: Princeton University Press.

Tullock, G. (1972). "Economic imperialism," in J.M. Buchanan and R.D. Tollison (eds.) *Theory of Public Choice: Political Applications of Economic*, Ann Arbor: The University of Michigan Press, pp. 317–329.

Vanberg, V. (1975). *Die zwei Soziologien: Individualismus und Kollektivismus in der Sozialtheorie*. Tübingen: J.C.B. Mohr (Paul Siebeck).

Vanberg, V. (1982). *Markt und Organisation — Individualistische Sozialtheorie und das Problem korporativen Handelns*. Tübingen: J.C.B. Mohr (Paul Siebeck).

Vanberg, V. (1983). "The rebirth of utilitarian sociology." *The Social Science Journal*, 20, 71–78.

Vanberg, V. (2002). "Rational choice vs. program-based behavior: Alternative theoretical paradigms and their relevance for the study of institutions." *Rationality and Society*, 14, 7–53.

Walras, L. (1954). *Elements of Pure Economics or the Theory of Social Wealth* (translated by W. Jaffé). Homewood, IL: Richard D. Irwin (originally published in 1874).

Wells, A. (ed.) (1978). *Contemporary Sociological Theories*. Santa Monica: Goodyear.

West, E. G. (1990). "Adam Smith and public choice," in E.G. West (ed.) *Adam Smith and Modern Economics: From Market Behaviour to Public Choice*. Aldershot: Edward Elgar Publishing, pp. 105–131.

Part IV
Fundamental Principles

Chapter 17
Constitutional Political Economy[1]

James M. Buchanan

Constitutional and Non-Constitutional Economics

There is a categorical distinction to be made between constitutional economics and non-constitutional, or ordinary, economics—a distinction in the ultimate behavioral object of analytical attention. In one sense, all of economics is about choice, and about the varying and complex institutional arrangements within which individuals make choices among alternatives. In ordinary or orthodox economics, no matter how simple or how complex, analysis is concentrated on choices made *within* constraints that are, themselves, imposed exogenously on the person or persons charged with making the choice. The constraints that restrict the set of feasible choice options may be imposed by nature, by history, by a sequence of past choices, by other persons, by laws and institutional arrangements, or even by custom and convention. In the elementary textbook formulation of demand theory, for example, the individual consumer-purchaser confronts a range of goods available at a set of prices, but is restricted by the size of the budget. This budget is not within the choice set of the consumer-purchaser during the period of choice under scrutiny. Indeed, it would seem unnatural or bizarre, within the mind-set fostered by ordinary economics, to consider or limit the set of available choice options. Within this mind-set, the utility of the chooser is always maximized by allowing for choices over the whole range allowed by the exogenously determined constraints.

It is precisely at this critical point that constitutional economics, in its most inclusive definition, departs from the conventional framework of analysis. Constitutional economics directs analytical attention to the *choice among constraints*. Once stated in this fashion, economists will recognize that there is relatively little in their established canon that will assist in analyzing choices of this sort. To orthodox economists, only the elementary reality of scarcity makes choice necessary; without

---

[1] This chapter is a reprint of an essay that appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 60–67. An earlier version of this chapter was first published under the title: "The Domain of Constitutional Economics' in *Constitutional Political Economy* 1, (1990), 1–18.

scarcity there would be no need to choose. And it would appear to be both methodologically and descriptively absurd to introduce the artificial creation of scarcity as an object for behavioral analysis. Such bedrock conservatism presumably explains much of ordinary economists' inattention and disinterest in constitutional questions, at all levels.

If we move beyond the models of orthodox economics, however, even while remaining at the level of individual behavior, we observe that individuals do, in fact, choose their own constraints, at least to a degree and within some limits. Within recent decades, a few innovative thinkers from economics and other social sciences have commenced to study the choice processes that are involved here (Elster, 1979; Schelling, 1978; Thaler and Shefrin, 1981). The *economics of self-control* has reached the status of a respectable, if minor, research program, which may be destined to become more important in this era of emphasis on diet, exercise, health, and the environment. We must surely be sufficiently catholic to allow analysis in this *individual constitutional economics* to qualify for inclusion in the domain.

As they carry on within their own guaranteed private spaces, however, individuals would presumably subject themselves to a relatively small set of prior constraints. Individuals basically *trust themselves* to choose rationally when confronted with the externally imposed constraints that are dictated in their historically emergent conditions. If the choice among constraints, in all its complexity, is limited to the economics of self-control, or stated conversely, to the economics of temptation, there might be little to be gained in delineating a constitutional economics enterprise.

It is essential to acknowledge, near the outset of discussion, that individuals choose to impose constraints or limits on their own behavior primarily, even if not exclusively, as a part of an *exchange* in which the restrictions on their own actions are sacrificed in return for the benefits that are anticipated from the reciprocally extended restrictions on the actions of others with whom they interact along the boundaries of private spaces and within the confines of acknowledged public spaces. That is to say, a domain of constitutional economics would exist even if individuals, in their private spaces, chose never to impose constraints on their own behavior. Note that by interpreting the individual's choice of a generalized constraint that restricts the actions both of others and himself as a part of a reciprocal exchange, we have moved toward the familiar domain of orthodox economics. So interpreted, the individual who joins in a collective decision to impose a generally applied constitutional rule is not, at base, acting differently from observed behavior in a setting that involves giving up one desired good, apples, for another desired good, oranges. In the latter example, we can, without violating the meaning of words, say that the individual chooses to constrain or to limit the potential consumption of apples in exchange for the expanded opportunity to consume oranges. Expressed in this way, all that is required is that we classify the restrictions on others' actions as *good* in the individual's preference function along with the more natural classification or restrictions on his own actions as *bad*.

In this simplistic and individualistic perspective, the choice of a reciprocally binding constraint by individuals who are related one to another in an anticipated

set of interactions becomes fully analogous to trade in ordinary goods and services, and, so treated, becomes quite different from the choice of a self-imposed constraint in the much more difficult economics of self-control, briefly discussed above. Note, in particular, however, that the analysis of individual choice behavior is necessarily shifted from the subject realm of the private to the public or political. The analysis becomes *political economy* almost by definition. *Constitutional economics* morphs into *constitutional political economy*.

Why have the practitioners of orthodox economics seemed so reluctant to extend analysis to include the reciprocal exchange of liberties that are central to the domain of constitutional political economy? In part, such reluctance stems from the artificial splitting between the academic disciplines of economics and political science late in the nineteenth century. Economists have been content to confine their attention to market relationships. I can advance several other and related reasons. Economists, along with their peers in the other social sciences as well as other academic disciplines, have had no difficulty, through the ages, in implicitly classifying restrictions on some of the activities of some persons in the body politic to be *good*. But the classification procedure has been quite different from the subjective evaluations presumed to be embodied in individuals' preference functions. The non-constrained voluntary behavior is not classified to be *bad* because an individual simply dis-prefers such behavior in the ordinary way. Some such behavior is deeded to be bad, and hence its rectification to be good, on the basis of an externally derived criterion of *goodness* or *truth*. The attributes or qualities of goodness and/or badness applied to actions of persons are treated as if they are intrinsically public, in the Samuelsonian taxonomic sense. An action cannot, properly, be adjudged to be good by one person without an implied generalization of such judgment to other persons. In this conceptualization, persons must, ideally, be brought into agreement on some ultimate classification of actions through a process that resembles scientific discourse. Agreement does not emerge from a trading process where different interests are essentially compromised, with each party reckoning to enjoy some benefits while suffering some sacrifice of preferred position.

In some respects, it is surprising that economists have "jumped out" of their own analytical framework so readily when they consider the possible imposition of generalized constraints on behavior. They have expressed little curiosity in deriving justification for such constraints from a calculus of individual interests. Economists have, instead, been willing intellectual captives of idealistic political philosophers, and they have readily embraced variants of the Platonic and Helenian mind-sets. Amartya Sen's (1970) usage of the term *meddlesome preferences*, by sharp contrast with such terms as *merit goods* and *merit wants*, tends to focus analysis back toward a straightforward calculus of interest and away from non-individualistic attributes of either goods or actions.

A second, and related, reason for economists' general failure to use the exchange setting when they consider the possible imposition of generalized constraints on individual behavior lies in the methodological dominance of the maximization paradigm. In the latter, *the economic problem* is defined as one of allocating scarce means (resources) among alternative ends. Choice is made necessary by the scarcity

of means, and that which is desired (utility) is maximized when like units of resources yield equivalent returns in all uses to which they are put. In this elementary formulation, emphasis is almost exclusively placed on the choices that are made within the scarcity constraints that are, themselves, presumed to be beyond the scope for chooser selection. There is little or no attention paid to the identification of the choosing unit in this abstracted definition, and this feature allows for a relatively unnoticed transference of analysis from individual choice to *social, political*, or *collective* choice on the basis of some implicit presumption that collectivities choose analogously to individuals.

This shift from individual to supra-individual choice was supported, and indirectly justified, by the emergence of macro-aggregation and macroeconomic theory and policy during the early decades of the post-Robbins half century. Target levels of macro-aggregates (national product, rate of growth, levels of employment) were established to be objectively *good* and to serve as guideposts for choices to be made by collective entities (governments) subject only to the constraints imposed by natural scarcities and technological limits. By some implicit extension of the model for individual choice behavior, constrained only by external forces, governments came to be viewed romantically and were deemed capable of achieving the *good*, as defined for them by the economists and other social philosophers. Microeconomists had long been ready at hand to proffer policy advice to governments concerning ways and means to promote greater overall economy efficiency.

A third reason for economists' general failure to extend their analytical apparatus to the derivation of institutional constitutional structure is to be found in their presumption that structural constraints are not, themselves, subject to deliberative choice, and, hence, to change. Economists have not neglected to recognize the relevance of institutional rules in affecting patterns of human behavior. Property rights economics, in particular (Alchian, 1977), has opened up a research program that concentrates attention directly on the effects of alternative structures. For the most part, however, the emphasis here is on existing arrangements rather than on the comparative analysis involved in extension to structures that might be designed and implemented.

Constitutional political economy differs from non-constitutional or orthodox economics along each of the dimensions that may be inferred from the reasons for neglect detailed above. Analysis is consistently individualistic, in the several senses that are relevant. The derivation of institutional constraints is based on a calculus of individual interests, which, in turn, requires the introduction and use of an exchange paradigm as opposed to the idealists' search for the unique *good*. Furthermore, there is no extension of the choice calculus from the individual to collectivities, as such. Collective *choice* is factored down into the participatory behavior of individual members. Finally, emphasis is centered directly on the selection of rules, or institutions, that will, in turn, limit the behavior of the persons who operate within them. Institutions, defined broadly, are variables subject to deliberative evaluation and to explicit choice (Buchanan and Tullock, 1962).

As noted, at one extreme, constitutional analysis may be applied to the individual in total isolation, who may act solely in private space. At the other extreme, constitutional analysis is applied to the whole set of persons who make up the

membership of the polity. This subcategory of research emphasis is the most familiar, since the very word *constitutional* tends to convey political connotations. The derivation of constraints on government does, indeed, occupy much of our attention. But the inclusive domain of constitutional economics also includes the derivation and analysis of and justificatory argument for rules that constrain both individual and collective behavior in a wide array of membership groupings, larger than the one-unit limit but smaller than the all-inclusive limit of the whole polity. Clubs, trade unions, corporations, parties, universities, associations—these, and many more, exist and operate under constitutions that are amenable to scientific inquiry.

Constitutional *Economics* and Constitutional *Politics*

In Section 1, I have attempted to distinguish between *constitutional* and *non-constitutional* economics or political economy. I propose, in this section, to distinguish between constitutional *economics* and constitutional *politics*, as the latter term may be generally and widely interpreted. As I have noted, most constitutional inquiry and analysis is concentrated at the level of the politically organized collectivity and is, in this sense, political. The distinction to be emphasized, however, is one of perspective rather than one that relates directly to either form of organization or to the type of activity. If an exchange rather than a maximizing paradigm is taken to be descriptive of the inclusive research program for the discipline, then *economics* involves inquiry into *cooperative* arrangements for human interaction, extending from the simplest of two-person, two-good trading processes through the most complex quasi-constitutional arrangements for multinational organizations. As noted in the first section, orthodox economics has rarely been extended to noncommercial or political activity, as such, but the exchange perspective readily allows this step to be taken.

The cooperative perspective, however, must be categorically distinguished from the contrasting *conflictual* perspective, which has been applied, almost automatically, to all political interactions, whether or not these are classified as constitutional. It will be useful here to examine the differences between the cooperative and the conflictual perspectives more carefully. The very term *politics* tends to conjure up a mental image of potential conflict among those persons who are members of the politically organized community. This conflict may be interpreted to be analogous to scientific disputes, in which separate participants or groups seek to convince one another of the *truth* of their advanced propositions. The age-old tradition of idealism in political philosophy conceives of all of politics in this light and, as noted earlier, the dominance of this model of politics has tended to discourage economists from political extensions of the exchange or cooperative paradigm. But, even if the teleological interpretation is rejected, politics may seem, by its very nature, to involve conflict between and among individuals and groups within a polity.

From the institutionally determined characteristics of collective decisions, the characteristics that dictate mutual exclusivity among the alternatives for selection (only one candidate can be electorally chosen) imply some ultimate division of the

membership into two subsets, *winners* and *losers*. This perspective almost directly suggests that politics is primarily, if not exclusively, a distributional game or enterprise—a process that involves transfers of value (utility) among and between separately identified coalitions of persons.

Note that the predominance of the distributional elements in the conflictual model of politics need not imply that the game be zero sum, although this limiting case may be useful for some analytical purposes. Conflictual politics may be positive, zero, or negative sum, as gains and losses are somehow aggregated over all participants (members). And this seems to be the natural model for analyzing politics so long as rules for reaching collective decisions require less than full agreement. If a majority, whether simple or qualified, is allowed to be decisive and impose its will on a majority, then the observed opposition of the minority to the alternative preferred by the majority can be taken to indicate that members of the minority expect to suffer utility losses, at least in a lost opportunity sense. In this model of conflictual politics, which appears to be descriptive of ordinary political activity, there seems to be no direct way of introducing a cooperative interpretation. A necessary condition for cooperation in social interaction is the prospect for positive expected gains by all parties, or, in the gainer—loser terminology, the prospect that there be no net losers. At a first descriptive cut, this condition seems to be foreign to the whole political enterprise.

It is precisely at this point, however, that constitutional politics, or politics at the constitutional level of choices among alternative sets of basic rules or constraints, rescues the cooperative model, at least in some potential explanatory and normative sense. As it operates and as we observe it to operate, ordinary politics may remain conflictual, in the manner noted above, while participation in the inclusive political game that defines the rules for ordinary politics may embody positively valued prospects for all members of the polity. In other words, constitutional politics does lend itself to examination in a cooperative analytical framework, while ordinary politics continues to lend itself to analysis that employs conflict models of interaction.

Generalized agreement on constitutional rules that allow for the reaching of ordinary collective decisions by means that do not require general agreement is surely possible, as is empirically demonstrated in the context of almost all organizations. The analytical-scientific inquiry that involves comparisons of the working properties of alternative sets of rules along with the examination of processes through which agreement on rules may be attained defines the domain of primary concern. The usage of the terminology *constitutional political economy* rather than the somewhat more accurate *constitutional politics* is prompted by the linkage in scientific heritage between *economics* and *cooperation*, by the inference of the appropriateness of the exchange as opposed to the conflict paradigm.

The Intellectual Traditions of Constitutional Political Economy

In sections 17.1 and 17.2, I have attempted to set the research program in constitutional political economy apart from ongoing programs within the interrelated and more inclusive disciplines of economics and political science. It would

be totally misleading, however, to infer from my discussion that this research program has emerged full blown, as if divorced from any traditions of intellectual inquiry. As I have noted, constitutional political economy did indeed blossom only in the second half of the century. But the program was not based either on a new scientific discovery, at least as usually defined, or on a new set of analytical tools. Constitutional political economy is best interpreted as a re-emphasis, a revival, a rediscovery of basic elements of earlier intellectual traditions that have been put aside, neglected, and sometimes forgotten in the social sciences and social philosophy. These traditions are those of classical political economy and contractarian political philosophy.

Classical political economy, represented especially in the works of Adam Smith (1776), was directed toward offering an explanation and understanding of how an economy (set of markets) would work without detailed political interventions and control. Smith's aim was to demonstrate that the *wealth* of the nation would be larger under a regime of minimal politicization than under the alternative closely controlled mercantilist regime. And the whole thrust of the argument was to the effect that all groups in the economy and especially the laboring classes, could be expected to share in the benefits promised upon the shift in regimes. The emphasis was on the generalization of expected gains over all persons and classes. The suggested change in the structure, or basic rules, that depoliticization involves was, therefore, within the feasible limits of potential agreement by all parties. The normative focus, again especially in Adam Smith, was not explicitly distributional. Only with Marxian extensions of Ricardo's abstract analysis did interclass conflict enter into classical attention.

It is also important to recognize that the Smithean emphasis was not allocational in the modern economists' meaning of this term. The analysis was not designed to show that economic resources would be more effectively allocated to higher valued uses under a market than under a politicized regime, as measured by some external and objective standard of value. The aim was, instead, to show that the market order would allocate resources such that the evaluations (preferences) of individuals would be more fully satisfied, *regardless of what these evaluations might be*. In terms of his familiar example of the butcher, Smith's lesson was to show that self-interest in the marketplace works to supply meat for supper, provided that meat is what consumers want. There is no implication here that self-interest in the marketplace works to supply meat because meat is valuable in some nutritional sense as defined by experts.

So interpreted, therefore, Adam Smith's enterprise falls squarely within the domain of constitutional political economy. In a strictly positive sense, his analysis described both how the existing regime worked and how an alternative regime might work. And, since the alternative seemed to generate more wealth to all parties, as measured by their own standards, the normative extension of the positive analysis was quite straightforward. In this extension, the object upon which collective attention must be placed is the set of rules or constraints within which persons behave in their capacities as consumers-buyers and producers-sellers. The laws and institutions that define the economic-political order become the variables subject to possible adjustment and reform.

I have selected elements from the tradition of classical political economy that seem to provide precursory foundations for the modern research program in constitutional political economy. My treatment would surely be accused of bias, however, if I failed to indicate the presence of considerable ambiguity and confusion in the philosophical underpinnings of the classical economics enterprise. An interpretation of that enterprise in terms of classical utilitarianism would be quite different from my own; this alternative interpretation would stress quite separate elements of the tradition. The interpersonal comparability and aggregate measurability of utility were not explicitly rejected by the classical economists and, in a selected reading, these may be attributed, as presumptions, to their analyses. In this case, the whole enterprise becomes precursory to the maximizing rather than to the exchange paradigm in economics, with both allocational and distributional implications, and with a wholly different avenue for moving from the individual to the collective levels of choice. The categorical distinction between choices among rules and choices within rules all but disappears in the utilitarian configuration.

The elements of Adam Smith's intellectual enterprise become directly precursory to the research program in constitutional economics only when these elements are imbedded within the tradition of contractarian political philosophy, the tradition that was developed prior to but became competitive with and quite different from classical utilitarianism. From the seventeenth century, from the works of Althusius, Hobbes, Spinoza, and Locke in particular, attempts were made to ground justificatory argument for state coercion on agreement by those individuals who are subject to coercion. This intellectual tradition invented the autonomous individual by shucking off the communitarian cocoon. The assignment to the individual of a capacity for rational independent choice, as such, allowed a *science* of economics and politics to emerge—a *science* that embodied a legitimatizing explanation for the emergence of and existence of the state. In agreeing to be governed, explicitly or implicitly, the individual exchanges his own liberty with others who similarly give up liberties in exchange for the benefits offered by a regime characterized by behavioral limits.

The contractarian logic leaves open any specification of the range and scope for consent-based coercive authority. The early contractarians, and notably Hobbes, had no understanding of the efficacy of market order as it might function under the umbrella of the protective or minimal state. This understanding was provided only in the eighteenth century and was fully articulated only in the great work of Adam Smith. Classical political economy, as appended to the contractarian intellectual foundations, allowed the development of a scientifically based analysis aimed at comparing alternative structures of political-legal order—analysis that could introduce and use principles of rational choice behavior of individuals and without resort to supra-individualistic norms. Utilitarianism also rejected all supra-individual norms, as such, and grounded all norms in a calculus of pleasure and pain. Nonetheless, this Benthamite intrusion created ambiguity in the efforts to add up utilities over persons. In this way, the contractarian justification derived from conceptual agreement was obscured, and the way was opened for a non-transcendental utilitarian supersession of individualistic norms. The contractarian philosophical basis upon

which classical political economy should have been exclusively developed was, at least partially, undermined and neglected for almost two centuries, only to be rediscovered in the research program of constitutional economics.

The Hard Core and Its Critics

Throughout this article I have referred to constitutional political economy as a *research program*, thereby deliberately using the Lakatosian classification. In this scheme, there exist elements in the hard core of the program that are rarely, if ever, challenged by those scholars who work inside the intellectual tradition defined by the program. These central elements are taken as presuppositions, as relatively absolute absolutes, and, as such, they become, themselves, the constraints (the constitution) within which the scientific discourse is conducted. External intellectual challenges to the whole enterprise tend to be directed at these elements in the core of the program. The ongoing research within the constraints can, of course, proceed without concern for these external criticisms, but practitioners need to be aware of the core-imposed limits on the persuasive potential of the internalized analytical exercise.

For constitutional political economy, the foundational position is summarized in *methodological individualism*. Unless those who would be participants in the scientific dialogue are willing to locate the exercise in the choice calculus of individuals, *qua* individuals, there can be no departure from the starting gate. The autonomous individual is a *sine qua non* for any initiation of serious inquiry in the research program. Individual autonomy, as a defining quality, does not, however, imply that the individual chooses and acts as if he or she exists in isolation from and apart from the community or communities of other persons with whom he or she may be variously associated. Any form of community or association of individuals may reflect some sharing of values, and, further, any individual's formation of values may be influenced by the values of those with whom he or she is variously associated in communities. The communitarian challenge to methodological individualism must go beyond the claim that individuals influence one another reciprocally through presence in communities. The challenge must make the stronger claim that individuation, the separation of the individual from community, is not conceptually possible, that it becomes meaningless to think of potential divergence between and among individual interests in a community. Stated in this way, it is evident that methodological individualism, as a presupposition of inquiry, characterizes almost all research programs in economics and political science; constitutional economics does not depart from its more inclusive disciplinary bases in this respect.

The communitarian critique does not often appear in such blatant guise. For constitutional political economy, in particular, the critique apparently leaves the individualistic postulates unchallenged, while either implicitly or explicitly asserting the existence of some supra-individualistic source of evaluation. Individual

evaluations are superseded by those emergent from God, natural law, right reason, or the state. This more subtle stance rejects methodological individualism, not on the claim that individuation is impossible, or that individual evaluations may not differ within a community, but rather on the claim that it is normatively improper to derive collective action from individual evaluations. To the communitarian who posits the existence of some supra-individualistic value scale, the whole analysis that builds on a base of an individualistic calculus can only be useful as an input in schemes of control and manipulation designed to align individualized preferences with those orderings dictated by the overarching norms for the community.

Concomitant with methodological individualism as a component of the hard core is the postulate of rational choice—a postulate that is shared over all research programs in economics. The autonomous individual is also presumed to be capable of choosing among alternatives in a sufficiently orderly manner as to allow a quality of rationality to be attributed to observed behavior. For constitutional economics, the capacity for rational choice is extended to include a capacity to choose among constraints, both individually and collectively applied, within which subsequent choices may be made.

Rationality implies that choices may be analyzed as if an ordering of alternatives exists, arrayed in accordance with some scalar of *preferredness*. We may, but need not, use the term utility to designate that which the individual calls upon to make up the ordinal ranking. At the analytical level, there is no need that the ranking correspond with any array of the choice alternatives that may be objectively measurable by some outside observer. The test for individual rationality in choice does require, however, the minimal step of classifying alternatives into *goods* and *bads*. The central rationality precept states only that the individual chooses more rather than less of *goods*, and less rather than more of *bads*. There is no requirement that rationality dictates choice in accordance with the individual's economic interest, as this might be measured by some outside observer of behavior.

The individualistic postulate allows the interests or preferences of individuals to differ, one from another. And the rationality postulate does not restrict these interests beyond the classificatory step noted. *Homo economicus*, the individual who populates the models of empirical economics, may, but need not, describe the individual whose choice calculus is analyzed in constitutional political economy. When selecting among alternative constitutional constraints, however, the individual is required to make some predictions about the behavior of others than himself. And, in such a setting, there is a powerful argument that suggests the appropriateness of something akin to the *Homo economicus* postulate for behavior (Brennan and Buchanan, 1985).

I have briefly discussed the individualistic and the rationality presuppositions for the research program. These elements are not controversial, and they would be listed as components of the hard core both by practitioners and critics of constitutional economics. A less obvious element that is, however, equally fundamental involves the generalization of the individualistic and the rationality postulates to *all* persons in the political community. All individuals must be presumed capable to make rational choices among alternatives in accordance with individually

autonomous value scales. And this generalization does not allow derivation of collective action, whether or not directed toward choices among constraints, from individual evaluations on anything other than an *equal weighting*. To introduce a weighting scheme through which the evaluation of some persons in the community are deemed more important than other persons would require resort to some supra-individualistic source, which is, of course, ruled out by adherence to the individu-alistic postulate. In this sense the whole of the constitutional economics research program rests squarely on a *democratic* foundation.

Perception, Vision, and Faith

Nietzsche used the metaphor of viewing the world of reality through differing windows (Kaufman, 1950, p. 61), and Ortega y Gasset went so far as to define ultimate reality itself as a perspective (Ortega y Gasset, 1961, p. 45). In a sense, any research program involves a way of looking at, and thereby imposing an order on, that which is perceived. This characterization applies particularly to any program in social science, where the ultimate object of inquiry is behavior in a social interac-tion process. I have on several occasions referred to the *constitutional perspective*, which I have acknowledged to be different from other perspectives that might be used in examining and evaluating the interaction of individuals in social and/or political settings. This elementary fact that perspectives differ, or may differ, raises difficult issues in epistemology that cannot be ignored.

Consider, first, perception at its simplest level. Presumably, individuals are sufficiently alike, one to another, biologically that we see, hear, taste, smell, and feel physical phenomena similarly, if not identically. We all see a wall as a barrier to movement, and no one of us makes an attempt to walk through walls. Someone who failed to perceive a wall as the others of us would be classified to be abnormal in at least one of the basic perceptual senses. As phenomena come to be increas-ingly complex, however, individuals may come to differ in their perceptions, despite the fact that, biologically, they continue to possess the same perceptual apparatus. Elementary sense perception must be accompanied by imaginative constructions that require some mental processing before a basis for evaluation, and ultimately for action, can be established.

As phenomena increase in complexity, the imaginative elements in perception increase relative to those that emerge directly from the senses. In this progression from the simple to the complex, the similarity in perceptions among persons must decrease. What may be called the *natural* way of observing phenomena fades away at some point along the spectrum. Individuals may then be brought into agreement on that which they observe only by entry into some sort of association of shared values or norms, which members, either explicitly or implicitly, choose. This statement may seem contradictory when first made; it may seem to state that persons choose how they see reality. But the statement becomes less challenging to ordinary notions when we replace *see* with *think about*.

I have been accused of committing the naturalistic fallacy, in some of my own works, of failing to respect properly *the fact—value, positive—normative* distinction, and, hence, of deriving the *ought* from the *is*, at least implicitly. I submit, however, that my critics mount such charges only because of their own confusion about the nature of perception of complex phenomena. If there exists no *natural* way of observing reality, some evaluation and choosing process is a necessary complement to the imaginative step that allows apparent chaos to be converted into order. We select the *is* that defines the hard core of our research program, and this holds true whether or not we are professional scientists. Within this *is*, we can adhere strictly to the precepts laid down for positive analysis. But the normative implications that may be drawn are, indeed, derivative from the chosen perceptive framework, and could not, or would not, be otherwise available.

Constitutional political economy is a domain of inquiry and discourse among scientists who choose to perceive social interaction as a set of complex relationships, both actual and potential, among autonomous persons, each of whom is capable of making rational choices. The domain, as such, cannot be extended to include inquiry by those who choose to perceive social interaction differently. There is simply no common basis for scientific argument, and ultimately agreement, with those who choose to perceive social interaction either in purely conflictual or purely idealistic visions. These visions are, indeed, alternative 'windows' on the world, and the process through which individuals choose among such windows remains mysterious. How can empirical evidence be made convincing when such evidence must, itself, be perceived from only one vantage point at a time? The naivete of modern empirical economists in this respect verges on absurdity.

When all is said and done, *constitutional political economy* must be acknowledged to rest upon a pre-commitment to, or a faith in, man's cooperative potential. Persons are neither bees in hives, carnivorous beasts in a jungle, nor angels in God's heaven. They are independent units of consciousness, capable of assigning values to alternatives, and capable of choosing and acting in accordance with these values. It is both physically necessary and beneficial that they live together, in many and varying associations and communities. But to do so, they must live by rules that they can also choose.

References

Alchian, A. (1977). *Economic Forces at Work*. Indianapolis: Liberty Fund.

Althusius, J. (1932). In C.J. Friedrich (ed.) *Politica Methodica Digesta*. Cambridge: Harvard University Press.

Brennan, G. and Buchanan, J.M. (1985). *The Reason of Rules: Constitutional Political Economy*. Cambridge: Cambridge University Press.

Buchanan, J.M. (1987). *"Constitutional economics." The New Palgrave*. London: Macmillan.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Elster, J. (1979). *Ulysses and the Sirens*. Cambridge: Cambridge University Press.

Hobbes, T. (1943). *Leviathan*. London: Everymans Library.

Kaufman, W. (1950). *Nietzsche*. Princeton: Princeton University Press.

Locke, J. (1955). *Second Treatise of Civil Government*. Chicago: Gateway.

McKenzie, R. (ed.) (1984). *Constitutional Economics*. Lexington, MA: Lexington Books.

Ortega y Gasset, J. (1961). *Meditations on Quixote*. New York: Norton.

Schelling, T. (1978). "Egonomics, or the art of self management." *American Economic Review*, 68, 290–294.

Sen, A.K. (1970). "The impossibility of a paretian liberal." *Journal of Political Economy*, 78, 152–157.

Smith, A. (1979). *The Wealth of Nations*. Oxford: Clarendon Press.

Spinoza, B. (1854). *A Treatise in Politics*. London: Holyoake. (Trans. William McCall.)

Thaler, R. and Shefrin, H.M. (1981). "An economic theory of self-control." *Journal of Political Economy*, 89, 392–406.

Chapter 18
Spatial Theory[1]

Melvin J. Hinich and Michael C. Munger

> Assume that every voter's preferences are single-peaked and slope downward monotonically on either side of the peak (unless his peak lies at one extreme of the scale). ...The best way [for each party] to gain more support is to move toward the other extreme, so as to get more voters outside of it—i.e., to come between them and its opponent. As the two parties move closer together, they become more moderate and less extreme in policy in an effort to win the crucial middle-of-the-road voters, i.e., those whose views place them between the two parties. This center area becomes smaller and smaller as both parties strive to capture moderate votes; finally the two parties become nearly identical in platforms and actions. (Downs, 1957, pp. 116–117)

One of the fundamental building blocks in the analysis of political phenomena is the representation of preferences. Without some means of capturing the essence of goals and trade-offs for individual choices, the mechanics of the public choice method are stalled. While there are many ways of representing preferences, the single most commonly used approach is the "spatial" model. The idea of conceiving preference in a kind of "space" is actually quite ancient, as the quote from Aristotle's *Politics* below shows. Furthermore, there are hints of several topics of modern spatial theory, including the power of the "middle," and the problem of instability in political processes.

Origins of Spatial Political Competition

It is important to recognize that the spatial model is not just an "as-if" form of reasoning about political phenomena. One can quickly find myriad references to "left" and "right" in political discourse, both in the media and in elite accounts. Although "space" is a metaphor, it is one that is used so widely that it must connect quite closely with human cognition about political representation.

---

[1] This chapter is a reprint of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 305–311.

The origin of the "left" and "right" metaphors, as is well known, is a reference to the physical positions occupied by different factions in the French National Assembly after 1789. The *Girondins* on the "right" of the huge meeting hall held power, and ran the government. The more radical *Jacobin* allies of Robespierre sat in the "mountain" on benches rising up the wall on the far "left". The *Jacobins* on the left were constantly agitating for change, while the *Girondins* on the right defended stability and the status quo. With only a very little adjustment for time and circumstance, these meanings still attach to "left" and "right" in political discourse today.

An alternative meaning, mapping an ideological left and right onto positions with respect to ownership of capital (right) and defense of labor (left), was created by Karl Marx, and is used today in a wide variety of surveys. This meaning, however, is at best misleading and is often simply incorrect. The former Soviet Union, particularly Russia, had a clearly defined left and right at the end of the twentieth century. But "left" was understood to mean liberal reformers who favored markets and democracy, whereas the "right" was composed of former communists who demanded a return to central planning and a secure and stable, if unelected, government.

The Problem of Representation

In economics, the problem of representation of preferences has been refined to the point that is simply a mathematical problem. Suppose that there are many alternatives, and that for each pair of alternatives, I prefer one, or like them equally. Then it is possible (assuming transitivity) to construct an aggregate weak ordering that allows the individual to "rank" alternatives from best to worst, with each alternative either uniquely or with a group of other alternatives associated with an ordinal level. If I like A better than B, we say, "A is preferred to B". If I like them equally well, then "I am indifferent between A and B".

"Representing" the preferences implied by this ranking requires the assignment of any mathematical function f that has the following properties: (1) If A is preferred to B, then $f(A) \geq f(B)$. (2) If A and B are equally preferred, then $f(A) = f(B)$. As should be obvious, if there exists at least one function that represents these preferences, there will be infinitely many (since, for example, f and $f' = (f/2) + 37$ both have the same ordering of the index numbers associated with alternatives). In other words, any order-preserving transformation of a function that represents the preferences is equally good.

The work on preferences in economics has shown that relatively few assumptions are required to ensure representability. One common, and plausible, type of preferences that is not representable by a mathematical function is "lexicographic" preferences, but most preferences that obey simple convergence criteria are representable. Economic preferences, however, generally assume either non-satiety or
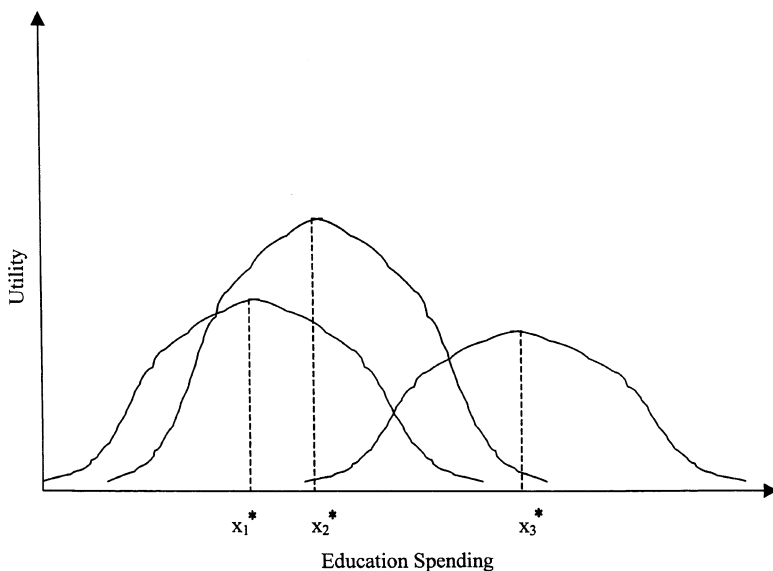
Fig. 18.1 Utility functions with interior ideal points x_i^* for education, three citizens

free disposal. Can something like the same approach be used to "represent" *political* preferences, which may very well require interior ideal points?

To understand the problem, consider the difference between preferences for apples and preferences for education. We generally model preference for apples as nondecreasing, so that more is preferred to less. What about education, or more accurately education budget? If asked, most citizens will not say that they think that the education budget should be infinite. Instead, they will select some finite number of dollars they think is the "best" budget, and will argue that either a larger or smaller budget is less preferred. This "interior" ideal point is illustrated in Fig. 18.1. (The utility functions graphed in the figure are symmetric for the sake of simplicity, but there is nothing in spatial theory that requires symmetry.)

One important research question in public choice is the relation of political preferences to economic preferences. More specifically, can political preferences with interior ideal points be derived from economic preferences, with the connection being the opportunity cost of taxes used to finance public programs? Interestingly, though the initial findings were hopeful (Barr and Davis 1966), the answer turns out to be "no" (Denzau and Parks 1977, 1979; Slutsky 1977; Hinich and Munger 1994, Ch. 2).

Consequently, the basis of political preferences in "representation" is more tenuous than for economic preferences: the only way to justify interior ideal points rigorously is to connect preferences for the good with the recognition that the financing scheme requires that citizens also pay for the good. This means that preferences are being defined simultaneously over the underlying good and the price of the good, rather than just preferences for the good alone.

On the other hand, the mathematical underpinnings for political preference representation in a "space" are well defined and consistent, requiring only minimal assumptions about the convexity of the sets of alternatives enclosed by indifference curves (Schofield, 1984). To the extent that government decisions on property rights, security, and a currency system are logically antecedent to the problem of representing economic preferences, the problem may go the other way. That is, there is a failure of duality in the representation problem: it is perfectly easy to take preferences in either the public or private sectors as primitive, and then use utility functions to represent them. As the literature cited in this section shows, once one starts with economic preferences, there is no consistent way to "induce" public sector preferences. However, it is equally true that if one takes political preferences as primitives, then it is the "induced" economic preferences that exhibit inconsistency. We will take spatial theory as a primitive, rather than induced or derived, means of representing public sector preferences.

The First Spatial Models

As Hinich and Munger (1994) point out, the first clear use of the spatial "model" appears in Aristotle's *Politics*, written down before 325 B.C.E., and perhaps amended and modified in several ways in the centuries that followed. Still, as the work comes down to us, it is clear that there is both a deep understanding of politics and stability, and a connection to the idea of a dimension, or simple space, that organizes political conflict.

> Now in all states there are three elements: one class is very rich, another very poor, and a third in a mean. It is admitted that moderation and the mean are best, and therefore it will clearly be best to possess the gifts of fortune in moderation; for in that condition of life men are most ready to follow rational principle...[T]hose states are likely to be well-administered, in which the middle class is large, and stronger if possible than both the other classes, or at any rate than either singly, for the addition of the middle class turns the scale, and prevents either of the extremes from being dominant...The legislator should always include the middle class in his government; if he makes his laws oligarchical, to the middle class let him look; if he makes them democratical, he should equally by his laws try to attach this class to the state. There only can the government ever be stable where the middle class exceeds one or both of the others, and in that case there will be no fear that the rich will unite with the poor against the rulers. (Aristotle, 1979, pp. 138–142)

As was discussed earlier, this understanding of politics seems to come naturally to human beings, with the clearest example deriving from the language used to describe the conflict in the French Assembly. But what of models? How are we to think of the idea of a "space", or dimension of conflict, in a way that gives us testable propositions about political behavior and institutions?

The early literature in economics on "spatial" competition addressed what seem like similar considerations. Hotelling (1929), Lerner and Singer (1937), and Smithies (1941) all addressed the problem of location, in the sense that a set of firms

selling zero-cost, undifferentiated products might compete by choosing the physical setting for the business. The classic metaphor is the choice of two hot dog stands on a street or beach, with potential patrons distributed along the linear dimension of competition. The key assumption is that, since the products are undifferentiated (all hot dogs are of the same make), patrons will choose solely on the basis of location. The equilibrium set of locations, as was shown by various means in this literature, was achieved when (in the case of two firms) the businesses converged to a "central place". With more than three competitors, the results are ambiguous (there are many possible equilibria), and with arbitrarily many firms very little can be said.

The interesting thing about the early spatial models in economics was the fact that the authors worked to develop normative implications. An important controversy was Hotelling's rather strong claim that capitalism was "wasteful", at least compared to a planned economy. The prediction of convergence implied that both (or all) firms ended up as close together as they could manage, increasing the average distance traveled by consumers. Hotelling concluded: "Our cities become uneconomically large and the business districts within them are too concentrated. Methodist and Presbyterian churches are too much alike; cider is too homogeneous" (p. 57).

Lerner and Singer (1937) disputed this claim, pointing out that it rested on tenuous assumptions (particularly about transport charges and the extreme inelasticity of demand assumed by Hotelling). Smithies (1941) pursued the matter further, showing that under some plausible assumptions there exist nonconvergent equilibria.

Unfortunately, the problems of spatial location for firms and spatial preference representation in politics are not isomorphic. The analogies in results are not very useful, and can be misleading. The idea that voters might choose the candidate "closer" to their own ideal seems plausible enough, but it is by no means clear what "close" means once the idea of simple Euclidean distance is dispensed with. Euclidean distance makes good sense in the hot dog stand competition, since it takes just as long to walk 100 yards north as it does to walk 100 yards south. But it is by no means clear that we would want to build on this extreme kind of symmetry in representing political spatial preferences.

The problem is worse if there are multiple dimensions. Euclidean distance makes two assumptions about preferences:

(1) Separability—My evaluation of issue i is not affected by the level of issue j I expect to result from the decision process.
(2) Equal salience—Marginal changes in issue i have the same increment/decrement for my utility as marginal changes in issue j.

Neither of these is a problem for the spatial location problem, because my reaction to having to travel is based on distance, not whether the distance is in any one direction. But if we are to use a policy "space" to represent political preferences, the assumptions of separability and equal salience are both empirically unrealistic and theoretically limiting.

The extension of this kind of reasoning to political problems, particularly of party competition, was accomplished by Downs (1957). It is clear, however, that

Downs' analysis is of a piece with the earlier work; consider Smithies' first paragraph:

> The very fact that Professor Harold Hotelling's pioneer article explained so successfully the close similarity of the Republican and Democratic platforms in 1928 indicates that something more was needed in 1936. It was probably true to say in 1928 that by moving to the center of electoral opinion neither party risked losing its peripheral support. The situation at the present time requires no elaboration; suffice it to say that neither party feels itself free to compete with the other for the undecided votes at the center, in full confidence that it will retain its support from the extremes of political opinion.

This is a very sophisticated statement, recognizing that equilibria, if they exist, will depend on the reliability of turnout and support from those at the extremes. If, to use the economic analogy for the last time, the "elasticity of demand" of citizens is high, moving toward the center may actually reduce one's vote share, as the ardent supporters out in the wings lose interest. To be fair, Downs concentrated on the problems of turnout, and information, but Downs has come to be associated with the result that candidates converge to the middle, or median, in two-party elections. It has since been shown (Berger et al., 2000) that the convergence result is actually very fragile under the plausible set of "Downsian" assumptions, and unlikely to be observed empirically.

A Rigorous Representation: Spatial Theory in the 1960s

The first rigorous statement of the spatial model as a representation of preferences, at a level of generality analogous to that of economics, was the result of the collaboration of Otto Davis and Melvin Hinich. In three papers (Davis and Hinich, 1966, 1967, 1968), they laid the groundwork for what is now thought of as spatial theory.

Using a generalized quadratic form for representing preferences, they were able to account for nonseparability and differences in salience in an elegantly simple way. Further, in all three papers, but particularly in the 1968 piece, they addressed the normative problem of the "good" in the democratic choice problem: if we accept the idea of Aristotle's "mean" as the best choice for a democracy, there is a benchmark against which predicted outcomes in the spatial model can be compared. The most widely recognized paper in this collaboration, Davis et al. (1970) is a general exposition of all the results in the series of papers, with some extensions, and is the generic original reference in the spatial theory literature.

Social Choice Theory

Social choice theory and spatial theory are related subjects, but there are many important differences. Social choice theory tends to focus on the consequences of aggregation of individual "lists", using different aggregation ("voting") rules.

Very little restriction is placed on the form that these lists can take, other than each weak order is transitive. Some of the most important work, such as Arrow (1963), actually assumes explicitly that preferences are characterized by "universal domain", so that any ordering over elements of the choice set is possible. There is no requirement, in social choice theory, that the preferences are "representable". Instead, social choice theorists work directly with preference orderings themselves.

Spatial theory, on the other hand, focuses on preferences that are single-peaked, and which are amenable to mathematical representation. The simplest kind of spatial preferences, Euclidean preferences, make a very restrictive set of assumptions about the kind of function that can represent the underlying ordering over alternatives. The simplest way to think of the difference, then, is that social choice theory takes preferences as primitive, and unknown, with any ordering equally likely. Spatial theory uses the notion that "closer" alternatives are more preferred, though spatial theory can account for weighted Euclidean distance, so that the function representing preferences exhibits nonseparability and different salience for different issues.

There are some important overlaps between spatial theory and social choice theory. An early work, in many ways ahead of its time, was Black and Newing (1951). This book introduced something very close to the analytical tool now called "win sets", but at the time too little was known about the problems of aggregation to give a coherent account. Black recognized the limitations in the earlier work, and published his seminal *Theory of Committees and Elections* in 1958, though this book took more of a social choice than a spatial perspective.

Probably the best known example of the intersection of spatial and social choice theory is Plott's (1967) then revolutionary exposition of the problem of the nonexistence of equilibrium under most arbitrarily chosen configurations of voter ideal points (Enelow and Hinich, 1983). This paper led to a new research agenda, trying to identify some subset of the policy space that is likely to contain outcomes, if not unique equilibria, of majority rule voting processes. For example, Schofield (1978, 1984) offered a mathematically more general treatment than that of Plott, but worked mainly within the logic of the spatial representation of preferences. McKelvey (1976a, b, 1979, 1986) generalized the concept of spatial equilibrium, and distilled some important solution concepts, including covering and dominance. The notion of the "uncovered set" in a spatial context derives from Miller (1980); for a review and some extensions, see Cox (1987).

Extensions

The spatial model has been extended in a number of useful ways, a review of which would extend beyond the scope of this short chapter. Useful, though very different, reviews of the literature can be found in Coughlin (1992), Enelow and Hinich (1984, 1990), Hinich and Munger (1997), and Ordeshook (1986, 1997). But a brief list of extensions is worthwhile.

- One of the earliest, and most interesting, is the extension of the spatial model to account for the turnout decision, allowing for rational abstention. Hinich et al. (1973) gather together many strands of literature, and raise some important questions about the notion of equilibrium in the spatial model.
- The idea of treating voter actions as outcomes of an idiosyncratic probability distribution function arises naturally from the Hinich, Ledyard, and Ordeshook work, and was taken up by Hinich (1977), and Enelow and Hinich (1989).
- The restriction of the "space" of conflict to only a few dimensions, based on the empirical phenomenon of clustering of issues, has resulted in two related, yet distinct, theoretical extensions of the spatial model. The idea that "ideologies" are important for explaining mass behavior was developed by Enelow and Hinich (1984), and given a firmer theoretical foundation by Hinich and Munger (1994). The claim that "ideology" is an important empirical predictor of both the vote of members in Congress and of the structure of the space of competition itself can be found in Poole and Rosenthal (1996), which reviews Poole and Rosenthal's many previous contributions to the development of this idea.
- The spatial model has an important policy implication for agenda control, because it allows analysis of the role of the "setter". There have been many contributions on this point, but the most important is Romer and Rosenthal (1978). A review of the larger literature, and its importance, can be found in Rosenthal (1990).
- Finally, the spatial model has given rise to a number of tests using experimental methods and human subjects. A review of this literature can be found in McKelvey and Ordeshook (1990). The important thing about experimental work in the spatial model is that it can suggest patterns of outcomes empirically, since many of the theoretical results are simply negative, because of the absence of equilibria.

Conclusion

Spatial theory is the single most important analytical construct for representing citizen preferences over policies, public goods, and government actions. Though the mathematical generality of spatial models falls short of the standards of preference representation in economics, it is important to recognize three things. First, the problem of representing political "preferences" is inherently more difficult than representing economic preferences. Thus, it is not clear whether our models are not very good, or the problem is just very hard. Second, spatial models perform very well in a wide variety of useful theoretical settings, and can be used to investigate the precise properties of different institutional arrangements, ranging from committee systems in legislatures to the assignments of ministry portfolios in parliamentary governments, and encompassing voting by the mass public on referenda or elections.

Finally, the spatial model is appealing because of its inherent verisimilitude. The notion of "left" and "right" as a description of the "location" of candidates or parties is nearly universal. The notion of "moving to the center" or "outflanking on the left/right" pervades media and elite discourse about politics. For all these reasons, knowledge of the basic results of spatial theory is one of the foundations of public choice theory.

References

Aristotle ([c. 350 B.C.] 1979). *Politics and Poetics* (translated by Benjamin Jowett and S.H. Butcher). Norwalk, CT: Easton Press.

Arrow, K. J. (1951, 1963). *Social Choice and Individual Values*. New Haven: Yale University Press.

Barr, J. and O. Davis (1966). "An elementary political and economic theory of the expenditures of local government." *Southern Economic Journal*, 33, 149–165.

Berger, M., M. Munger, and R. Potthoff. (2000). "The Downsian model predicts divergence." *Journal of Theoretical Politics*, 12, 78–90.

Black, D. ([1958], 1987). *The Theory of Committees and Elections*. Dordrecht: Kluwer.

Black, D. and R. A. Newing. (1951). *Committee Decisions With Complementary Valuation*. London: Lowe and Brydon.

Coughlin, P. (1992). *Probabilistic Voting Theory*. New York: Cambridge University Press.

Cox, G. (1987). "The core and the uncovered set." *American Journal of Political Science*, 31, 408–422.

Davis, O. A. and M. Hinich. (1966). "A mathematical model of policy formation in a democratic society," in J. Bernd (ed.) *Mathematical Applications in Political Science*, Volume II. Dallas: Southern Methodist University Press, pp. 175–208.

Davis, O. A. and M. Hinich. (1967). "Some results related to a mathematical model of policy formation in a democratic society," in J. Bernd (ed.) *Mathematical Applications in Political Science*, Volume III. Charlottesville: University of Virginia Press, pp. 14–38.

Davis, O. A. and M. J. Hinich. (1968). "On the power and importance of the mean preference in a mathematical model of democratic choice." *Public Choice*, 5, 59–72.

Davis, O. A., M. J. Hinich, and P. C. Ordeshook. (1970). "An expository development of a mathematical model of the electoral process." *American Political Science Review*, 64, 426–448.

Denzau, A. and R. Parks. (1977). "A problem with public sector preferences." *Journal of Economic Theory*, 14, 454–457.

Denzau, A. and R. Parks (1979). "Deriving public sector preferences." *Journal of Public Economics*, 11, 335–352.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Enelow, J. M. and M. J. Hinich. (1983). "On plott's pair-wise symmetry condition for majority rule equilibrium." *Public Choice*, 40(3), 317–321.

Enelow, J. and M. J. Hinich (1984). *Spatial Theory of Voting: An Introduction*. New York: Cambridge University Press.

Enelow, J. M. and M. J. Hinich. (1989). "A general probabilistic spatial theory of elections." *Public Choice*, 61, 101–113.

Enelow, J. M. and Hinich, M. J. (eds.) (1990). *Advances in the Spatial Theory of Voting*. New York: Cambridge University Press, 1–11.

Hinich, M. J. (1977). "Equilibrium in spatial voting: the median voting result is an artifact." *Journal of Economic Theory*, 16, 208–219.

Hinich, M. J., Ledyard, J. O., and Ordeshook, P. C. (1973). "A theory of electoral equilibrium: a spatial analysis based on the theory of games." *Journal of Politics*, 35, 154–193.

Hinich, M. and M. Munger. (1994). *Ideology and the Theory of Political Choice*. Ann Arbor, MI: University of Michigan Press.

Hinich, M. and M. Munger. (1997). *Analytical Politics*. New York: Cambridge University Press.

Hotelling, H. (1929). "Stability in competition." *Economic Journal*, 39, 41–57.

Lerner, A. P. and Singer, H. W. (1937). "Some notes on duopoly and spatial competition." *The Journal of Political Economy*, 45, 145–186.

McKelvey, R. (1976a). "General conditions for global intransitivities in formal voting models." *Econometrica*, 47, 1085–1111.

McKelvey, R. (1976b). "Intransitivities in multidimensional voting bodies and some implications for agenda control." *Journal of Economic Theory*, 30, 283–314.

McKelvey, R. (1979). "Covering, dominance, and institution-free properties of social choice." *American Journal of Political Science*, 30, 283–314.

McKelvey, R. (1986). "General conditions for global intransitivities in formal voting models." *Econometrica*, 47, 1085–1111.

McKelvey, R. and P. Ordeshook (1990). "A decade of experimental results on spatial models of elections and committees," in Enelow and Hinich (eds.) *Advances in the Spatial Theory of Voting*. New York: Cambridge University Press, pp. 99–144.

Miller, N. (1980). "A new solution set for tournament and majority voting." *American Journal of Political Science*, 24, 68–96.

Ordeshook, P. C. (1986). *Game Theory and Political Theory*. New York: Cambridge University Press.

Ordeshook, P. C. (1997). "The spatial analysis of elections and committees: Four decades of research," in D. Mueller (ed.) *Perspectives on Public Choice: A Handbook*. Cambridge: Cambridge University Press, pp. 247–270.

Plott, C. R. (1967). "A nation of equilibrium and its possibility under majority rule." *American Economic Review*, 57, 787–806.

Poole, K. and H. Rosenthal. (1996). *Congress: A Political-Economic History of Roll-Call Voting*. New York: Oxford University Press.

Romer, T. and H. Rosenthal. (1978). "Political resource allocation, controlled agendas, and the status quo." *Public Choice*, 33, 27–43.

Rosenthal, H. (1990). "The setter model," in Enelow and Hinich (eds.) *Advances in the Spatial Theory of Voting*. New York: Cambridge University Press, pp. 199–234.

Schofield, N. (1978). "Instability of simple dynamic games," *Review of Economic Studies*, 45, 575–594.

Schofield, N. (1984). "Social equilibrium and cycles on compact sets," *Journal of Economic Theory*, 33, 59–71.

Slutsky, S. (1977). "A voting model for the allocation of public goods: Existence of an equilibrium." *Journal of Economic Theory*, 14, 299–325.

Smithies, A. (1941). "Optimum location in spatial competition." *The Journal of Political Economy*, 49, 423–439.

Chapter 19
Social Choice, Coase Theorem, Contracts, and Logrolling[1]

Peter Bernholz

The problems connected with logrolling (Bernholz, 1974) and vote-trading (Kramer, 1973; McKelvey, 1976; Plott, 1967) are special cases of much wider phenomena (Bernholz, 1981; Schwartz, 1981, 1986). These phenomena are, in fact, the only reasons for the inconsistencies of nondictatorial societies described by Arrow's *General Impossibility Theorem* (Arrow, 1963/51; see Sen, 1987, for a review of Social Choice Theory) "if more than one issues are implied, if individuals have separable preference orderings and if such inconsistencies are not present concerning the alternatives of single issues". Moreover, the respective "paradoxes" of voting and logrolling, usually distinguished in the literature, are identical (Bernholz, 2000). Also, the implied social inconsistencies can occur even if participating individuals have identical preferences but face different restrictions (Breyer, 1980). Finally, the problems put into the center of attention by Arrow would not exist without the presence of negative externalities (in its broad sense, that is, including also political externalities) (Bernholz, 1982). But then, as suggested by the Coase Theorem, generalized below, stable Pareto-optimal outcomes do result in the absence of transaction costs in spite of the validity of Arrow's Theorem, provided that binding contracts are possible. Consequently, the following statement by Sen (1987, p. 383) is true only if contracts are not binding: "... it would appear that there is no way of arriving at a social choice procedure specifying what is to be chosen..., satisfying the appropriately interpreted (i.e., in terms of choice) conditions specified by Arrow...".

Definitions and Assumptions

To consider the general nature of the phenomena, consider a decentralized society, in which $M = \{M_1, M_2, \ldots, M_n\}$ is the *set of issues* among whose *alternatives* humans can select. Each issue comprises at least two alternatives $a_{ik(i)}$,

$(i = 1, 2, \ldots, n; k(i) = 1, 2, \ldots)$. An *outcome* is defined as containing one alternative out of each issue: $a_s = (a_{1k(1)}, a_{2k(2)}, \ldots, a_{nk\ (n)})$. Consequently, $s = 1, 2, \ldots, q$, where $q = |M_1|^* M_2|^* \ldots |^* M_n|$ Further, let $V_i \leq V$ denote the n subsets of society to which the rights to decide the n issues M_i have been assigned. V is the set of all adult people in society. We assume that these m individuals have weak, ordinal, complete, and transitive preferences over all outcomes. In some cases we will assume that individual preference orderings are separable. This means that if an individual prefers alternatives of one or a number of issues to other alternatives of the same issue(s), where the alternatives of all other issues remain constant, then this is also true for different alternatives of the other issues held constant. Formally, consider four different vectors $a_f^h, a_f^{n-h}, a_g^h, a_g^{n-h}$ such that $a_f \equiv (a_f^h, a_f^{n-h})\ a_g \equiv (a_g^h, a_g^{n-h})$, $a_s \equiv (a_g^h, a_f^{n-h})\ a_t \equiv (a_f^h, a_g^{n-h})$. The four vectors contain h and $n - h$ different alternatives, respectively, one out of each issue. Denote by R_j that individual j either prefers the first alternative to the second or is indifferent between them. Assume that $a_f R_j a_s$. Then individual preferences are separable if also $a_t R_j a_g$ holds. Or, similarly, if $a_s R_j a_g$, then $a_f R_j a_t$ is valid.

Return to the decision-making subsets of society. If $V_i > 1$ we call this an *organization*. An organization is supposed to have some consistent decision rule like simple majority voting, voting with the majority of stocks, or unanimous decision making to decide among alternatives. Moreover, let $C_{i1}, C_{i2} \ldots \subseteq V_i$ be the "winning coalitions" of the organization, that is the subset of people who, according to its decision rule, can take decisions for the organization. Only one such coalition exists, namely $C_{i1} = V_i$, if unanimity is required. This is also the case, if V_i contains only one individual. Finally, dictatorship is excluded by assuming that there exists no individual $j \in V$ for which $\{j\} = C_{ij}$ for all i.

These definitions and notations comprise a very broad range of institutional settings. For instance, if $V_i = V$ for all i, and if simple majority voting is used throughout, we may speak of a *Total Direct Democracy*. For, in this case all issues are collectively decided by all citizens by applying simple majority voting. On the other hand, if $n > m$, if each individual j has assigned to him the right to decide at least one issue, and if all issues are only decided by individuals, one may call this *Pure Individualistic Liberalism*. Another form of *Pure Liberalism* would be present, if $V_i < V$, if $|V_i| \geq 1$ for all i, if $V_i \neq V_k$ $(i, k = 1, 2, \ldots, m; i \neq k)$, and if each individual belonged to at least one V_i.

Logrolling and Cyclical Social Preferences

We prove first that a logrolling agreement beneficial to its participants always implies cyclical social preferences. It is assumed that individuals have complete, weak, transitive, and separable individual preference orderings (the separability assumption makes the proof easier, but it can be removed, see below). A logrolling situation is given if all decisions are taken by majority (or qualified majority) voting of the members of a group: if two or more subsets of this group who prefer

intensively certain alternatives of different issues to others would remain in a minority concerning their favored issues, but could form a majority by agreeing to vote for each other's alternatives in an exchange of votes. This exchange of votes for the alternatives of different issues implies, however, that the decision in favor of the unwanted alternatives of other issues is not as important as that relating to the alternatives of the issues favored by the members of the own subset of the group. This means that to assume separable individual preferences is adequate in such a situation. Subsequently only an agreement among two subsets of society will be analyzed, but an extension to the case in which more than two subsets have to join to win a majority is straightforward.

Formally, assume that

$$\left(a_g^h, a_f^{n-h}\right) \equiv a_s R a_f \equiv \left(a_f^h, a_f^{n-h}\right) \tag{19.1}$$

$$\left(a_f^h, a_g^{n-h}\right) \equiv a_t R a_f \equiv \left(a_f^h, a_f^{n-h}\right) \tag{19.2}$$

where R means "the group prefers or is indifferent to". Note that "group indifference" can arise not only because all group members are indifferent among the respective outcomes, but also since no majority can be found for one of the two outcomes. This is especially possible if a qualified majority is required by the decision rule.

Now assume further that

$$\left(a_f^h, a_f^{n-h}\right) \equiv a_f P a_f \equiv \left(a_g^h, a_g^{n-h}\right) \tag{19.3}$$

Then a logrolling situation allowing a successful agreement among a majority is present. For, the last assumption means that a majority (or a qualified majority) of individuals in the group prefers a_f to a_g, $a_f P_m a_g$. But this implies, together with the first two assumptions and the assumption of separable individual preferences, that this majority is composed out of two subsets of the group who each form a minority and for whose members $a_f P_j a_s$, $a_f P_j a_t$, respectively.

The existence of cyclical social preferences can now be derived easily. For, first, from (19.1) and (19.2), respectively, we get by using the assumption of separable individual preference orderings:

$$\left(a_g^h, a_g^{n-h}\right) \equiv a_g R a_t \tag{19.4}$$

$$\left(a_g^h, a_g^{n-h}\right) \equiv a_g R a_s \tag{19.5}$$

And second, it follows from (19.3) together with (19.5) and (19.1), and with (19.4) and (19.2):

$$a_f P a_g R a_s R a_f, \tag{19.6}$$

and

$$a_f P a_g R a_t R a_f. \tag{19.7}$$

It is important to realize that the preferred, but dominated and consequently unstable, outcome a_f can be reached in two different ways by the winning coalition composed out of the two subsets of society. First, with a_g being the status quo, the

coalition can first vote on this outcome as compared to a_s or a_t in favour of one of the latter, and then in favour of a_f compared to these outcomes. This will be called *explicit logrolling* in contrast to *implicit logrolling*. Implicit logrolling takes place if the coalition votes directly in favor of a_f put against the status quo a_g. This presupposes that the changes in the alternatives of the issues at stake can be bundled together by the coalition into one bill. The different problems connected with implicit and explicit logrolling will be taken up in the next section.

Logrolling by Majorities and the Paradox of Voting

Let us turn next to prove the identity of logrolling and the paradox of voting, assuming that two or more issues are present. Consider the graphical representation in Fig. 19.1. Two issues M_1 and M_2 with infinitely many alternatives are assumed. Society comprises only three members as voters. Also let $V_1 \equiv V_2 \equiv V_3 \equiv V \equiv \{1, 2, 3\}$ i.e., all members of society have the right to decide the two issues. As a decision rule, simple majority rule is assumed for both issues. The most preferred outcomes for voters 1, 2 and 3 are x_1, x_2, x_3, respectively. The circles or ellipses around these points combine all outcomes for which the voter in question is indifferent. The further away such a curve is from the most preferred outcome the less the outcomes on it are estimated by the voter.

   This implies that all three voters have single-peaked preferences. But it is well known that this is not a condition sufficient to secure a stable outcome if simple (and in this case even two-thirds) majority voting is used. It is true that if voting
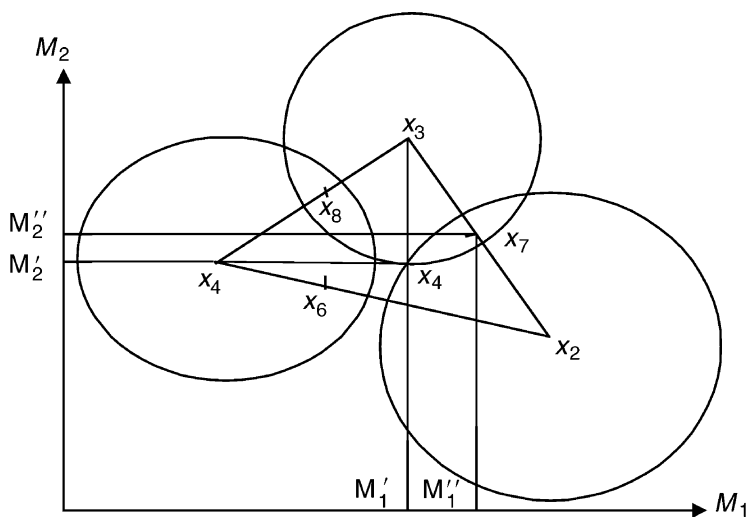
Fig. 19.1 Logrolling and the paradox of voting

would take place independently on the two issues, $x_4 = \{M_1', M_2'\}$ which combines the two most preferred positions of the two different median voters, would be selected. But voters 2 and 3 prefer x_7 to this outcome, so that it will receive a majority if put to a vote against x_4. But the former is itself dominated by x_8, which is preferred by 1 and 3. Similarly, x_8 is dominated by x_6, itself dominated by x_4. Thus a cycle exists and no outcome is stable. It is easy to show that this is just one of infinitely many possible cycles, that the outcome may wander anywhere (McKelvey, 1976), and also, that the conditions for a stable outcome are very restrictive (Plott, 1967; Kramer, 1973).

Note next that by moving from one outcome to another, the respective majority is creating in each case a negative externality to the outvoted voter (minority). Thus, as will be proved in the following section, negative externalities are here as always a precondition for a cycle. Moreover, the majorities in question have to follow an agreement as to their voting behavior. This can be done in two ways: (i) an explicit logrolling contract is concluded, for instance, by 2 and 3 to bring about x_7 instead of x_4: both agree to vote first for M_1'' of the first, and then for M_2'' of the second issue. In this case, voting on the two issues takes place separately; (ii) the two issues are bundled together (e.g., in one bill), so that only one vote takes place between x_4 and x_7, where $x_7 = \{M_1'', M_2''\}$ This is a contract which Gordon Tullock has named implicit logrolling. Explicit logrolling is problematic, since 2 would prefer not to keep his part of the bargain in the second vote and vote for M_2'. But even if x_7 is brought about (as is certainly the case if implicit logrolling is used), it is still unstable. For, other contracts are better for different majorities, so that further logrolling agreements are favorable, which leads to cyclical social preferences. Thus logrolling contracts are a precondition for cyclical preferences. Moreover, it follows from the example, given the case of implicit logrolling, that the paradox of voting for two or more issues is identical to the fact that majority logrolling is possible.

With many voters present, it will usually not be possible for them to conclude logrolling agreements. But in this case parties may exist that propose packages in their election platforms containing alternatives of several issues. If such a package is beneficial to a majority of voters, we have again a case of implicit logrolling, but with the contract now proposed by outsiders, the parties. As a consequence, cyclical social preferences are present again. And this means that any party program proposed by one party can be defeated by that of another party, as already recognized by Downs (1957, p. 55–60).

The General Relationship between Contracting and Cyclical Social Preferences

It has already been stated that logrolling is just a special case of a much more general phenomenon, which has now to be analyzed. Assume that all individuals have separable individual preference orderings, and that no cyclical social

preferences exist concerning single issues. That is, there is no collection of at least three different winning coalitions C_{ir} ($r = 1, 2, 3,...$), such that for outcomes *different only in alternatives of issue i*:

$$a_{i1}Ra_{i2}R...Ra_{ih}Pa_{i,h+2}R...Ra_{i1}$$

is valid. Recall that P means "preferred by society", and R that "society" either "prefers" an outcome to the following one or is "indifferent" between them. Note again that the latter can also occur when no winning coalition exists for one of the two outcomes following each other in the sequence.

Given these assumptions, it follows that cyclical social preferences can occur only if two or more winning coalitions exist, which may be called without loss of generality $C_{11}, C_{21}, C_{31}, ...$, who have the right by using their respective decision rules to decide *different* issues $M_1, M_2, M_3, ...$, and who agree to bring about an outcome preferred by all of their members to the status quo. This means that only by agreeing on such an outcome, that is by concluding a contract, cyclical social preferences arise.

This result holds for the following reasons. First, given separable individual preferences, it follows from the absence of cyclical social preferences in single issues that such cycles can also not occur if only the alternatives of other issues are different. For, since this does not influence the preferences of the members of the winning coalitions concerning their issues because of the assumption of separable individual preferences, this is also true for their decisions and consequently for the "preferences of society". Second, none of the winning coalitions can bring about the new desired outcome on its own, since it does not control the other issues, for which it is not a winning coalition. Otherwise, the outcome would already have resulted before. Consequently, if cyclical social preferences occur, they can only result from contracts referring to more than one issue controlled by the winning coalitions to bring about preferred outcomes.

Let us illustrate this result with the help of an example. Assume three issues, each with two alternatives, and a society comprising three individuals, i.e., $V = \{1, 2, 3\}$. The number of outcomes is, therefore, 8. Moreover, we assume strong individual preference orderings (i.e., individuals are not indifferent among any pair of outcomes), and for individuals 1 and 3 *separable* individual preference orderings concerning issues M_1 and M_3, respectively, and for all individuals concerning issue M_2 (Table 19.1).

The right to decide issues 1 and 3 is assigned to individuals 1 and 3, respectively. All members of the society have the right to decide issue 2 by a simple majority vote. Thus we assume $V_1 = \{1\}$, $V_3 = \{3\}$, $V_2 = V = \{1, 2, 3\}$. The resulting situation is sketched in Fig. 19.2. Here the arrows point from the outcomes preferred by society to those dominated by them. For instance for individual 1, who has the right to decide the first issue, $a_7P_1a_5$, so that "society" also "selects" the former outcome. Concerning the second issue, $a_2P_2a_3$, $a_2P_3a_3$, so that the former is chosen by majority voting. Note also, that all parallel arrows point in the same direction because of the separable preferences of individuals 1 and 3 concerning issues M_1 and M_3, and of all individuals concerning issue M_2, respectively.

Table 19.1 Preference orderings of the members of society

Individual	1	2	3
	a_3	a_2	a_8
Strong individual	a_1	a_1	a_6
Preference	a_4	a_6	a_2
Orderings over	a_7	a_4	a_4
Outcomes	a_5	a_5	a_5
	a_8	a_8	a_1
	a_6	a_7	a_7
	a_2	a_3	a_3
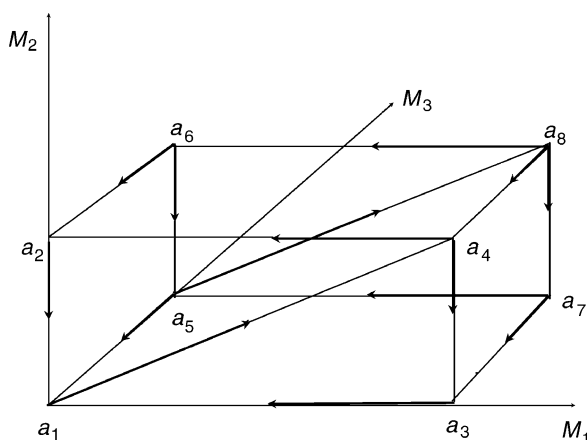
Fig. 19.2 Contracting as causing cyclical social preferences

Now without any contracts among different winning coalitions controlling different issues, no cyclical social preferences are present and a_8 is the resulting stable outcome. But, $a_5 P_1 a_8$, $a_5 P_2 a_8$. Note that 1 has the right to decide the first issue, and that 1 and 2 form a winning coalition for the second issue. As a consequence, they can conclude an agreement to bring about a_5, which is, however, dominated by a_6, a_7, which are dominated by a_8. Thus cyclical social preferences are caused by the possibility to conclude a contract beneficial for its participants.

It remains to prove that cyclical social preferences will always result whenever a contract between winning coalitions controlling different issues benefits all their members. This is easy to prove. Without contracts and no cyclical social preferences among the alternatives of single issues, there can also be no cyclical social preferences comprising alternatives of two or more issues. This follows from the separability of individual preferences, which implies that the decision by those controlling an issue will remain the same whatever decisions others take concerning the other issues they have the right to decide. This implies that in Figure 19.2 all parallel arrows point in the same direction. As a consequence, a stable outcome results. But the conclusion of a contract changes this situation, since it leads to another outcome which is not stable relative to the decisions taken by the winning

coalitions controlling individual issues. Otherwise, it would have been selected before. Thus the decisions taken independently by these winning coalitions lead back to the originally stable outcome, which is itself dominated by that selected through the contract.

The situation is different and more complicated if nonseparable individual preference orderings are present. This is the case for the preference orderings of individual 1 concerning issue M_3 in the above example (Table 19.1), for though $a_1 P_1 a_5$, $a_3 P_1 a_7$, $a_4 P_1 a_8$ we see that $a_6 P_1 a_2$. Similarly, inseparability of preference ordering concerning M_1 holds for individual 3. These inseparable preferences were unimportant for the relationships sketched in Fig. 19.2 since 1 and 3 only had the right to decide M_1, M_3, respectively. But this is no longer the case if a different assignment of rights is considered. As an example, and to bring logrolling into the discussion, assume that $V_1 = V_2 = V_3 = V$ and that all issues are decided by majority voting. Then the situation of Fig. 19.3 emerges for society.

As can be seen, there exist now cyclical social preferences, namely $a_2 P a_4 P a_8 P a_6 P a_2$, without any contracts and this means, in this case, without a logrolling agreement being concluded among different winning coalitions controlling different issues. This is a consequence of the fact that the nonseparable preferences of individual 3 are now decisive concerning M_2.

But note also that if a contract, in this case a logrolling agreement, is concluded between 1 and 2 to bring about $a_1 P a_4$, since they both have corresponding preferences and form together a majority, then an additional social preference cycle is created.

Consider next a third example, in which the preference ordering of individual 1 is changed concerning one pair of outcomes only, namely from $a_6 P_1 a_2$ to $a_2 P_1 a_6$. Moreover, whereas it is assumed as before that $V_1 = V_2 = V = \{1, 2, 3\}$, with majority voting as a decision rule, the right to decide M_3 is now assigned to individual 1, $V_3 = \{1\}$. Note that inseparable individual preference orderings are still present. In this example we get the situation of Fig. 19.4.
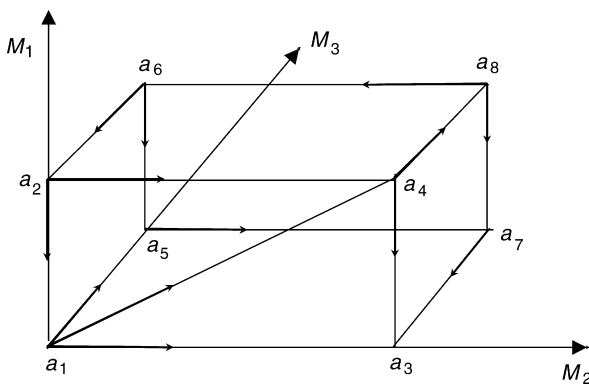
Fig. 19.3 Non-separable individual preferences and cyclical social preferences
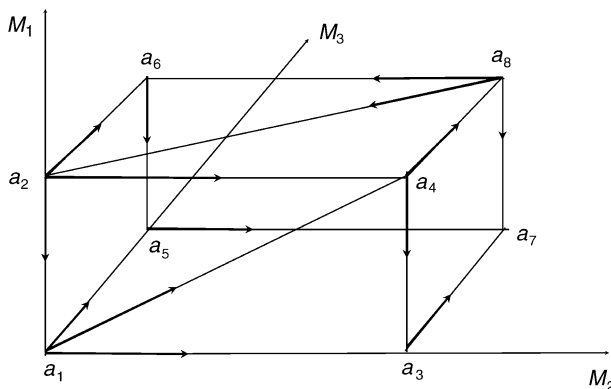
Fig. 19.4 Non-separable individual preferences and cyclical social preferences: Case 2

It follows that no cyclical social preferences are present if no contracts are concluded and that a_2 is the only stable outcome. Now let us check whether contracting leads again to cyclical social preferences. This would be not the case if 1 and 2 would conclude a contract implying a_1Pa_4, which they could do since they enjoy a majority for both issues M_1 and M_2. But such a contract would never be agreed on by 2, since the present stable outcome a_2 is preferred by it to a_1. This is different for a contract between 1 and 3 implying a_8Pa_2, since they both prefer the former outcome to the latter. Moreover, 1 has the right to decide M_3 and they both form a majority for a decision on M_2. And this contract leads again to cyclical social preferences. We conclude from this example that contracts bring about cyclical social preferences only if another outcome is "socially preferred" to an outcome that directly or indirectly dominates all other outcomes if no contracts are concluded. However, returning to the second example, this conclusion has to be changed somewhat. For, in this example outcomes a_2, a_4, a_6, a_8 are members of a cycle, but dominate directly or indirectly all other outcomes if no contracts take place. And again, since a_1Pa_4, one of this set of dominant outcomes is dominated itself if a contract is concluded. And as a consequence, a cyclical social preference cycle results.

The results derived with the help of the examples allow the formulation of the following theorem:

Theorem 1. Assume weak, ordinal, complete and transitive individual preferences and a society in which there exists a decentralized assignment of the rights to decide issues, together with appropriate decision rules for organizations. Then, if no cyclical social preferences are present without contracts being concluded on decision making concerning different issues, then each conclusion of such contracts leads to cyclical social preferences. Moreover, if a subset of outcomes exists for which cyclical social preferences are present if no contracts are concluded, and which dominate directly and indirectly all other outcomes, then additional cycles result if contracts are concluded by which one of the latter outcomes dominates one of the former (Schwartz, 1981, 1986).

Note that the special case of separable individual preferences is covered by Theorem 1. For, in this case one stable dominating outcome exists without contracts, if no cyclical social preferences exist among outcomes different only in single issues. It then follows from the theorem that the conclusion of contracts is the only reason for the emergence of cyclical social preferences.

Negative Externalities and Cyclical Social Preferences

In this section we prove the following theorem:

Theorem 2. Externalities are a necessary condition for the existence of cyclical social preferences.

Define first negative externalities: Denote by C'_{ij} a winning coalition that can decide among outcomes a_j, a_j. Note that this can either be a coalition which has the right according to the assignment of rights and the prevailing decision rule to decide among alternatives of one issue, so that the two outcomes are only different in this issue; or that the winning coalition could come about by a contract among such coalitions who agree on their preferences for the two outcomes that differ in more than one issue.

Definition of negative externalities: Assume that a C'_{ij} exists for all of whose members $a_i P_h a_j$, $(\forall h \in C'_{ij})$. In this case we get $a_j P a_j$. Thus this winning coalition can bring about the former outcome whenever the latter is present. Then negative externalities exist for other members of society not belonging to the coalition if $a_j P_h a_i (h \in V - C'_{ij})$.

This is a very broad definition: Negative externalities include not only the externalities usually considered in economics, but also the negative consequences for the individuals who have the right to participate in a decision, but who are "outvoted" according to the decision rules prevailing in the organization(s). This latter phenomenon has been mentioned earlier by Buchanan (1962) and Buchanan and Tullock (1962) as stemming from the fact that collective decision making not requiring unanimity may lead to negative externalities.

Let us now prove Theorem 2: We take into account that individuals may be indifferent between two outcomes, and denote by $a_j R_h a_j$ that individual h is either indifferent between these two outcomes or prefers the former to the latter outcome. Consider outcomes $a_0, a_1, \ldots, a_s$. Assume that cyclical social preferences exist:

$$a_s P a_{s-1} P \ldots P a_1 P a_0 P a_s \text{ with } 3 \leq s \leq q. \tag{19.8}$$

Then there exist winning coalitions $C'_{s,s-1}$, $C'_{s-1,s-2}, \ldots, C'_{10}$, C'_{0s} for whom

$$a_0 P_h a_s (\forall h \in C'_{0s}) \tag{19.9}$$

and

$$a_i P_h a_{i-1}, (h \in C'_{i,i-1})(i = 1, 2, \ldots, s) \tag{19.10}$$

is valid.

Equation (19.8) implies, because of the transitivity of individual preferences:

$$(\bigcap_{i \neq}^{s} C'_{i,i-1}) \cap C'_{0s} = \emptyset \tag{19.11}$$

which implies the absence of dictatorship.

We assume now that no negative externalities exist for the members of society not belonging to the above winning coalitions, and show that this leads to a contradiction. It follows from the absence of negative externalities that

$$a_0 R_h a_s, (\forall h \in V - C'_{0s}) \tag{19.12}$$

and

$$a_i R_h a_{i-1}, (\forall h \in V - C'_{i,i-1})(i = 1, 2,\ldots,s). \tag{19.13}$$

From Equations (19.9) and (19.12) and from (19.10) and (19.13), respectively,

$$a_0 R_h a_s, (\forall h \in V) \tag{19.14}$$

$$a_i R_h a_{i-1}, (\forall h \in V)(i = 1, 2,\ldots,s) \tag{19.15}$$

Now consider any $C'_{i,i-1}$ ($i = 1, 2,\ldots, s$) or C'_{0s} for all of whose members (19.10) or (19.9) is valid. Since $C'_{0s}, C'_{i,i-1} \subset V$, (19.14) and (19.15) are true for all their members. But then one can derive from (19.9) and (19.15) or from (19.10), (19.14) and (19.15), respectively:

$$a_s R_h a_{s-1} R_h \ldots R_h a_1 R_h a_0 P_h a_s, (\forall h \in C'_{0s}), \tag{19.16}$$

$$a_s R_h a_{s-1} R_h \ldots R_h a_i P_h a_{i-1} R_h \ldots R_h a_0 R_h a_s, (\forall h \in C'_{i,i-1})(i = 1, 2, \ldots ,s)\tag{19.17}$$

This result, however, contradicts the assumption of transitive individual preferences. Thus, cyclical social preferences can only exist if negative externalities are present.

Theorem 2 has far-reaching consequences. It suggests that the assignment of rights is of decisive importance for the occurrence of cyclical social preferences. Even more, according to the *Coase Theorem* the original assignment of rights is not important if no transaction costs for concluding contracts are present, since then all negative externalities can be removed by mutually beneficial agreements. Before returning to this problem in a later section , however, let us take up the influence of the kind of assignment of rights.

Cyclical Social Preferences and Assignment of Rights

Denote by a *profile of individual preference orderings* an m-tuple of individual preference orderings, one for each of the m members of society. Obviously there exist a huge number of such profiles, since people can have many different preference orderings. Subsequently we illustrate with the help of examples (Bernholz, 1986).

Theorem 3. For each possible profile of individual preference orderings, there exists a non-oligarchic assignment of rights to decide among all pairs of outcomes, such that no intransitive social preferences exist, and that a stable Pareto-optimal outcome results.

Note that rights are referred to in the theorem as rights to decide among pairs of outcomes, instead of among alternatives of issues. This is, however, not important. For, if we add to the right to decide issues the right to conclude contracts, both rights together imply a right to decide among certain pairs of outcomes for the winning coalitions agreeing on the respective contract.

Consider first the examples of Table 19.1 and of Figs. 19.2 and 19.3. In these cases the same individual preference orderings lead through two different assignments of rights and the conclusion of contracts to cyclical social preferences. According to Theorem 3, there should, however, given the profile of individual preferences, exist an assignment of rights preventing cyclical social preferences. And this is indeed the case for the following assignment: $V_1 = \{1\}$ $V_2 = \{3\}$ $V_3 = \{2\}$. The resulting situation is presented in Fig. 19.5, in which no cycles exist and in which a_1 is the stable Pareto-optimal outcome. The diagonal from a_1 to a_4 shows no arrow, since the preferences of the potential partners are opposed, so that no contract is concluded. For all other diagonals the same is true, or the potential contracts do not imply cyclical social preferences. Moreover, they would not be concluded since they would not lead for the participants of such contracts to a better or another outcome than a_1.

Moreover, when analyzing whether there exists an assignment of rights preventing cyclical social preferences in the example of Fig. 19.4, we find that the assignment just given, and thus the situation depicted in Fig. 19.5, is again a solution. This is not surprising, for the preferences of the three members of society are the same as in Table 19.1, except that for individual 1 $a_2 P_1 a_6$ instead of the opposite preference among these outcomes. But 1 has no right to participate in the decision concerning issue M_3, so that the situation of Fig. 19.2 remains unchanged.
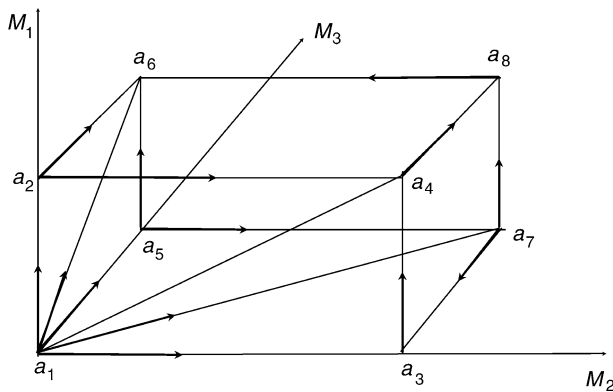
Fig. 19.5 Reassignment of rights to decide issues prevents cyclical social preferences

It is important to point out that within the framework of total direct democracy present in the logrolling case no solution can be found for removing cyclical social preferences, except with unanimity as a voting rule, whereas an individualistic liberal reassignment of rights offers a solution. This is not the accidental result of our specific example, since it has been shown (Bernholz, 1986, p. 256f.) that:

Theorem 4. For any possible profile of individual preferences there exists a purely liberal assignment of the rights to decide among all pairs of outcomes, such that no intransitive or cyclical social preferences exist and that any outcome to which no other outcome is preferred is Pareto optimal. But the same result does not hold if we assign the rights to decide among all pairs of outcomes to all members of society (Total Direct Democracy), deciding with simple or qualified majorities.

Since the general approach presented here does include pure individualistic liberalism, it also covers Sen's (1970) well-known example concerning the conflict between mother and daughter whether they should read *Lady Chatterley's Lover*. Decision rights for the two issues are only assigned to individuals, namely mother and daughter. It will be shown that in this example, too, negative externalities are present, that the conclusion of a contract leads to cyclical group preferences, and that they can be removed by a different assignment of rights. Following Sen, let us assume $V_1 = \{1\}$ and $V_2 = \{2\}$, $V = \{1, 2\}$ where 1 denotes the mother and 2 the daughter. The two issues M_1 and M_2 refer to reading the book by mother and daughter, respectively. The two alternatives of each issue are "reading" and "not reading" by mother and daughter. Table 19.2 describes the strong preference orderings of mother and daughter over the four outcomes. As can be seen, the decisions by the mother lead to negative externalities for the daughter, and vice versa.

The resulting situation is sketched in Fig. 19.6. The stable outcome without contract, a_3, is Pareto-inferior. This is Sen's celebrated theorem of the *Impossibility of a Paretian Liberal*. But if we allow a contract between mother and daughter, such an outcome will be agreed on, since both prefer a_2 to a_3. But the contract implies a social preference cycle, because the mother prefers a_1 (both do not read), and the daughter a_4 (both read the book) (Theorem 1). But a stable outcome is reached without a contract if the right to decide M_1 is assigned to the daughter and that to decide M_2 to the mother (Theorems 3 and 4).

It has been pointed out correctly that the results presented by Theorems 3 and 4 may be interesting logically, but that their relevance for reality is rather limited, since an assignment preventing cyclical social preferences would have to be based

Table 19.2 Preferences for reading lady Chatterley's lover

	Members of society	
	Mother: 1	Daughter: 2
Strong	a_1	a_4
Separable	a_2	a_2
Individual	a_3	a_3
Preference	a_4	a_1
Orderings		
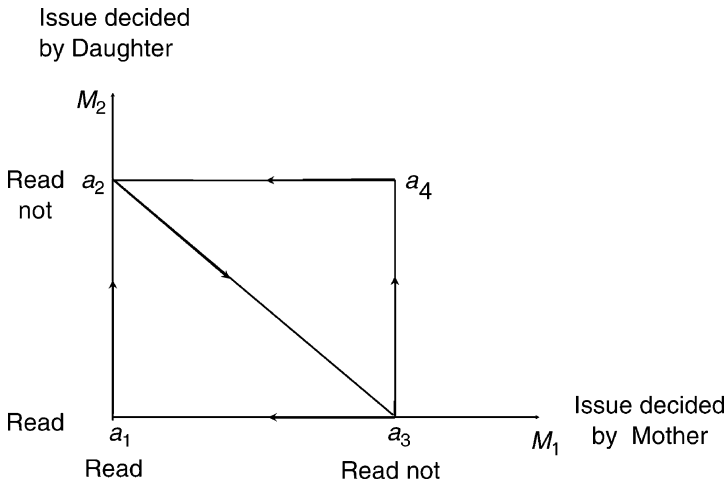
Fig. 19.6 Reading Lady Chatterley's Lover

on a knowledge of all individual preference orderings in society. As a consequence, one has to ask whether the same or a similar result cannot be reached by individual decisions, given some original assignment of rights. This is, indeed, the case, as will be shown in the next section.

Cyclical Social Preferences, Negative Externalities, and the Coase Theorem

Negative externalities are a precondition for the existence of cyclical social preferences. But then, the possibility to conclude contracts should, according to the Coase Theorem (Coase, 1960), lead to stable Pareto optimal outcomes, whatever the original assignment of rights, provided that there are no transaction costs. And social choice theory, as well as the theory of logrolling, does not assume the presence of such costs. It seems to follow that the Coase Theorem contradicts Theorem 1, which asserts that contracts are a major reason of cyclical social preferences.

But this seeming contradiction can be removed, if *contracts are binding*, and if the Coase Theorem is generalized. Subsequently, a difference will be made between *internal* and *external contracts*. An external contract is a contract agreed on by organizations or individuals with other organizations or individuals. An internal contract, by contrast, is a contract concluded by the members of a winning coalition who have the right to decide an issue for an organization according to its decision rule, to bring about a certain alternative. Note that this may but need not be done by agreeing to an external contract.

Let us now illustrate with several examples that the generalized Coase Theorem holds if internal as well external contracts are binding. Consider first the example of

Table 19.2 and Fig. 19.6. In this case only *external contracts* can be concluded, since no organization is present. The contract to bring about a_2 is advantageous to both mother and daughter. If it is also binding, then this Pareto-optimal outcome is also stable. For outcomes a_1, a_4, cyclical social preferences could only come about if mother or daughter, respectively, could break the contract. This result also demonstrates that Sen's theorem of the *Impossibility of a Paretian Liberal*, though true, is based on too narrow a conception of liberalism. For, if the possibility of binding contracts is included, the supposed paradox disappears.

Consider next the logrolling example of Table 19.1 and Fig. 19.3. In this case, only *internal contracts* can be agreed on by winning coalitions, since $V_1 = V_2 = V_3 = V$. 1 and 3, for both of whom $a_8 P_i a_2$, are minorities for votes on some alternatives of issues M_2, M_3, if they are voted on separately. They thus have to conclude a contract to bring about a winning logrolling coalition for outcome a_8. But since 1 prefers a_4, it is motivated to break this *internal* contract. This is, however, prevented if the contract is binding, and no cyclical social preferences result. Note also, that the resulting outcome is Pareto optimal. The same analysis is valid for the example of Fig. 19.4.

Things are a little bit more complicated for the example of Table 19.1 and Fig. 19.2. In this case only $V_2 = V$, whereas $V_1 = 1$. As a consequence, an external and an internal contract are involved in bringing about a_1 instead of a_4: An external contract concluded between V_1 and V_2, and an internal contract between 1 and 2 to vote for the respective alternative of M_2, though for 2 $a_2 P_2 a_1$. But a_1 is not stable, even if both contracts are binding. For, 3 has the right to decide M_3 and prefers a_5. Obviously, Pareto-optimal outcome a_4 is a candidate for an external contract by all V_i and an internal contract by 1, 2, and 3 to vote for the respective alternative of issue M_2. Note, that in this case all participants to the contract have an incentive to break it. 3 prefers a_8 and has the right to bring it about. 1 and 2 would prefer to move to a_1 and could do it by the contract described above. This would imply that 1 would break the external contract for V_1, and that both would break the internal contract on how to vote on issue M_2, which would also violate the external contract concluded for V_2. But again, if both external and internal contracts are binding, stable Pareto-optimal outcome a_4 comes about.

The results just discussed suggest (Bernholz, 1997, 1999):

Theorem 5. Assume that weak, ordinal, complete, and transitive individual preference orderings are present, that at least one finely divisible good exists, that any decentralized original assignment of rights to subsets of society is given to decide issues, and that no cyclical social preferences are present without contracts. Then if binding external and internal contracts can be concluded, a stable Pareto-optimal outcome results from the decisions of the members of society.

Conclusions

In the present essay it has been shown that the problems laid open by Social Choice Theory, such as Arrow's theorem of the impossibility of a nondictatorial social preference ordering or Sen's impossibility of a Paretian liberal, are a consequence

of not allowing binding contracts between individuals and organizations. With the generalized Coase Theorem applied above, Pareto-optimal solutions are present for all these cases provided that there are no transaction costs and that binding contracts are possible. But binding and enforceable contracts are a prominent characteristic of developed economies.

Unfortunately, these results have until now not been taken up by mainstream Social Choice Theory (Arrow, 2002; Gaertner 2007). The only exception is provided by Dennis Mueller (2003), who discusses these relationships extensively. He shows that the original Coase Theorem may not hold for three or more potential participants in a contract, if the core is empty. But, in following our argument, he demonstrates that the introduction of binding external and internal contracts helps solve this problem.

The conclusions drawn in the above essay underline (i) the importance of a clear and complete assignment of rights to decide issues, which is also the case of property rights. They show (ii) the importance of the freedom to conclude binding contracts, and stress (iii) the need for a legal system to prevent any violation of contracts by sufficiently strong sanctions. The results, therefore, cast doubt on the social opprobrium against concluding contracts concerning political decisions, and the absence of fines and penalties if such contracts are broken. Even the prohibition to exchange or sell votes seems to be problematic. The possibility to enforce "political" contracts under certain conditions within the framework of repeated games may not be sufficient to overcome the implied disadvantages in many empirical cases. Finally, the meta-rights of some organizations like parliament, government, or bureaucracy to reassign decision rights more or less at their discretion, may have severe negative consequences for society. On the other hand, new issues and alternatives are coming up all the time, so that some organization is needed to assign them. Connected to this may be a necessity to reassign some rights concerning the old issues. But much work remains to be done to clarify these problems.

Moreover, great caution is advisable concerning the interpretation of the conclusions. First, no strategically motivated decisions to bring about a preferred outcome were taken into account. Second, the absence of transaction costs has been assumed. This is a severe restriction, since binding contracts implying up to n issues may be needed to bring about the stable Pareto-optimal outcomes. And transaction costs will generally increase with the number of issues and people involved. As a consequence, many binding contracts may not be concluded because of such costs. Note also that the presence of transaction costs may be a sufficient condition for the evolution of organizations in history. For, we have shown that up to n-issue contracts may be necessary to get stable Pareto-optimal outcomes. But if transaction costs are no longer neglected, it is clear that they generally increase with the number of issues and of people involved. It follows that this is already a sufficient condition for the evolution of organizations in history, which control one or more issues, so that the number of people participating in negotiating contracts and thus contract costs are reduced. From this conclusion it follows that it is not possible to agree with Williamson's statement (1981, p. 1545) that "but for

the *simultaneous* existence of *bounded rationality and opportunism*, all economic contracting problems are trivial and the study of economic institutions is unimportant. Thus, but for bounded rationality, all economic exchange could be effectively organized by contract."

References

Arrow, K.J. (1963/51). *Social Choice and Individual Values*. New York: Wiley.

Arrow, K.J. (ed.) (2002). *Handbook of Social Choice and Welfare*. Amsterdam: Elsevier.

Bernholz, P. (1974). "Logrolling, arrow-paradox and decision rules: A generalisation." *Kyklos* 27, 49–72.

Bernholz, P. (1981). "A General Social Dilemma: Profitable Exchange and Intransitive Group Preferences." *Zeitschrift fuer Nationaloekonomie* 40, 1–23; reprinted in J.M. Buchanan and R.D. Tollison (eds.) (1984) *Public Choice II*, 361–381. Ann Arbor: University of Michigan Press.

Bernholz, P. (1982). "Externalities as a necessary condition for cyclical social preferences." *Quarterly Journal of Economics* 47(4), 699–705.

Bernholz, P. (1986). "A general constitutional possibility theorem." *Public Choice* 51, 249–265.

Bernholz, P. (1997). "Property rights, contracts, cyclical social preferences and the coase theorem: a synthesis." *European Journal of Political Economy* 13(3), 419–464.

Bernholz, P. (1999). "The generalized coase theorem and separable individual preferences: A comment." *European Journal of Political Economy* 15, 331–335.

Bernholz, P. (2000). "Instability of voting outcomes, logrolling, Arrow, Coase, and all that: A different interpretation," in P. Fishback et al. (eds.) *Public Choice Essays in Honour of a Maverick Scholar: Gordon Tullock*. Boston: Kluwer, pp. 83–97.

Breyer, F. (1980). "On the relevance of preference similarities for the paradox of voting—identical utility functions and cyclical majorities." *Kyklos* 33, 523–530.

Buchanan, J.M. (1962). "Politics, policy and the pigovian margins." *Economica N.S.* 29, 17–28.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Coase, R.H. (1960). "The problem of social cost." *Journal of Law and Economics* 3, 1–44.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Gaertner, Wulf (2007). *A Primer in Social Choice Theory*. Oxford: Oxford University Press.

Kramer, G. (1973). "On a class of equilibrium conditions for majority rule." *Econometrica* 41, 285–297.

McKelvey, R.D. (1976). "Intransitivities in multidimensional voting models and some implications for agenda control." *Journal of Economic Theory* 12, 472–482.

Mueller, D. (2003). *Public Choice III*. Cambridge: Cambridge University Press.

Plott, C.R. (1967). "A notion of equilibrium and its possibility under majority rule." *American Economic Review*, 57, 787–806.

Schwartz, T. (1981). "The universal instability theorem." *Public Choice* 37, 487–501.

Schwartz, T. (1986). *The Logic of Collective Choice*. New York: Columbia University Press.

Sen, A. (1970). "The impossibility of a paretian liberal." *Journal of Political Economy* 78, 152–157.

Sen, A. (1987). "Social choice," in J. Eatwell, et al. (eds.) *The New Palgrave. A Dictionary of Economics*. Volume 4, London: Macmillan, pp. 382–393.

Williamson, O.E. (1981). "The modern corporation: Origins, evolution, attributes." *Journal of Economic Literature* 19, 1537–1568.

Chapter 20
Interest Group Behavior and Influence[1]

Frans A.A.M. van Winden

Introduction

During the last two decades economics has witnessed a remarkable upsurge in theoretical as well as empirical studies of the behavior and of the political influence of interest groups. Recent books by Sloof (1998), Drazen (2000), Persson and Tabellini (2000), and Grossman and Helpman (2001) refer to a wealth of evidence of the significance of organized interests in the political arena, besides presenting surveys of theoretical studies. Political economics definitively seems to move away from the common assumption of atomistic demand in 'political markets' (the median voter model) towards a more realistic framework. In a sense it is picking up and deepening some older strands of literature inspired by classical writers on political economy (like Marx and Pareto), the so-called pluralists in political science (like Bentley and Truman), and others, who were concerned with the political impact of particular social groups under the label of 'factions', 'classes', or 'elites' (see e.g., Bottomore 1970; Moe 1980). The modern political economic literature to be surveyed in this paper, however, is characterized by much greater rigor, through the use mathematical modeling, and keener attention for individual incentives. Strict adherence to methodological individualism would require the modeling of the following chain of events regarding the interaction between policymakers and interest groups: group formation/adjustment $\rightarrow$ group decision making $\rightarrow$ group activity $\rightarrow$ political decision making $\rightarrow$ government policies (plus other relevant events) $\rightarrow$ group formation/adjustment. Due to the complexity involved, group formation and adjustment (influenced by policy outcomes) are typically neglected by taking the existence of interest groups as given, thereby sidestepping the thorny issue of individual incentives for participation in collective action (Olson 1965). In addition, interest groups are commonly assumed to act as single (unitary) actors. Nevertheless, our conclusion will be that there has been

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 118–129.

substantial theoretical progress, opening up many promising paths for important and exciting research.

In this paper we will focus on formal theoretical models of interest group behavior and influence, with emphasis on the positive aspects.[2] Early modeling of interest groups, during the 1970s and the beginning of the 1980s, had difficulty in dealing simultaneously with the behavior of interest groups and policymakers. In response short cuts were taken in the form of higher levels of abstraction or by focusing on one side of the interaction between the agents. The former short cut is used in the cooperative game and compromise function models described in section 2, the latter by the so-called influence and vote function models discussed in section 3. In the wake of the rise of non-cooperative game theory in the 1980s the modeling of interest group behavior became much more general and sophisticated. Two strands of literature will be highlighted. Section 4 discusses common agency models of contributions offered to policymakers in exchange for policies or to help finance electoral campaigns, while section 5 deals with models of strategic information transmission. Section 6 is concerned with extended models investigating the multiple means and channels of influence that are, in general. available to groups. Section 7 concludes.

Cooperative Games and Compromise Functions

Characteristic of cooperative game models is the focus on coalitions rather than individual agents, and outcomes (reasonable compromises) rather than strategic moves. Although less explicit, it avoids problems of arbitrariness in the specification of moves. By requiring collective rationality, policy outcomes of these models are (constrained) efficient, that is, they are in accordance with the maximization of a weighted representation of the utilities of the players involved. More formally, suppose that n interest groups can be meaningfully distinguished for the policy x, and that $v^i(x)$ represents the related net benefits or utility of group i ($I = 1, \ldots, n$). Let μ denote the 'political influence weight' of the group. Then, the behavioral assumptions underlying the models imply that x follows from the maximization of the function $P(x) = \sum_i \mu^i v^i(x)$. Although this function looks like a social welfare function, it should not be labeled as such because the influence weights are based on a positive instead of a normative (ethical) analysis. We will therefore call it a *political welfare function*. Two types of models will be discussed: the *power to tax* model (Aumann and Kurz 1977) and the *interest function approach*

(van Winden 1983). These models differ in the assumptions underlying the function $P(x)$ and the nature of the influence weights.

The *power to tax* model concerns a redistribution game where the so-called Harsanyi-Shapley-Nash value is used as a solution concept. The income distribution is determined by majority voting. Players in the game are all n individuals in society (making up the set N), who are endowed with a pre-tax income y^i. Redistribution is constrained by the requirement that total after-tax income ($\sum_i x^i$) equals total pre-tax income ($\sum_i y^i$). Groups enter the picture because a majority coalition is required for redistribution. Any majority coalition, C, can redistribute all income from those outside the coalition, $N\backslash C$, to itself. The crucial point is that the outside coalition $N\backslash C$ can threaten to destroy its own pre-tax income, leaving nothing to be redistributed to C. The outcome of this game is determined by using the Nash Bargaining Solution (which assumes that players can make binding agreements, committing themselves to carry out threats if no agreement is reached). Proceeding in this way for all possible coalitions, an individual's 'power' (Shapley value) can be derived from the individual's (expected) contribution to all possible coalitions. They show that this power over the *resulting* income distribution ($x = x^1, \ldots, x^n$) corresponds with $\mu^i = 1/v_x^i$, that is, an individual's influence weight equals the reciprocal of her or his *ex post* marginal utility v_x^i. Since commitments or possible threats, are never carried out, because they are anticipated by the players, preventing inefficient outcomes. Furthermore, no coalitions (interest groups) actually form. Thus, one could say that x results from the anticipation of pressure activities that could but do not actually occur. The model has been extended in several directions. For example, Aumann, Kurz and Neyman (1983) apply a similar analysis to public good, Gardner (1981) introduces a government as player, Osborne (1984) studies the differential taxation of goods that can (like labor-time via strikes) or cannot (like land) be 'destroyed', while Peck (1986) takes incentive effects of taxation into account.[3]

The *interest function approach* takes a less abstract perspective on policymaking. It is argued that in capitalist economies, analytically, four basic social groups can be distinguished, based on their position with respect to production in the economy: capitalists, private sector workers, public sector workers (politicians and bureaucrats), and dependants (unemployed, disabled, retired). The political interests of a group are represented by an 'interest function' $v^i(x, y)$. The value of x is determined by the public sector workers, while $y = (y^1, y^2)$ stands for the actions taken by the capitalists and private sector workers, respectively. The two latter groups play a non-cooperative game, where each group takes the actions of the government and the other group as given. This determines their actions as a function of x: $y = (y^1, y^2) = y(x)$. The crucial assumption is that public sector workers, when deciding on x, will to some extent, take into account the interests of the other groups. The extent to which they will do so is related to the potential

---

[3] Dougan and Snyder (1996) present another cooperative game model of income redistribution. Zusman (1976) deals with consumer and producer groups in a regulated market.

influence of 'ideology' (including altruism), multiple positions (simultaneous membership of different groups), mobility (probability of becoming a member of a different group), and pressure (influence attempts by private sector groups).[4] The resulting policy x is assumed to have the character of a compromise (a generalized Nash Bargaining Solution), equivalent to the maximization of the 'complex interest function' $P(x)$ above, where the influence weights are determined by the aforementioned factors. However, no explicit behavioral model is provided for the relationship between these weights and the proposed determinants of pressure (threat potential, group cohesion, and an information factor). Later models, discussed below, do provide such a microfoundation.[5] The approach has been theoretically as well as empirically applied in several ways. For example, dynamic models including elections—showing politically induced economic cycles of various lengths—are analyzed by van Winden (1983) and van Winden, Schram, and Groot (1987). Borooah and van der Ploeg (1983) and van Velthoven (1989) study macroeconomic models with endogenous government behavior (see also Przeworski and Wallerstein 1988). van Velthoven and van Winden (1985) and Verbon (1989) focus on social security. Renaud (1989) presents (empirical) analyses of fiscal federalism and public sector growth. Mazza and van Winden (1996) study the impact of labor migration, and Drissen (1999) analyzes a computable general equilibrium model with redistribution and public production. Also, with some empirical support (Renaud and van Winden 1988; van Velthoven 1989) the relative numerical strengths of these groups have been used to study with a theoretical model the dynamics of endogenous influence weights (van Velthoven and van Winden, 1985).

Another strand of literature, with roots in Stigler's (1971) theory of regulation and its formalization by Pelzman (1976), simply postulates a compromise function to endogenize policy, using as arguments typically the weighted surpluses of consumers and producers. Maximization by the policymaker is usually (implicitly) justified by the presumed goal of maximal electoral support. However, as noted by

---

[4] Mobility can be an important reason why the interests of dependants are taken into account (see Renaud and van Winden, 1988). Another reason why social groups may count is 'structural coercion', that is, the systematic reactions by private sector agents to government policies when these are taken as given. In that case policymakers may be induced to sort these agents into groups (which need not be organized). In this way interest groups play a role in *probabilistic voting models*, for example (see Coughlin et al., 1990; Coughlin, 1992; Hettich and Winer, 1999). To illustrate, suppose that an incumbent party has to choose its platform x in a forthcoming election. The electorate comprises N groups, where each member of a group (say, group i, with n_i members) derives the same utility $v_i(x)$ from the party's policy. In addition, member j has a personal utility 'bias' b_{ij} in favor of (>0) or against (<0) this party, where b_{ij} is uniformly distributed over the interval (l_i, r_i). Let utility from the challenging party be zero, then voter ij votes for the incumbent if $v_i(x) + b_{ij} > 0$. Interestingly, maximization of expected plurality (assuming $l_i < v_i(x) < r_i$) implies that $P(x)$ above is maximized, with $\mu_i = n_i/(r_i - l_i)$. Thus, numerical strength and group homogeneity determine the influence weights.

[5] See the common agency model in section 4. Further support is provided by the (pressure) model of strategic information transmission of Potters and van Winden (1990). See also the probabilistic voting model discussed in the previous note.

Hirshleifer (1976), policymakers (regulators) themselves constitute an interest group with an interest in wealth, implying that political support can only be an instrumental and partial aim.

Influence and Vote Functions

Policies can be affected by interest groups in two ways: directly, by influencing the behavior of policymakers, and indirectly, by influencing the behavior of voters. The influence function and vote function models discussed next are concerned with these two channels of influence. Characteristic is the focus on interest group behavior, whereas the impact on policymaking or voting behavior is simply assumed. Furthermore, while the precise nature of the activity is often left obscure in the first type of models, campaign contributions are focused on in the latter.

Influence Functions

Political decision making is often modeled as a kind of all-pay-auction. Policy-makers offer certain policies (public goods, transfers, regulation), while demand comes from interest groups. The 'price' the latter have to pay is determined by the resources spent on the acquisition of the goods. Let x represent the policies, y^i the resources spent by interest group i, and $v^i(x, y^i)$ its net benefits. Many studies assume a fixed positive relationship between policies and resources spent, an *influence function*: $x = I(y; z)$, where both y and z are vectors and z represents exogenous variables (like group sizes). Examples are Findlay and Wellisz (1983), Cairns (1989), and Coggins, Graham-Tomasi, and Roe (1991). In one part of the literature, based on the pressure model of Becker (1983), x represents the amount of a transfer or public good. In the rent seeking literature, originating with Tullock (1967, 1980), x usually denotes the probability that a particular good (a monopoly license, for instance) is obtained. The equilibrium level of the resources expended by the groups is determined under the assumption of non-cooperative (Cournot-Nash) behavior. In both literatures, the resources spent by the interest groups typically entail a pure social cost, that is, their activity has no productive aspect. Competition has a better side in Becker's model, where efficiency costs of the policies (transfers) as such are taken into account. Under some reasonable assumptions, an increase in the efficiency cost of taxes (subsidies) induces an increase (decrease) in the resources spent by the taxed (subsidized) group, leading to a fall in the tax and subsidy level. Another interesting result follows if an increase in group size induces free riding. If the negative free riding effect is sufficiently strong, this will lead to fewer resources being spent. The implication is that "groups can more readily obtain subsidies when they are small relative to the number of taxpayers"

(Becker, 1983, p. 395).[6] This second result qualifies the importance of sheer numbers in politics. However, this result only bites if influence via elections (votes) is dominated by interest group pressure.[7] If not, larger groups can be expected to focus relatively more on pressuring politicians interested in votes than on bureaucrats. Also, larger groups will be relatively more inclined to produce pressure in the pursuit of group-specific public good (like a trade tariff), because of the fewer spoils to the individual member in case of private good (like transfers).[8]

An important issue that rent-seeking models are concerned with is the extent to which the benefits of the rent (x) are dissipated in the competition among groups to obtain the rent. Other issues explored are the effects of: risk attitude, nature of the rent (private or public good), groups versus individuals as players, intergroup mobility, multiple rents (prizes), endogeneity of the order of moves, asymmetry of information (e.g., regarding valuation or capabilities), budget constraints, and sharing of rules (for surveys, see Nitzan 1994; Tollison 1997).[9]

Although competition among interest groups may be less detrimental to efficiency than what the rent seeking literature suggests, in the Becker model "all groups could be made better off by reduced expenditures" (Becker 1983, p. 387), because of the assumed wasteful character of these expenditures. This brings us to an important limitation of the models of influence function. Since the influence of expenditures (pressure) is assumed but not explained, it is not clear why policymakers would behave this way. The government is a 'black box', and there is no benchmark showing the consequences of having no interest group activity. It is also not clear on what kind of activities resources are spent by the (exogenously given) interest groups.

Vote Functions

More specific regarding interest group activity, are models focusing on campaign contributions. Although the importance of this type of activity is not undisputed, for the US at least, a relative abundance of data makes this focus attractive.[10] Two types of models can be distinguished. In *exchange models* contributions to a candidate are assumed to elicit a preferred policy response (Welch 1980).[11] Because of the simply

---

[6] This result also follows if the efficiency cost to taxpayers decreases when the tax per individual falls due to an increase in the number of taxpayers. The opposition by taxpayers to subsidies decreases in that case.

[7] According to Becker this happens via the persuasion of 'rationally ignorant' voters.

[8] The last two results concerning group size are demonstrated in Potters and van Winden (1996).

[9] Neary (1997) compares rent seeking models with economic models of conflict. For a recent rent seeking model incorporating a constitutional stage, see Sutter (2002).

[10] According to Wright (1990) the ratio of campaign contributions to lobbying expenditure is about 1 to 10.

assumed positive relationship between contributions and policies (platforms) these models are similar to the models just discussed. One interesting outcome is that groups will generally split contributions between candidates, while contributions will rise with the probability of electoral success (assumed to be given).

In contrast, the *support models* of campaign contributions assume that interest groups take policies as given but try to increase the number of votes for the favored candidate (Brock and Magee 1980; Hillman and Ursprung 1988; Pedersen, 1995; Persson and Tabellini, 2000). In this case, the probability of electoral success is assumed to be positively related to contributions. Under some plausible additional assumptions the following results are obtained (Potters and van Winden 1996): (a) groups will only contribute to the favored candidate, (b) the more preferred the policy of the favored candidate the higher the contribution, (c) no contributions are made if platforms are identical, and (d) contributions are higher the 'closer' the election. Regarding the optimal behavior of the candidates it is typically assumed that (informed) voters will punish candidates for adjusting policies in the direction favored by the campaign donors. Consequently, candidates may on balance (at some point) start to lose votes when (further) catering to interest groups to raise campaign contributions (Denzau and Munger 1986).[12]

Compared to influence function models, a strong point of these models is not only their explicitness regarding interest group activity but also that they open up the 'black box' of policymaking by introducing candidates. The assumption of a vote function introduces another 'black box,' however, concerning the nature of the mechanism through which money buys votes.[13]

Common Agency Models of Contributions

One approach actually explaining why influence occurs is the common agency or menu auction model of Bernheim and Whinston (1986). To illustrate, suppose that part of the electorate is organized in n interest groups or lobbies. Let the joint welfare of the members of interest group i be denoted by $v^i(x)$, and that of the unorganized by $v^u(x)$, where x represents government policy. Before the policy is determined, the lobbies offer contributions to the policymaker contingent on the value of x, denoted by the contribution schedules $c^i(x)$. The net welfare of group i

---

[11] In some models contributions are exchanged for services which are assumed to be independent from policies (e.g., Baron, 1989). Apart from the fact that it is difficult to visualize such services (cf. Morton and Cameron, 1992), these models are of little help in analyzing the influence of interest groups on policies (for more discussion, see Austen-Smith, 1997).

[12] Usually, candidates are assumed to play Nash amongst each other and to act as Stackelberg leaders with respect to the interest group(s). Edelman (1992) reverses the latter assumption.

[13] Austen-Smith (1987) assumes that campaign expenditures enable a candidate to clarify her or his policy position, which is appreciated by risk-averse voters because it reduces the variance of the perceived policy. This reduction is exogenously given, though, and only bites out of equilibrium. See also Coate (2004).

equals $w^i(x) = v^i(x) - c^i(x)$. The policymaker is assumed to care about total contributions $c(x) = \sum_i c^i(x)$ (for campaign spending or other reasons) and aggregate welfare $v(x) = \sum_i v^i(x) + v^u(x)$ (due to re-election concerns, for instance; see below). More specifically, it is assumed that the policymaker's objective is to maximize $c(x) + \gamma v(x)$, with $\gamma \geq 0$.[14] The game between the lobbies and the policymaker consists of two stages: first, the interest groups simultaneously commit to a contribution schedule, followed by the policymaker committing to a policy. In equilibrium, contribution schedules $\{c^i(x)\}$ are such that each lobby maximizes the net joint welfare of its members, given the schedules of the other groups and the anticipated policy response of the policymaker, while the policy x is such that it maximizes the policymaker's objective, taking the contribution schedules as given. Focusing on 'truthful Nash equilibria'[15] this model has the interesting property that the policymaker sets policy x in accordance with the maximization of $(1 + \gamma)\sum_i v^i(x) + \gamma v^u(x)$, which is clearly a function of the form $P(x)$ above. Thus, it provides a micro-foundation for such a political welfare function and an explicit behavioral model for the link between influence weights and pressure in the interest function approach. Note that the welfare of individuals represented by the lobbies has a larger weight, and that the numerical strength of social groups plays a role (since v^i and v^u denote joint welfare). Not surprisingly, competition by other groups can dramatically affect the benefits from lobbying. Only a single (monopolistic) lobby can capture all the surplus from lobbying—by just compensating the policymaker for selecting a different policy—because it leaves no alternative for the policymaker.

Applications of the model concern international trade policies (Grossman and Helpman 1994, 1995), electoral competition (Grossman and Helpman 1996; Prat, 2002a), public goods (Besley and Coate 2001), redistribution (Dixit, Grossman, and Helpman 1997; Grossman and Helpman 1998), local public good and fiscal federalism (Mazza and van Winden 2002; Persson 1998; Persson and Tabellini 1994), capital taxation (Marceau and Smart 2003), environmental policies (Aidt 1998), labor market policies (Rama and Tabellini 1998), and legislative bargaining (Persson 1998; Dharmapala 1999a).

Extensions are presented by Dixit, Grossman, and Helpman et al. (1997) who allow for preferences that are not quasi-linear,[16] Bergemann and Välimäki (2003) who extend the model to a multi-period game, and Prat and Rustichini (1998) who consider a multi-agent setting. Variants of the model include sequential lobbying (Prat and Rustichini 1998), and so-called 'natural equilibria' where principals offer

---

[14] A similar function holds if the policymaker cares about net (of contributions) welfare, given that contributions are valued higher than the same amount in the public's purse.

[15] Bernheim and Whinston (1986) show that the set of best responses to any strategies played by opponents includes a strategy that is 'truthful', which means that it reflects the true preferences of the interest group; moreover, such equilibria are 'coalition proof', in the sense that players cannot improve their lot through costless pre-play communication which carries no commitment.

[16] Quasi-linear preferences imply constant marginal utility of income which frustrates a concern for redistribution (via money transfers).

contributions for at most one instead of all possible policy alternatives (Kirchsteiger and Prat 2001).

Interestingly, Grossman and Helpman (1996) demonstrate that the function maximized by the policymaker can be endogenously obtained in an electoral competition model where parties maximize their seat shares in a legislature and where contributions can influence platforms as well as voting behavior (through campaign expenditures).[17] Dixit, Grossman, and Helpman (1997), furthermore, show that more efficient policy instruments will be used in equilibrium when they are available, which supports the argument of Becker (1983). However, in contrast with Becker's 'black box' model, interest groups may prefer the government to be institutionally restricted to inefficient redistributive policies, because distortions (accompanied by welfare losses) make it more difficult to exploit them.

Although providing an explicit behavioral model of interest group influence, which is a major achievement, existing common agency models rely on some strong assumptions. For example, interest groups are exogenously given, of fixed size and are assumed to behave as unitary actors. Also, players are supposed to stick to their choices, which may be due to reputation concerns in a repeated game (Harrington 1993), but is simply assumed here. Moreover, essentially complete information is assumed, a major restriction which is relaxed in the models discussed next.

Information Transmission Models

An important kind of interest group activity neglected in the models discussed so far is the transmission of information. Think of the endorsement of electoral candidates or the information conveyed to candidates regarding issues that are important to electoral groups. Moreover, the essential role they play in informing policymakers of the likely consequences of policies is not restricted to elections. Interest groups are often better informed about issues that are relevant to them. However, due to conflicts of interests, strategic behavior (dissembling) by interest groups may be expected, which makes the study of this topic not at all trivial. To illustrate, I will discuss the basic signaling model of lobbying of Potters and van Winden (1992). Suppose that a policymaker has to choose between two policies, x_1 and x_2. The payoffs of these policies to the policymaker and an interest group are determined by the 'state of the world', which is either t_1 or t_2, in the following way:

	t_1	t_2
x_1	$a_1, 0$	$0, 0$
x_2	$0, b_1$	a_2, b_2

---

[17] Note that this microfoundation of a political welfare function hinges on the sorting of individuals into organized interest groups, and not on the grouping of individual voters by policymakers because of their shared characteristics (as in probabilistic voting models).

with a_i (b_i), denoting the normalized payoff to the policymaker (interest group), assumed to be positive $(a_i, b_i > 0, i = 1, 2)$.[18] Thus, if the state is t_i, the policymaker prefers x_i while the interest group always prefers x_2: there is a partial conflict of interest.[19] Which state prevails is assumed to be private information to the group; that is, the group knows its 'type', which is either 't_1' or 't_2'. The policymaker only knows the probability, p $(1 - p)$, that the group is of type t_2 (t_1). Assuming that $p < a \equiv a_1/(a_1 \cdot a_2)$ the policymaker will pick x_1 on the basis of her prior belief p However, before the policymaker decides, the group can either send a message (m) against a fixed cost $(c > 0)$, or no message (n), which is costless. Let s_i denote the probability that type t_i sends a message (m), and $r(s)$ the probability that the policymaker responds with x_2 after signal $s = m,n$. Then, the following (sequential) equilibrium of this signaling or sender-receiver game is obtained: (1) if $b_1 < c < b_2$ $s_1 = 0, s_2 = 1, r(n) = 0$ and $r(m) = 1$; (2) if $c < b_1 < b_2$ $s_1 = p(1 - a)/(1 - p)a$, $s_2 = 1, r(n) = 0$ and $r(m) = c/b_1$.[20] In regime (1) lobbying costs are prohibitive for the 'bad' type t_1 (who wants to dissemble), but not for the 'good' type t_2 (who wants to convey the truth). Consequently, only the latter sends a message, enabling the policymaker to make fully informed decisions. If lobbying costs are not prohibitive, regime (2), the good type (with the larger stake) again always lobbies, whereas the bad type does so only from time to time. Fully mimicking the good type would induce the policymaker to stick to x_1, because she would not be able to distinguish between the types. By sometimes responding to a message with this policy, however, the policymaker discourages the bad type from doing so. Since a message may come from both types, lobbying is clearly less informative in this regime. Note that lobbying increases with p (reflecting the inclination of the policymaker to choose x_2), with lower costs, and higher stakes (via a switch from regime (1) to (2)). The influence of lobbying, $r(m)$, increases with higher costs, and lower stakes (b_1). In this equilibrium (with $p < a$) lobbying can never be detrimental to the policymaker or the interest group (*ex ante*, that is). However, in the case that $p > a$ an equilibrium exists, where the group, irrespective of its type, always lobbies, although the response of the policymaker (x_2) remains the same as with no lobbying. This shows that lobbying can be a pure social waste.

---

[18] In terms of payoff functions $v^i(x)$, with the policymaker (interest group) denoted by $i = 1(2)$, it is assumed that: $v^1(x_1;t_1) - v^1(x_2;t_1) > 0$, $v^1(x_2;t_2) - v^1(x_1;t_2) > 0$ and $v^2(x_2;.) - v^2(x_1;.) > 0$.

[19] if $b_1 < 0 < b_2$, there is no conflict of interests and no problem for information transmission, since the group has no incentive to dissemble. Even costless messages can be effective, then. If $b_2 < 0 < b_1$, there is a full conflict of interests with no scope for information transfer, because the group always wants to dissemble which is anticipated by the policymaker. The fact that 'cheap talk' (costless messages) can be informative if the sender's preferences regarding the receiver's actions are dependent on the former's private information is shown more generally in the seminal paper by Crawford and Sobel (1982).

[20] Note the condition that $b_2 > b_1$. If the reverse of this 'sorting condition' would hold, the 'bad' type t_1 has a larger stake in persuading the policymaker. Since the latter will then be inclined to interpret a message as coming from t_1 rather than t_2, no messages will be sent in that case.

The model illustrates that lobbies should somehow be able to distinguish themselves in order to influence policies through information transfer. Fixed lobbying costs provide one such opportunity. The model can be extended in several directions, generally increasing the scope for information transfer (for surveys, see Austen-Smith 1997; Sloof 1998; Grossman and Helpman 2001). Consider, for instance, endogenous lobbying costs. If the interest group can determine the cost, a full revelation equilibrium can always be obtained by having the good type profitably outspend the bad type (in the example, by choosing c at least equal to b_1; Potters and van Winden 1992). However, also the policymaker can make lobbying costly, by demanding a fee or contributions for access (Austen-Smith 1995; Lohmann 1995). The reason may be a time constraint, the intrinsic valuation of contributions, or to screen the lobbies. Also in this way the scope for information transfer increases, by forcing lobbies to reveal their preferences. Other extensions, with a similar outcome, include multiple senders (Potters 1992; Austen-Smith and Wright 1992), multiple receivers (Ainsworth and Sened 1993), multidimensional policies (Battaglini 2002), receiver uncertainty about whether the sender is informed (Austen-Smith 1994), auditing and verification by the policymaker or an intermediary agent (Potters and van Winden 1992; Austen-Smith and Wright 1992; Rasmusen 1993), and persuasion games (Lagerlöf 1997; Bennedsen and Feldmann 2002a). In a persuasion game the sender can transmit or withhold evidence, but cannot 'lie.' This assumption is sometimes justified by referring to reputational concerns in a repeated game.[21] In the above example, the interest group would only be able to reveal its type (t_i) or to abstain from lobbying. This obviously increases the scope for information transfer. Actually, a persuasion game can be seen as one extreme of a more general static model with exogenous cost of lying (which are infinite, then), and the basic signaling game (where lying is costless) as the other extreme. These costs can be endogenized in a repeated signaling game model, where an interest group may want to report truthfully to build up or maintain its reputation. Moreover, apart from costly messages ('words'), sanctions through the enforcement of threats ('deeds') become available then as a means of influence. See the integrated model of Sloof and van Winden (2000).

Applications concern: fiscal policies and regulation (Potters and van Winden 1992; Lohmann 1998; Esteban and Ray 2000), legislative voting and its institutional features (Austen-Smith and Wright 1992; Ainsworth 1993; Austen-Smith 1993; Bennedsen and Feldmann 2002a, 2002b), international trade negotiations (Milner and Rosendorff 1985), the emergence of lobbyists (Ainsworth and Sened 1993), legislative control of bureaucracy (Epstein and O'Halloran 1995; Sloof 2000), and issues related to political campaigning, like contributions and endorsements (Cameron and Jung 1992; Austen-Smith 1995; Lohmann 1995; Potters et al. 1997; Grossman and Helpman 1999; Sloof 1999; Prat 2000, 2002a).

---

[21] This may also justify the (exogenous) cost of lying in the model of Austen-Smith and Wright (1992). In this model (two) interest groups have to pay a cost to get informed (observed by the policymaker) but can subsequently send a costless message (which would be uninfluential 'cheap talk' were it not for the anticipated cost of lying).

Models of information transmission also typically assume that interest groups are of fixed size and behave like a unitary actor.[22] Their comparative strength relates to the fact that they deal with a crucial problem in actual politics, the lack of information. Furthermore, often no (exogenous) commitment assumption is relied on. However, this is bought with simplicity in terms of issues and institutions investigated, which restricts their usefulness. Another worrisome feature concerns the strong rationality assumptions (Sadiraj et al., 2005, 2006). Nevertheless, as a benchmark these models serve a useful purpose. For one thing, due to the relationship between lobby expenditures and influence—qualified by the incentives of interest groups—an informational microfoundation is provided for the use and possibly the specification of an influence function (Lohmann 1995) as well as a political welfare function (Potters and van Winden 1990).

Multiple Means and Channels

So far attention has been focused on one means of influence (contributions or information transmission) and one channel of influence (mostly the nexus with politicians). In practice, however, interest groups can use multiple means and multiple channels. Drawing conclusions from studies focusing on just one means or channel can be treacherous, because the use and impact of these different instruments is not likely to be independent. For instance, common agency models predict that contributions buy policies. However, if contributions simultaneously transmit information on the lobby's type or only serve to gain access, signaling models suggest that this relationship is much more subtle and may even be absent. We now turn to the relatively few models dealing with this multiplicity.

Multiple Means

The following means of influence can be distinguished:[23] (1) lobbying, (2) pressure, (3) structural coercion, and (4) representation. Models of lobbying—the use of 'words'—typically involve costly messages in the transmission of information. However, if preferences are sufficiently aligned cheap talk messages may also be informative and influential. Austen-Smith and Banks (2002) focus on the consequences of adding the option for a sender to inflict self-imposed utility losses and demonstrate that the scope for information transfer and influence increases. In case

---

[22] An exception is Sadiraj, Tuinstra, and van Winden (2006). In this dynamic 'bounded rationality' model the participation of voters in interest groups is endogenous. The fees paid by those who join are (conditionally) supplied to political candidates to finance polling (for learning the preferences of voters).

[23] See van Winden (1983, in particular pp. 16 and 94). Pressure is here distinguished from lobbying.

of pressure—the use of 'deeds'—(opportunity) costs are inflicted on the policy-maker. Contributions in common agency models are one example, where in general contributions may stand for anything that is valued by the policymaker and costly to the interest group (campaign contributions, bribes, ghost writing, etc.). Another example is punishment (instead of reward) through the enforcement of a threat, like a labor or investment strike or a terrorist act. Bennedsen and Feldmann (2006) combine a common agency model with a persuasion game to allow an interest group the choice between lobbying and pressure via contributions. According to their analysis contributions are a more effective means of influence, which may crowd out the search for and transmission of information. Sloof and van Winden (2000) investigate the choice between lobbying and pressure via the enforcement of threats in a repeated signaling game. It turns out that pressure—in contrast to lobbying—only occurs when the interest group's reputation is 'low' (think of a new group). Moreover, (repeated) lobbying cannot completely substitute for pressure, but may be necessary to maintain a reputation. It is concluded that pressure is typically exerted to build up a reputation while lobbying is used to maintain a reputation.

Structural coercion refers to constraints on the behavior of a policymaker which are not related to influence attempts. The behavior of voters (with negligible individual influence) forms a constraint of this type. Through the use of endorsements, or campaign contributions after policies have been determined, interest groups may affect voting and thereby influence the political process. Potters, Sloof, and van Winden (1997) investigate an interest group's choice between endorsement and contributions, using a signaling game, and show that the group may prefer contributions (indirect endorsements) when the preferences of the group and the voter are not sufficiently aligned. This model also provides a micro-foundation for the impact of campaign expenditures on voting.

In case of 'representation', finally, interest groups try to get their interests directly represented among the policymakers.[24] This may be achieved in different ways: through 'multiple positions' (a form of penetration where, for example, via an election a position of policymaker is obtained), 'revolving doors' (offering future career opportunities), and the development of 'social ties' (affective bonds; see Harsanyi 1962). To our knowledge, there are no models yet incorporating this means of influence. Extension of the so-called citizen-candidate model of representative democracy (Osborne and Slivinski 1996; Besley and Coate 1997) may be helpful, though, to deal with penetration, while the model of van Dijk and van Winden (1997) may be useful for social ties.

---

[24] Interestingly, the empirically often observed lobbying of friendly legislators instead of opponents may be related to a bias in representation, since committees tend to share the same biases as the interest groups surrounding them (Kollman, 1997).

Multiple Channels

In practice, interest groups have many different channels of influence available. For example, they can choose between different legislators, bureaucrats,[25] or political candidates (at home, but also abroad).[26] They may also approach several of them to expand supportive coalitions. Moreover, policymakers may be targeted at different tiers within a single governmental body (e.g., the legislative and the bureaucratic tier) as well as at different governmental levels (like the municipal, state, or national level). In addition, an interest group can go for it alone, hire professionals, form an alliance with others, or support an intermediary organization.

Austen-Smith (1993) studies the lobbying of legislators at the agenda setting stage (committee) and the voting stage (House). His signaling model predicts that only agenda stage lobbying is generically influential. Dharmapala (1999a, b) demonstrates with a common agency model the impact of legislative committee structure on policy outcomes when interest groups can offer contributions to different legislators. Prat (2002b) provides a micro-foundation for split contributions to candidates (cf. section 3), using a common agency model. Models concerning the efforts of interest groups to expand supportive coalitions are lacking (Hojnacki and Kimball 1998).

Several studies investigate the choice between legislators and bureaucrats. Moore and Suranovic (1992) look at the case where import-competing industries can pursue import relief via administered protection or via lobbying politicians directly (assuming exogenous probabilities of success). Their analysis suggests that reform restricting one of these options may cause welfare losses through substitution effects. Mazza and van Winden (forthcoming), using a common agency model, look at various issues related to the interaction between a legislator (deciding on a budget) and a bureaucrat (deciding on the allocation of the budget) when both can be offered contributions by interest groups. Their results show that competition between interest groups may function as a substitute for legislative control, while the budget may be used as a second-best instrument of control (a smaller government being the legislative response to bureaucratic capture). Sloof (2000) studies a politician's decision whether or not to delegate policy authority to a bureaucrat when both can be lobbied. His signaling game analysis shows that politicians may prefer a biased bureaucracy and an interest group with a large stake, because the informational gains may outweigh the distributional losses. Moreover, interest groups would typically lobby politicians to further delegation.

Hoyt and Toma (1989), using an influence function model, consider the choice between state and local governmental levels as targets for interest groups, when states mandate expenditure and revenues of local governments. Their analysis

---

[25] Relatively few studies focus on the influence of interest groups on the bureaucracy; see e.g., Spiller (1990), Laffont and Tirole (1991), and Banks and Weingast (1992).

[26] See Hillman and Ursprung (1988), Hillman (1989).

suggests that payoffs from influence at the state level generally will exceed that at the local level.

Interest groups may also delegate the influencing of decision making at another level to a policymaker. For example, policymakers are used as an intermediary when campaign contributions are offered to candidates to affect voting behavior. The choice between working alone and hiring a lobbyist is modeled by Johnson (1996), while van Winden (1983) addresses the budget allocation decision regarding the alternative of joining an alliance (like a trade organization); both authors use influence functions. These choices clearly relate to the internal organization of an interest group, a neglected topic in the literature (Moe, 1980, Rothenberg 1988).

Concluding Remarks

An important achievement of the literature surveyed in this paper is the successful incorporation of interest group behavior and influence in the formal positive analysis of political decision making. It has helped to redress the imbalance in Public Choice created by a disproportional attention for the electoral nexus between policymakers and voters. Interest groups also impact government policies outside elections by employing resources that are valuable in both contexts (particularly, money and information). Gradually, a more rigorous and also more positive view of the functioning of interest groups has been established.

Notwithstanding the progress made, there are still many blind spots in our understanding of the political economic role played by interest groups. Firstly, notwithstanding the huge number of empirical studies there are relatively few 'stylized facts,' basically showing that contributions and lobbying, the size of organized membership, and an interest group's stake are positive determinants of influence, whereas the presence of an oppositional force in the political arena, electoral pressures, and the presence of a well-informed electorate are negative determinants (Potters and Sloof, 1996). The main problems are a serious lack of data and a shortage of hypotheses derived from theoretical models that provide structure and a base for embedding. Laboratory experimentation forms an important complementary research method, because of the opportunity it offers to study behavior under controlled conditions and to check the robustness of findings through replication. However, until now only few experiments have been carried out, with mixed success for the models (van Winden 2002). Also, to compensate for the extreme simplicity of many formal models, which may lead to a distorted view of the multi-faceted interaction between interest groups and policymakers (Saint-Paul 2000), computer experiments (simulations) allowing for greater complexity and dynamics should receive more attention (Fafchamps et al. 1999; Sadiraj et al. 2005, 2006).

Secondly, existing models typically assume a level of rationality which seems unrealistic, though useful as a benchmark. There is mounting evidence that people are quite myopic, use rather simple adaptive rules of decision making, and

concentrate on issues at hand (Ortoni et al. 1988; Kagel and Roth 1995, Camerer 2003). A related issue concerns the impact of emotions and feelings. Although investigated in a few studies of political behavior,[27] the subject has been neglected in political economics (van Winden, 2007) and, more particularly, the literature on interest groups. Interestingly, allowing for affective social ties in the interaction between a policymaker and an interest group would not only imply that the former may be willing to benefit the latter without compensation, but also that the interest group may care about the interests of the policymaker.[28]

Thirdly, research needs to go beyond the common assumption of exogenously given groups that are of fixed size and behave as unitary actors. The formation, dynamics, and internal politics of interest groups are badly neglected topics.[29] Why do only some interests get organized, or are induced to do so by policymakers? Why, for instance, are the retired in the US well organized in the intergenerational 'redistribution game' while there is no comparable organization for the younger people (Lohmann 1998)? Furthermore, what are the nature, cause, and impact of the decision-making procedures maintained by organized interests? And how do government policies feed back into the development of groups?

The considerable theoretical progress made in recent years will serve as a fresh source for the derivation and testing of competing hypotheses and for structuring the search for new data. In addition, it has helped the developing of a framework for the interpretation, coordination, and planning of future research. Notwithstanding the substantial progress, much remains to be done.

References

Abelson, R.P., Kinder, D.R., Peters, M.D., and Fiske, S.T. (1982). "Affective and semantic components in political person perception." *Journal of Personality and Social Psychology*, 42, 619–630.

Aidt, T. (1998). "Political internalization of economic externalities and environmental policy." *Journal of Public Economics*, 69, 1–16.

Ainsworth, S. (1993). "Regulating lobbyists and interest group influence." *The Journal of Politics*, 55, 41–56.

Ainsworth, S. and Sened, I. (1993). "The role of lobbyists: entrepreneurs with two audiences." *American Journal of Political Science*, 37, 834–866.

Aumann, R. and Kurz, M. (1977). "Power and taxes." *Econometrica*, 45, 1137–1161.

---

[27] A theoretical example is the 'minimax regret strategy' (Ferejohn and Fiorina, 1974) to explain voting in large-scale elections. For empirical studies, see Abelson, Kinder, Peters, and Fiske (1982), Marcus and Mackuen (1993), and Bosman and van Winden (2002). Further See van Winden (2007).

[28] See van Dijk and van Winden (1997) for a theoretical model. Empirical support is presented in van Dijk, Sonnemans, and van Winden (2002) and Brandts, Riedl, and van Winden (2007).

[29] Some recent attempts to endogenize group formation include Dougan and Snyder (1996), Mitra (1999), and Damania and Frederiksson (2000). Felli and Merlo (2001) focus on the opportunity for a policymaker to select lobbies to bargain with.

Aumann, R., Kurz, M., and Neyman, A. (1983). "Voting for public goods." *Review of Economic Studies*, 50, 677–693.

Austen-Smith, D. (1987). "Interest groups, campaign contributions, and probabilistic voting." *Public Choice*, 54, 123–139.

Austen-Smith, D. (1993). "Information and influence: lobbying for agendas and votes." *American Journal of Political Science*, 37, 799–833.

Austen-Smith, D. (1994). "Strategic transmission of costly information." *Econometrica*, 62, 955–963.

Austen-Smith, D. (1995). "Campaign contributions and access." *American Political Science Review*, 89, 566–581.

Austen-Smith, D. (1997). "Interest groups: money, information and influence," in D.C. Mueller (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press.

Austen-Smith, D. and Banks, J.S. (2002). "Costly signaling and cheap talk in models of political influence." *European Journal of Political Economy*, 18, 263–280.

Austen-Smith, D. and Wright, J.R. (1992). "Competitive lobbying for a legislator's vote." *Social Choice and Welfare*, 9, 229–257.

Banks, J.S. and Weingast, B.R. (1992). "The political control of bureaucracies under asymmetric information." *American Journal of Political Science*, 36, 509–524.

Baron, D. (1989). "Service-induced campaign contributions and the electoral equilibrium." *Quarterly Journal of Economics*, 104, 45–72.

Battaglini, M. (2002). "Multiple referrals and multidimensional cheap talk." *Econometrica*, 70, 1379–1401.

Becker, G. (1983). "A theory of competition among pressure groups for political influence." *Quarterly Journal of Economics*, 98, 371–400.

Bennedsen, M. and Feldmann, S.E. (2002a). "Lobbying legislatures," *Journal of Political Economy*, 110, 919–946.

Bennedsen, M. and Feldmann, S.E. (2002b). "Lobbying and legislative organization, the effect of the vote of confidence procedure." *Business and Politics*, 4, 187–203

Bennedsen, M. and Feldmann, S.E. (2006). "Informational lobbying and political contributions." *Journal of Public Economics*, 90, 631–656.

Bergemann, D. and Välimäki, J. (2003). "Dynamic common agency." *Journal of Economic Theory*, 111, 23–48.

Bernheim, B.D. and Whinston, M.D. (1986). "Menu auctions, resource allocation, and economic influence." *Quarterly Journal of Economics*, 101, 1–31.

Besley, T. and Coate, S. (1997). "An economic model of representative democracy." *Quarterly Journal of Economics*, 112, 85–114.

Besley, T. and Coate, S. (2001). "Lobbying and welfare in a representative democracy." *Review of Economic Studies*, 68, 67–82.

Borooah, V. and van der Ploeg, F. (1983). *Political Aspects of the Economy*. Cambridge: Cambridge University Press.

Bosman, R. and van Winden, F. (2002). "Emotional hazard in a power-to-take experiment." *The Economic Journal*, 112, 147–169.

Bottomore, T.B. (1970). *Elites and Society*. Harmondsworth: Penguin Books.

Brandts, J., Riedl, A. and van Winden, F. (2007). *On Competition and Well-being*. mimeo, CREED, University of Amsterdam.

Brock, W. and Magee, S. (1980). "Tariff formation in a democracy." in J. Black and B. Hindley (eds.) *Current Issues in Commercial Policy and Diplomacy*. London: MacMillan.

Cairns, R. (1989). "Dynamic rent-seeking." *Journal of Public Economics*, 39, 315–334.

Camerer, C.F. (2003). *Behavioral Game Theory*. Princeton: Princeton University Press.

Cameron, C.M. and Jung, J.P. (1992). *Strategic Endorsements*. mimeo, New York: Columbia University.

Coate, S. (2004). "Political competition with campaign contributions and informative advertising." *Journal of the European Economic Asscociation*, 2, 772–804.

Coggins, J., Graham-Tomasi, T., and Roe, T. (1991). "Existence of equilibrium in a lobbying economy." *International Economic Review*, 32, 533–550.

Coughlin, P. (1992). *Probabilistic Voting Theory*. Cambridge: Cambridge University Press.

Coughlin, P., Mueller, D., and Murrell, P. (1990). "Electoral politics, interest groups, and the size of government." *Economic Inquiry*, 29, 682–705.

Crawford, V. and Sobel, J. (1982). "Strategic information transmission." *Econometrica*, 50, 1431–1451.

Denzau, A. and Munger, M. (1986). "Legislators and interest groups: how unorganized interests get represented." *American Political Science Review*, 80, 89–106.

Dharmapala, D. (1999a). "Comparing tax expenditures and direct subsidies: the role of legislative committee structure." *Journal of Public Economics*, 72, 421–454.

Dharmapala, D. (1999b). *Legislative Bargaining and Vote Buying Under the Influence of Lobbying,*. mimeo, Cambridge: Harvard University.

Dixit, A., Grossman, G.M., and Helpman, E. (1997). "Common agency and coordination: general theory and application to government policy making." *Journal of Political Economy*, 105, 752–769.

Dougan, W.R. and Snyder, J.M., Jr. (1996). "Interest-group politics under majority rule." *Journal of Public Economics*, 61, 49–71.

Drazen, A. (2000). *Political Economy in Macroeconomics*. Princeton: Princeton University Press.

Drissen, E. (1999). *Government Decisions on Income Redistribution and Public Production*. Amsterdam: Thela Thesis.

Edelman, S. (1992). "Two politicians, a PAC, and how they interact: two extensive form games." *Economics and Politics*, 4, 289–305.

Epstein, D. and O'Halloran, S. (1995). "A theory of strategic oversight: congress, lobbyists, and the bureaucracy." *Journal of Law, Economics, and Organization*, 11, 227–255.

Esteban, J. and Ray, D. (2000). "Wealth constraints, lobbying and the efficiency of public allocation." *European Economic Review*, 44, 694–705.

Fafchamps, M., De Janvry, A., and Sadoulet, E. (1999). "Social heterogeneity and wasteful lobbying." *Public Choice*, 98, 5–27.

Ferejohn, J.A. and Fiorina, M.P. (1974). "The paradox of not voting: a decision theoretic analysis." *American Political Science Review*, 68, 525–536.

Findlay, R. and Wellisz, S. (1983). "Some aspects of the political economy of trade restrictions." *Kyklos*, 36, 469–481.

Gardner, R. (1981). "Wealth and power in a collegial polity." *Journal of Economic Theory*, 25, 353–366.

Grossman, G.M. and Helpman, E. (1994). "Protection for sale." *American Economic Review*, 84, 833–850.

Grossman, G.M. and Helpman, E. (1995). "Trade wars and trade talks." *Journal of Political Economy*, 103, 675–708.

Grossman, G.M. and Helpman, E. (1996). "Electoral competition and special interest politics." *Review of Economic Studies*, 63, 265–286.

Grossman, G.M. and Helpman, E. (1998). "Intergenerational redistribution with short-lived governments." *The Economic Journal*, 108, 1299–1329.

Grossman, G.M. and Helpman, E. (1999). "Competing for endorsements." *American Economic Review*, 89, 501–524.

Grossman, G.M. and Helpman, E. (2001). *Special Interest Politics*. Cambridge: MIT Press.

Harrington, J.E. (1993). "The impact of reelection pressures on the fulfilment of campaign promises." *Games and Economic Behavior*, 5, 71–97.

Harsanyi, J. (1962). "Measurement of social power, opportunity costs, and the theory of two-person bargaining games." *Behavioral Science*, 7, 67–80.

Hettich, W. and Winer, S.L. (1999). *Democratic Choice and Taxation*. Cambridge: Cambridge University Press.

Hillman, A. (1989). *The Political Economy of Protection*. Chur: Harwood.

Hillman, A. and Ursprung, H. (1988). "Domestic politics, foreign interests, and international trade policy." *American Economic Review*, 78, 729–745.

Hirshleifer, J. (1976). "Comment (on: Peltzman 1976)." *Journal of Law and Economics*, 19, 240–244.

Hojnacki, M. and Kimball, D.C. (1998). "Organized interests and the decision of whom to lobby in Congress." *American Political Science Review*, 92, 775–790.

Hoyt, W.H. and Toma, E.F. (1989). "State mandates and interest group lobbying." *Journal of Public Economics*, 38, 199–213.

Johnson, P.E. (1996). "Corporate political offices in a rent-seeking society." *Public Choice*, 88, 309–331.

Kagel, J.H. and Roth, A.E. (eds.) (1995). *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Kirchsteiger, G. and Prat, A. (2001). "Inefficient equilibria in lobbying." *Journal of Public Economics*, 82, 349–375.

Kollman, K. (1997). "Inviting friends to lobby: interest groups, ideological bias, and congressional committees." *American Journal of Political Science*, 41, 519–544.

Laffont, J. and Tirole, J. (1991). "The politics of government decisionmaking: a theory of regulatory capture." *Quarterly Journal of Economics*, 106, 1089–1127.

Lagerlöf J. (1997). "Lobbying, information, and private and social welfare." *European Journal of Political Economy*, 13, 615–637.

Lohmann, S. (1995). "Information, access, and contributions: a signaling model of lobbying." *Public Choice*, 85, 267–284.

Lohmann, S. (1998). "An informational rationale for the power of special interests." *American Political Science Review*, 92, 809–827.

Marceau, N. and Smart, M. (2003). "Corporate lobbying and commitment failure in capital taxation." *American Economic Review*, 93, 241–251

Marcus, G.E. and Mackuen, M.B. (1993). "Anxiety, enthusiasm, and the vote: the emotional underpinnings of learning and involvement during presidential campaigns." *American Political Science Review*, 87, 672–685.

Mazza, I. and van Winden, F. (1996). "A political economic analysis of labor migration and income redistribution." *Public Choice*, 88, 333–363.

Mazza, I. and van Winden, F. (2002). "Does centralization increase the size of government? The effects of separation of powers and lobbying." *International Tax and Public Finance*, 9, 379–389.

Mazza, I. and van Winden, F. (forthcoming). "An endogenous policy model of hierarchical government." *European Economic Review*.

Milner, H.V. and Rosendorff, B.P. (1996). "Trade negotiations, information, and domestic politics: the role of domestic groups." *Economics and Politics*, 8, 145–189.

Mitchell, W.C. and Munger, M.C. (1991). "Economic models of interest groups: an introductory survey." *American Journal of Political Science*, 35, 512–546.

Moe, T.M. (1980). *The Organization of Interests*. Chicago: University of Chicago Press.

Moore, M.O. and Suranovic, S.M. (1992). "Lobbying vs. administered protection." *Journal of International Economics*, 32, 289–303.

Morton, R. and Cameron, C. (1992). "Elections and the theory of campaign contributions: a survey and critical analysis." *Economics and Politics*, 4, 79–108.

Neary, H.M. (1997). "A comparison of rent-seeking models and economic models of conflict." *Public Choice*, 93, 373–388.

Nitzan, S. (1994). "Modelling rent-seeking contests." *European Journal of Political Economy*, 10, 41–60.

Olson, M. (1965). *The Logic of Collective Action*. Cambridge: Harvard University Press.

Ortoni, A., Clore, G.L., and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

Osborne, M.J. (1984). "Why do some goods bear higher taxes than others?" *Journal of Economic Theory*, 32, 111–127.

Osborne, M.J. and Slivinski, A. (1996). "A model of political competition with citizen candidates." *Quarterly Journal of Economics*, 111, 65–96.

Peck, R. (1986). "Power and linear income taxes: an example." *Econometrica*, 54, 87–94.

Pedersen, K.R. (1995). "Rent-seeking, political influence and inequality: a simple analytical example." *Public Choice*, 82, 281–305.

Pelzman, S. (1976). "Toward a more general theory of regulation." *Journal of Law and Economics*, 19, 211–240.

Persson, T. (1998). "Economic policy and special interest policy." *The Economic Journal*, 108, 310–327.

Persson, T. and Tabellini, G. (1994). "Does centralization increase the size of government?" *European Economic Review*, 38, 765–773.

Persson, T. and Tabellini, G. (2000). *Political Economics*. Cambridge: MIT Press.

Potters, J. (1992). "Lobbying and pressure." Tinbergen Institute Research Series No. 36, Amsterdam.

Potters, J. and Sloof, R. (1996). "Interest groups: A survey of empirical models that try to assess their influence." *European Journal of Political Economy*, 12, 403–442.

Potters, J., Sloof, R., and van Winden, F. (1997). "Campaign expenditures, contributions and direct endorsements: the strategic use of information and money to influence voter behavior." *European Journal of Political Economy*, 13, 1–31.

Potters, J., and van Winden, F. (1990). "Modelling political pressure as transmission of information." *European Journal of Political Economy*, 6, 61–88.

Potters, J. and van Winden, F. (1992). "Lobbying and asymmetric information." *Public Choice*, 74, 269–292.

Potters, J. and van Winden, F. (1996). "Models of interest groups: four different approaches," in N. Schofield (ed.) *Collective Decision-Making, Social Choice and Political Economy*. Boston: Kluwer, 337–362.

Prat, A. (2000). "An economic analysis of campaign finance." *World Economics*, 1, 13–28.

Prat, A. (2002a). "Campaign spending with office-seeking politicians, rational voters, and multiple lobbies." *Journal of Economic Theory*, 103, 162–189.

Prat, A. (2002b). "Campaign advertising and voter welfare." *Review of Economic Studies*, 69, 997–1017.

Prat, A., and Rustichini, A. (1998). "Sequential common agency." Discussion Paper No. (9895). CentER, Tilburg University.

Prat, A. and Rustichini, A. (2003). "Games played through agents." *Econometrica*, 71, 989–1026.

Przeworski, A. and Wallerstein, M. (1988). "Structural dependence of the state on capital." *American Political Science Review*, 82, 1–29.

Rama, M. and Tabellini, G. (1998). "Lobbying by capital and labor over trade and labor market policies." *European Economic Review*, 42, 1295–1316.

Rasmusen, E. (1993). "Lobbying when the decisionmaker can acquire independent information." *Public Choice*, 77, 899–913.

Renaud, P.S.A. (1989). *Applied Political Economic Modelling*. Berlin Heidelberg New York: Springer.

Renaud, P. and van Winden, F. (1988). "Fiscal behaviour and the growth of government in the Netherlands," in J. Lybeck and M. Henrekson (eds.) *Explaining the Growth of Government*. Amsterdam: North Holland, 133–156.

Rothenberg, L.S. (1988). "Organizational maintenance and the retention decision in groups." *American Political Science Review*, 82, 1129–1152.

Sadiraj, V., Tuinstra, J., and van Winden, F. (2005). "Interest groupsize dynamics and policymaking." *Public Choice*, 125, 271–303.

Sadiraj, V., Tuinstra, J., and van Winden, F. (2006). "A computational electoral competition model with social clustering and endogenous interest groups as information brokers."*Public Choice*, 129, 169–187.

Saint-Paul, G. (2000). "The new political economy: recent books by Allan Drazen and Torsten Persson and Guido Tabellini." *Journal of Economic Literature*, 38, 915–925.

Sloof, R. (1998). *Game-Theoretic Models of the Political Influence of Interest Groups*. Boston: Kluwer.

Sloof, R. (1999). "Campaign contributions and the desirability of full disclosure laws." *Economics and Politics*, 11, 83–107.

Sloof, R. (2000). "Interest group lobbying and the delegation of policy authority." *Economics and Politics*, 12, 247–274.

Sloof, R., and van Winden, F. (2000). "Show them your teeth first!"*Public Choice*, 104, 81–120.

Spiller, P.T. (1990). "Politicians, interest groups, and regulators: a multi-principals agency theory of regulation, or 'let them be bribed." *Journal of Law and Economics*, 33, 65–101.

Stigler, G. (1971). "The theory of economic regulation." *Bell Journal of Economics and Management*, 2, 3–21.

Sutter, D. (2002). "Constitutional prohibitions in a rent seeking model." *Public Choice*, 111, 105–125.

Tollison, R.D. (1997). "Rent seeking," in D.C. Mueller (ed.) *Perspectives on Public Choice*. Cambridge: Cambridge University Press, 506–525.

Tullock, G. (1967). "The welfare costs of tariffs, monopolies, and theft." *Western Economic Journal*, 5, 224–232.

Tullock, G. (1980). "Efficient rent-seeking," in J.M. Buchanan, R.D. Tollison, and G. Tullock (eds.) *Toward a Theory of the Rent-Seeking Society*. College Station: Texas A&M University Press, 97–112.

van Dijk, F., Sonnemans, J., and van Winden, F. (2002). "Social ties in a public good experiment." *Journal of Public Economics*, 85, 275–299.

van Dijk, F. and van Winden, F. (1997). "Dynamics of social ties and public good provision." *Journal of Public Economics*, 323–341.

van Velthoven, B. (1989). *The Endogenization of Government Behaviour in Macroeconomic Models*. Berlin Heidelberg New York: Springer.

van Velthoven, B. and van Winden, F. (1985). "Towards a politico-economic theory of social security." *European Economic Review*, 27, 263–289.

van Winden, F. (1983). *On the Interaction between State and Private Sector*. Amsterdam: North Holland.

van Winden, F. (1999). "On the economic theory of interest groups: towards a group frame of reference." *Public Choice*, 100, 1–29.

van Winden, F. (2002). "Experimental investigation of collective action," in S.L. Winer and H. Shibata (eds.) *Political Economy and Public Finance*. Cheltenham: Edward Elgar, 178–196.

Van Winden, F. (2007). "Affective Public Choice," in J. Casas Pardo and P. Schwartz (eds.), *Public Choice and the Challenges of Democracy*. Cheltenham: Edward Elgar, 45–61.

van Winden, F., Schram, A., and Groot, F. (1987). "The interaction between economics and politics: modelling cycles." *European Journal of Political Research*, 15, 185–202.

Verbon, H. (1989). *The Evolution of Public Pension Schemes*. Berlin Heidelberg New York: Springer.

Welch, W. (1980). "The allocation of political monies: economic interest groups." *Public Choice*, 35, 97–120.

Wright, J.R. (1990). "Contributions, lobbying, and committee voting in the U.S. House of Representatives." *American Political Science Review*, 84, 417–438.

Zusman, P. (1976). "The incorporation and measurement of social power in economic models." *International Economic Review*, 17, 447–462.

Chapter 21
Dictatorship[1]

Ronald Wintrobe

The literature which takes a public choice approach to dictatorship, largely barren before 1990 except for Tullock's *Autocracy* (1987), is now growing and may be entering a period of prosperity. This survey focuses on the most recent literature, and on three questions in particular: (1) The behavior of dictators, including the strategies that dictators use to stay in power; (2) The relative efficiency of dictatorship: Which is better, dictatorship or democracy, in promoting economic growth and efficiency?; and (3) What policies should the democracies adopt to deal with dictatorships if they are interested in promoting freedom?

The Behavior of Dictators

The Dictator's Dilemma

The standard view of the difference between democracy and dictatorship in political science (Friedrich and Brzezinski, 1965) is that dictators can use the tool of repression to stay in power. Thus dictators typically impose restrictions on the rights of citizens to criticize the government, restrictions on the freedom of the press, restrictions on the rights of opposition parties to campaign against the government, or, as is common under totalitarian dictatorship, simply prohibit groups, associations, or political parties opposed to the government. To be effective, these restrictions must be accompanied by monitoring of the population, and by sanctions for disobedience. The existence of a political police force and of extremely severe sanctions for expressing and especially for organizing opposition to the government such as imprisonment, internment in mental hospitals, torture and execution are the hallmark of dictatorships of all stripes.

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 77–90.

However, the use of repression creates a problem for the autocrat. This is the Dictator's Dilemma (Wintrobe, 1990, 1998a)—the problem facing any ruler of not knowing how much support he has among the general population, as well as among smaller groups with the power to depose him. The use of repression of course breeds fear on the part of a Dictator's subjects, and this fear breeds reluctance on the part of the citizenry to signal displeasure with the Dictator's policies. This fear on their part in turn breeds fear *on the part of the dictator*, since, not knowing what the population thinks of his policies, he has no way of knowing what they are thinking and planning, and of course he suspects that what they are thinking and planning is his assassination. The more the dictator rules by repression, i.e., through fear, the more problem is magnified. The more his repressive apparatus stifles dissent and criticism, the less he knows how much support he really has among the population.

From a theoretical point of view, the Dictator's Dilemma originates from the lack of an enforcement mechanism in politics. It is advantageous for the dictator to "buy off" some of his constituents, especially those who may be too powerful to repress, and those whose demands are easily satisfied. So a simple trade of rents or policies for support would solve the Dictator's dilemma, and also allow his subjects to rest easily. But there is no mechanism analogous to legal contractual enforcement which would enforce this trade. Another way to put it is that the dictator and his subjects have a mutual *signaling* problem. In general, the easiest way to overcome the problem of obtaining support is to "overpay" supporters, i.e., to pay them more than they are worth by distributing rents to them. The support of workers can be obtained through paying them excessive wages, of capitalists by giving them monopoly privileges, of particular regions by locating manufacturing facilities in places where they do not really belong but where they are politically valuable, of ethnic groups by giving them special privileges and so on. Of course, similar practices are widespread in democracy where they are known as "pork barrel politics". They are often described as a failure of democracy. But if democracy may be likened to a pork barrel, the typical dictatorship is a temple of pork! That is, these practices appear to be much more widespread under a dictatorship than under democracy.

In sum, while there is always a class of people who are repressed under a dictatorship, there is also, in any successful dictatorship, another class—*the overpaid*. As far as the people in the middle are concerned, the sad thing is that they can side with either group. The general population may be repressed in that their civil liberties may be taken away, but other aspects of the regime may compensate for this as far as they are concerned.

However, *the use of repression does not mean that dictators are unpopular*. Indeed, it sometimes appears from the historical record that the more repressive they were, the more popular they became! All the evidence indicates that Hitler was very popular. Communism was *popular* at one time; when it became unpopular, the regimes fell. Reports in the newspapers suggest that Castro and Saddam Hussein were often popular with their peoples.[2]

---

[2] See for example, John Deutsch, Options: Good and Bad Ways To Get Rid of Saddam, *The New York Herald Tribune*, February 24, 1999, p. 8 on Saddam Hussein's popularity.

That dictatorships use two instruments—repression and loyalty or popularity—to stay in power suggests a useful classification of regimes. Four types can be distinguished: tinpots (low repression and loyalty), tyrants (high repression, low loyalty), totalitarians (high levels of both), and timocrats (low repression, high loyalty). Thus, totalitarian regimes combine high repression with a capacity to generate loyalty. Under tyranny, the regime stays in power through high repression alone and loyalty is low. A tinpot regime is low on both counts. And timocracy implies that loyalty is high at low levels of repression. These four types or *images* have tended to recur over and over in the literature on dictatorship.[3]

This classification may be thought of in three ways. On one level, the regimes simply represent different combinations of the variables loyalty and repression. However, the classification also illuminates behavior, because *the regimes differ in their responses to economic change*. Suppose, for example, that there is an increase in economic growth which raises the Dictator's popularity. Tinpots and timocrats both respond to an increase in popularity by lowering the level of repression; tyrants and totalitarians, by raising it.

A third way to think about the regimes is that they simply represent different solutions (levels of repression and loyal support) to the same general model. Thus, assume that all dictators have the same utility function, where the arguments are consumption (C) and power (π).

$$U = U(\pi, C) \tag{21.1}$$

Power may be desired either for its own sake, or because dictators wish to impose their ideas of common good on society.[4] The tinpot ruler represents the special or "corner" solution where the sole aim is to maximize consumption. On the other hand, the leaders of tyrannies and totalitarian regimes represent the opposite extreme of dictators who maximize power. Finally, timocracy[5] represents the case of benevolent dictatorship, where the Dictator's objective is the welfare of his people. While many, if not all, dictators profess this objective, it is hard to think of an instance where it explains much about their behavior.[6]

---

[3] For details, see Wintrobe (1998a, b), chapter 1.

[4] The model does not distinguish between the desire for power for its own sake (as an end in itself) and the desire for power as a means to implement some other objective, e.g., in order to implement some personal or party preference as government policy. Bernholz (2001, discussed further below) stresses that dictators pursue power in order to implement a vision of society, e.g., Nazism with respect to racial objectives or communism with respect to equality. Wittman emphasized the same point with respect to politicians in a democracy: that they are interested in ideological objectives as well as being elected (Wittman, 1983). In my book I did not rule out other objectives for dictators besides power and consumption, but I tried to see how far one could go with this simple and basic public choice perspective. Of course, sometimes an ideology interferes with the pursuit of power; this could be incorporated into the model in the same way as consumption benefits already are. However, ideology is in part a tool (often incorporated in propaganda) to accumulate power, so the pursuit of power and that of an ideological objective are often difficult to distinguish in practice.

[5] The Greek root of timocracy is *Thymos—to* love. The term is borrowed from Plato's *Republic*.

[6] In my book I suggest the example of the Age of the Antonines, following Gibbon's description of this time as "the happiest the world has ever known", in his *Decline and Fall of the Roman Empire* (1981).

Combining this utility function with a constraint which shows how money can be converted into power and power into money provides an illuminating explanation of the limits to a Dictator's power. Totalitarian dictators in particular appeared to maximize the control of the state over the individual. For example in Nazi Germany, one official suggested that: "the only time an individual has a private life is when he is asleep." What limits a Dictator's pursuit of power? It would be arbitrary to specify that the Dictator's power is limited by a revenue-maximizing tax. For, so long as the dictator has sufficient power, he can raise more funds by imposing new tax bases and by finding other ways to raise money. In short, if there is no limit to his power, there is no limit to his resources either - and *vice versa*. In the end, the constraint on his behavior does not arise from an artificially fixed budget or from arbitrary limits to his power, but from the ultimately diminishing possibilities of transforming money into power and *vice versa*. So the limits to budgetary resources and to power must be simultaneously determined.

More precisely, the dictator is constrained in two ways. The first constraint—the costs of accumulating power—is governed by the political institutions of the regime, and the second—the capacity to use his power to increase revenue—by the Dictator's economy. These constraints are combined in the equation:

$$B(\pi) = P_\pi \pi (B - C) + C \qquad (21.2)$$

The left-hand side of the constraint (2) shows the Dictator's budget B as a function of power (π), i.e., it shows how the Dictator's power may be used (through taxation, regulation or the provision of public goods) to obtain budgetary resources. The right-hand side shows how the funds are "spent": either on consumption, C, or accumulating power π via the money-to-power relation π (B–C), with each unit of π multiplied by P_π—the "price" of power in terms of money.

The solution (first-order conditions) may be obtained by choosing Π and C to maximize (1) subject to the constraint (2). Rearranging the terms, it is expressed simply as

$$\frac{U_c}{U_\pi} = \frac{1}{P_\pi \left(1 - \dfrac{1}{\varepsilon^\pi}\right) - B_\pi} \qquad (21.3)$$

Equation (21.3) displays, in a particularly transparent way, the elements that enter into the Dictator's calculus—the marginal costs of accumulating power $p_\pi[1-(1/\varepsilon^\pi)]$, where $\varepsilon^\pi \equiv (\partial \pi / \partial P\pi)(P_\pi/\pi) > 0$, is the elasticity of π with respect to its price; the marginal effect of power on the Dictator's budget ($B\pi$), and U_C/U_π—is the Dictator's preferences for power vs. consumption.

The first term (P_π) is governed by the Dictator's political apparatus for building loyalty and for repression and the productivity of these instruments in producing power. The second (B_π) shows what the exercise of power does to the Dictator's budget via its effects on the economy, e.g., its effects on economic growth,

economic efficiency, and the capacity to implement taxes. Sometimes (e.g., if power is used to provide a public input, or to raise or implement taxes) the exercise of state power will raise revenue, i.e., $B_\pi > 0$. Sometimes (i.e., if power is used to impose inefficient regulation on industry) power will lower state revenue ($B_\pi > 0$). The third factor (U_C/U_π) simply represents the Dictator's preferences between consumption and power. Sometimes one can see some of the factors at work molding these preferences—e.g., how Party organization or the nature of the Dictator's support can drive him in the direction of maximizing power. But, perhaps more than any other political or economic agent, political dictators have some freedom to put their stamp on society.

This equilibrium provides the limit to power. At the equilibrium, the dictator cannot obtain more power (its marginal cost in money is larger than the extra power required to support this budget) and cannot obtain more revenue (the power required to do so is larger than the revenue necessary to support this level of power). Note the simultaneous equilibrium of money and power.

In turn the model also simultaneously determines the Dictator's consumption, equilibrium level of repression and loyal support. So this model of a utility maximizing dictator can be put together with various types of economic system (communist, apartheid, capitalist-authoritarian, etc. each of which contain values of B_π, ε_π and P_π) to derive implications about the behavior of different regimes—comparative static changes in Π and B, as well as in levels of repression, loyalty, etc. Put differently, the three elements in equation (21.3), determine the nature of the dictatorship—whether the regime resembles more closely that of a tinpot, totalitarian, tyrant, or timocrat.

As far as the economy is concerned, what turns out to be crucial is not whether the Dictator's intervention helps or hurts the economy on the whole, but the effects of *marginal* (further) interventions on economic growth, efficiency, or the Dictator's budget. If this marginal effect (B_π) is positive, whether the total effect is positive or negative within a considerable range, the dictator will tend to be oriented more towards power rather than towards consumption. On the other hand, if the use of power tends to retard growth and other dimensions of economic efficiency rather than favoring it, the dictator tends to be a tinpot. So the marginal economic effects of the Dictator's power helps to determine whether the dictator is tinpot, totalitarian or tyrant.

Islam and Winer (2001) test Wintrobe's theory of non-democratic regimes using a large sample of both non-democratic and democratic countries. Some additional hypotheses about the differences between democratic and non-democratic countries suggested, but not explicitly considered, by Wintrobe are also considered. The results indicate clearly that the relationship between an index of civil and political freedoms and economic growth varies substantially across all regime types. Other aspects of the theory are partially confirmed. In particular, positive growth leads to a reduction in the degree of freedom in totalitarian regimes (that attempt to maximize power), and negative growth (falling levels of per capita real income) appears to reduce freedom in tinpot regimes (that just attempt to maintain power), as predicted by the Wintrobe theory. On the other hand, positive growth in tinpots

and negative growth in totalitarians also reduces freedom, contrary to the theory, although in the case of tinpots, the absolute value of the effect on the index of freedom appears to be bigger for negative than for positive growth, as predicted by Wintrobe's model. Some results concerning differences across regimes in the effect of schooling on freedom are also provided.

New Work on Repression: Dynamics, Ideology and Genocide

The theory of repression has been extended by Philip Verwimp (2001), who attempts to understand the behavior of the Habyarimana regime in Rwanda, and in particular to explain the origins of the tragic genocide that took place there. The paper applies Wintrobe's model in a new way (by using the price of coffee as an index of the capacity of a dictatorial regime to generate loyalty) and it extends the model to explain genocide. Verwimp suggests that the Habyarimana regime, frustrated by its loss of power, attempted to split the population along ethnic lines and set one group against the other, culminating in rewarding Hutus for the extermination of Tutsis. Thus the genocide is interpreted as the attempt by the regime to remain in power by accentuating the ethnic split, splitting the population into two groups, ultimately singling out one for extermination by the other.

Spagat (2001) studies the optimal strategy for a dictator hanging onto power by choosing how much repression to apply in every period. State variables are the amount of "hate" and "fear" in society both of which increase with the amount of repression from the previous period. Hate, fear and a random shock determine the quantity of repression required for the dictator to survive. They show that in every period there are only two possible optimal choices: the minimal repression necessary to retain power ("no demonstration") or the maximum possible repression ("demonstration"). The state space can be divided into two regions separated by an increasing function such that "no demonstration" is optimal in one and "demonstration" in the other. Under some conditions the opportunity for international borrowing makes demonstration optimal when it would not have been so without this option.

Bernholz (2001) developed a model of the evolution of totalitarian regimes. In the model there are "believers" who are convinced that others have to be converted to the supreme values of their ideology for their well-being or are possibly, enemies of their creed whose presence is obnoxious to them. Believers spend resources on winning new converts and the secular power of the state. Whether they succeed in this endeavour depends on the costs of converting new believers and on the amount of resources they are prepared to spend for this purpose, given their incomes and their propensity to consume. Their chances of success are greater if a crisis occurs, an event which is usually outside of their control. Once secular power has been secured, the resources of the state can be used to win more converts, to drive into exile or to kill the inconvertibles and to try to reach the imperialistic aims implied by the ideology. If the latter is not the case, the regime may turn into a mature

"ideocracy" after having reached its domestic aims. This would for instance be the case, if all inconvertibles had been removed and all the other population been converted. In this case no further terror and (or) repressions characteristic of totalitarian regimes are required. If the ideology implies ambitious imperialistic aims, for instance the conversion of all people on earth (except for inconvertibles) or the domination of the whole globe by the believers, it is highly probable that these aims cannot be reached. As a consequence, either a war is lost leading to the removal of the totalitarian regime, or the ends have to be adapted to maintain the credibility of the ideology. But then the totalitarian state may again turn into a mature ideocracy, if the ideology has been reinterpreted to remove its unrealistic imperialistic aims. Or the change of the ideology weakens the regime in a way that it loses its proselytizing character altogether, and turns into an ordinary autocratic regime.

The Irony of Absolutist Monarchy

Another important analysis of the limits on the power of dictatorship is provided by the "irony of absolutism." The problem is described in a series of works by North, Weingast, Root and others (North, 1981; North and Weingast, 1989; Root, 1994). In North' (1981) model of the monarchy, the King maximizes revenue, and the central problem is that the structure of property rights which is appropriate for this purpose is not usually that which is efficient from the economic point of view. More subtly, there is a tradeoff between power and revenue. As Root describes the "Irony of Absolutism", absolute power gave the King the capacity to repudiate debts, but

> Creditors took into account the king's reputation for repudiating debts and therefore demanded higher interest rates than would otherwise have been needed to elicit loans. Actually, *because he was above the law, the king had to pay more for loanable funds than did his wealthy subjects*. In short, the Crown had a problem asserting its credit because it had a history of reneging on commitments. [Italics added.][7]

North and Weingast suggest that this problem gave rise to the Glorious Revolution in England, in which power over the Treasury was devolved on the Parliament. In this way the King could credibly commit to repay. No such devolution of power occurred in France. The result was that the English King solved the problem of how to raise funds and could finance his army and other expenditures while the French King did not, leading to the chronic shortage of revenue that was one of the factors leading to the French revolution.[8]

---

[7] Root (1994), p. 177.

[8] Note that the irony of absolutism is already incorporated into equation (21.3) above: it means that B_π,0, i.e., that an increase in the autocrat's power π reduces budgetary revenue B. Presumably this would be true at high levels of π.

Congleton (2002) extends this analysis by pointing out that all kings share power. He suggests a generalized template, "King and Council" for looking at these issues. In practice one rarely observes pure forms of dictatorship that lack a council, or pure forms of parliament that lack an executive. Generally government policies emerge from organizations that combine an executive branch of government, "the king," with a cabinet or a parliamentary branch, "the council." Congleton provides an explanation for this regularity: The bipolar "king and council" constitutional template has a number of properties which give it great practical efficiency as a method of information processing and collective choice. First, a council generally has a wider array of direct experience and/or knowledge than the king does, and therefore is in position to be a better estimator of "policy consequences" than the king alone tends to be. Second, a bipolar design can reduce losses from conflict in cases where significant power centers other than the king exist. Third, a king and council template which provides agenda control to the king, tends to reduce the extent to which majoritarian cycles may arise in the council. Fourth, the king and council templates allow gradual evolutionary shifts of power between the executive and parliament as circumstances change without the necessity of violent conflict. Insofar as a form of majority rule is used by the council and is stable, the recommendations of the council tend to be both robust as estimators and moderate in their policy recommendations.

Growth and Economic Efficiency Under Dictatorship

There has been a lot of research asking the question, which is better for the economy, democracy or dictatorship? The answer is complex, mainly because the economic systems under autocracies vary so much. Those who believe there is some simple formula for distinguishing the economy of dictatorship from that of democracy should compare, for example, the economies of Nazi Germany, Apartheid South Africa, Papa Doc's Haiti, Pinochet's Chile and the Former Soviet Union.

Democratic Inaction

One general proposition which is true of all these systems is that dictators have a greater capacity for *action*, good or bad. If a dictator wishes to raise taxes, declare war, or take tough measures vs. crime, he may have to deal with some opposition to these policies among his advisers, but by and large, he can do so. Democracies, on the other hand are often mired in *inaction*.[9] The basic reason for this is that

---

[9] Some formal conditions for the existence of equilibrium inaction, as well as the circumstances under which this is inefficient, are discussed in Howitt and Wintrobe (1995) and in Wintrobe (1998a, b), chapter 11.

democratic leaders can only act when they can build support for their policies and there may be no consensus as to what to do. Even on problems where there is agreement that something should be done, there may be no agreement on *what* should be done. In extreme cases, the political system of a democratic country may become paralyzed by conflicts or opposing viewpoints.[10] In these circumstances, politicians often prefer to do nothing, to shroud their positions in ambiguity, or to pretend to be on all sides of an issue. The result is that the population can become cynical, and lose trust in the promises of any politician. This can set in motion a downward spiral, since the more this happens and trust is lost, the harder it becomes for politicians to do something by forging a compromise. This is more likely to happen when the pressures for political action on an issue are particularly conflicting, when positions are far apart, when issues are particularly divisive, when the population is divided along racial or ethnic lines, and when there is relatively little trust in politicians by the citizens.

Economic Growth and Efficiency

Introduction

Some new theorizing and empirical work compares the economic performance of democracies and dictatorships directly. A convenient place to start is Barro's empirical work (1996a, b). Barro stresses the advantages of dictatorship, which are that it controls rent seeking and other redistributionist pressures, i.e., the autocrat, unlike the democratic politician, is capable of shutting down or simply ignoring the redistributionist demands of interest groups characteristic of democracy (Barro, 1996b, p. 2). His empirical work suggests that democracy raises growth more at low levels of political freedom but depresses growth when a moderate amount of freedom has already been attained. However, the effect of an increase in political freedom on economic growth is not very large and the overall effect is "not statistically different from zero" (Barro, 1996b, p. 6). Barro's results are only obtained once certain variables are held constant, including free markets, the rule of law, and small government consumption. So, really, again, only certain *kinds* of dictatorship are being discussed. The paper also finds, perhaps surprisingly, that democracy does not necessarily promote the rule of law.

Przeworski et al. (2000) find that basically there is no difference between the rates of growth in dictatorships and democracies in their comprehensive examination of the performance of these two kinds of regimes in 141 countries over the 40

---

[10] There is a large literature in political science which associates the historical breakdown of democracy in various countries with precisely these variables: inaction, lack of credibility, and their mutually reinforcing effects (for details and references see Wintrobe, 1998a, b, chapter 13).

years or so after the Second World War. But the same study confirms the importance of politics on economic growth. It shows that changes in office (political instability) and other forms of unrest such as strikes, demonstrations and riots reduce economic growth substantially under dictatorship, whereas while these are more frequent under democracy they do not cause a reduction in the rate of growth there (Przeworski et al. (2000) pp. 192–193).

Sen (1993) calls the general idea that dictatorship is better suited to economic development than democracy the Lee thesis, after Lee Kwan Yew, the autocratic but economic efficiency-minded ruler of Singapore for many years. Sen raises many questions about Lee's ideas and suggests instead that democracy is intrinsically important to the process of development. In particular, Sen's observation that famines only seem to occur under dictatorship is provocative. However, no general theoretical model is presented that compares democracy with dictatorship.

The Predatory State

The most prominent theoretical idea in this literature is undoubtedly Olson's concept of an autocrat as a "stationary bandit"—at one point he refers to it as "the other invisible hand"—that guides rulers to use their power at least to some extent in public interest. In his (2000) book, this concept is approached through a criminal metaphor. Each theft reduces the wealth of society and therefore the amount available for the thief to steal. Does this lead the thief to curtail his activity, in order to preserve the wealth of his prey? For the typical criminal, the answer is "no" because his interest is too narrow. The wealth of the society on which he preys is like a public good to the typical small scale criminal, his effort to preserve it would have only a minuscule effect, and so he is better off free riding rather than attempting to conserve it. On the other hand, the Mafia and other criminal organizations which have a monopoly on crime in their jurisdictions do have a sufficiently *encompassing* interest to take the effects of their thefts on the wealth of society as a whole. Thus, Olson asserts, they typically do not steal at all but engage in protection instead, charging the citizens a fee to ensure the safety of their victims both from others and from the protectors themselves.

This criminal metaphor then becomes the foundation for the origins of government. The logic is the same as that just outlined with respect to government by a "roving" vs. that by a "stationary" bandit: the stationary bandit, unlike the roving one, has an encompassing interest in preserving the wealth of the society from which he steals, and therefore limits his "theft" (taxes) and even provides public goods—both to the point where the marginal benefit to him is sufficient to account for his costs in terms of foregone income. The history of the forms of government is then simple to derive: autocracy (the stationary bandit) arises out of anarchy as the bandit(s) with the greatest capacity for violence takes over the area and substitutes an encompassing interest for a narrow one; democracy arises out of dictatorship when autocracy is overthrown and none of the individuals or leaders involved in the coup has sufficient power to make themselves autocrats.

In the end, just two variables are necessary to compare and analyze governments:

(1) how encompassing (breadth of self interest) is the interest of the ruler,
(2) how long (time horizon) is his interest.

Thus, in the same way that dictatorship is superior to anarchy because the dictator has an encompassing interest in the society he rules, so democracy is superior to dictatorship because democratic majorities are more encompassing than the interest of the dictator. Secondly, dictators or democracies with long time horizons have more of an interest in preserving or enhancing the wealth of the society they rule than those who rule only for the short term.

Some evidence is presented in Keefer et al. (1996), who argue that any incentive an autocrat has to respect property rights comes from his interest in future tax collections and national income and increases with his planning horizon. They find an empirical relationship between property and contract rights and an autocrat's time in power.

So, comparing dictatorships, the basic implication is that, the more encompassing it is, the better. Political scientists indeed have a classification that appears to match this: between "mobilizational" regimes which encourage political participation among the ruled and regimes which simply try to stamp out opposition. The problem with Olson's analysis is that, comparing dictatorships, the worst regimes in human history appear to be precisely those such as Nazi Germany, Soviet Russia, or Cambodia which appear to have been the most encompassing. The reason is simple: it was those regimes that wanted to remold the citizens and the societies under their rule and therefore intervened most dramatically and thoroughly into the lives of their citizens. Whether it is their brutal treatment of minorities or their record on the environment, it is an understatement to suggest that the historical record of these regimes offers little that is to be admired. So the theory appears to be capable, not just of misleading with respect to the understanding of autocratic regimes, but of "getting it wrong" in a spectacular fashion.

The same problem appears with respect to the second variable, the time horizon of the dictator. In Olson's model, the longer the time horizon, the better, i.e., the more the dictator tends to rule in the social interest. But regimes with a long time horizon have been precisely those in which the leaders had a tighter grip on power, and hence were more capable of molding the society and the individuals within it, i.e., the mobilizational regimes just discussed. Those where the regime is just interested in looting the society typically have a shorter time horizon.

In short, from the point of view of citizens of these regimes, or more specifically from that of the peasants under Stalin, the Jews under Hitler, the blacks in South Africa, and so on, it would no doubt have been better if their bandits had been less stationary!

The alleged superiority of dictatorship over anarchy is also challenged in a major article by Moselle and Polak (2001). In their model, the existence of a state can result in lower levels of both output and welfare than would occur under anarchy. This occurs if the state is "predatory" in the sense that the rulers extract taxes from

the population for their own ends. In this framework, even a weak state can be bad for output and welfare and a "corrupt" state that makes side deals with bandits can be especially bad.

Perhaps the most basic problem with Olson's framework is, I suspect, the lack of emphasis on competition. Once the struggle for power is assumed away, many of the most interesting aspects of the behavior of dictators become idiosyncratic features of their preferences, and hence largely unpredictable, instead of being derived from the principle of competition. Thus the wars among the monarchies, etc are all aspects of "princely consumption". And how would the model explain Stalin's war against the peasantry, Hitler's treatment of the Jews, and the persecution of minorities in other dictatorships? On the bandit model, the only way to understand these forms of behavior is that dictators have some monopoly power, and that they use this power to implement their preferences which happen to be weird preferences. The reason for this is that the model does not deal with the competitive struggle to acquire and maintain dictatorial powers. So the behavior of the dictator cannot be understood as motivated by competition or survival in office but simply as consumption.

Two other contributions address the problem of why some dictatorships, most notably regimes in East Asia and Chile, appear to be pro-growth while in others the autocrat is "predatory" and simply plunders the economy. Robinson (1997) argues that the likelihood of predatory behavior may be positively related to the extent to which a regime is encompassing and values the future. He develops a model in which whether or not a state is predatory hinges on the relationship between development and the distribution of political power. Development is typically inconsistent with the preservation of the political status quo and this gives those in power an incentive to oppose it. Also, the lower the level of income and the more unequal the society, predatory behavior is more likely. To put it bluntly, from the Dictator's point of view, ruining the economy can sometimes be a good thing!And the regimes of Mobutu and Papa Doc, who both did this, were extremely long lived. A democratic politician cannot hope to profit in the same way.

Michael Spagat's (2001) paper addresses this problem by suggesting that there is a "bifurcation point" or level of capital below which it does not pay the dictator to try and develop the economy, and above which the dictator pursues rapid growth in order to maximize his personal consumption over time. He develops this idea in a simple formal model. A particularly novel feature of it is that there is an endogenous probability of a political catastrophe which removes the dictator from power, and this in turn depends on the Dictator's capacity to satisfy certain groups which depends on the level of the capital stock. Hence a Dictator's economy sometimes grows faster than a social planner's might, as capital accumulation wards off the possibility of catastrophe. The authors use simulation analysis to show the existence of bifurcation and to show how it depends on various parameters, and they provide some empirical evidence using the Gastil data of the existence of bifurcation, and of their basic prediction that the variance of growth rates in dictatorship is higher than that under democratic regimes.

The Contest for Power

In contrast to economic models which stress the incentives of a ruler, *once he is in office*, Wintrobe (2002) focuses on the conditions under which the ruler obtains power, and how he can be deprived of it. All political systems contain a mechanism which determines the allocation of political power, and if and how it is reallocated when a transfer would improve the functioning of the system. Among the most obvious and commonly considered types of political system—democracy, dictatorship, anarchy, and hereditary monarchy—*only* democracy appears to possess a relatively low-cost procedure or mechanism which makes it possible to transfer political power on a regular and systematic basis, where the transfer is accepted by those who lose power as well as by those who gain it, and which offers some possibility that these reallocations will tend to shift power into the hands that can use it most effectively.

Thus there is a strict analogy between democracy, based on human rights, and capitalism, based on property rights: democracy makes power transferable just as capitalism makes the ownership of capital assets transferable. This gives democracy an enormous advantage over these other political systems.

To elaborate, the main economic advantage of the election mechanism would seem to be that it allows for the transfer of power at a relatively low cost. It solves the *contest for power* problem. If there are no elections, the only ways to transfer power are by such means as revolutions, insurrections, coups and wars. Compared to these, democratic elections on the basis of inalienable human rights would seem to be, in a word, cheap. Thus the economic attractiveness of the election mechanism is simple: it provides a formal and agreed-upon procedure to decide on the allocation of political power, and one that is explicitly accepted by or consented to in advance by the parties who lose the contest. Among the most commonly discussed systems of governments, anarchy, hereditary monarchy, dictatorship and democracy, only democracy possesses this advantage.

Granted that democracies can transfer power at relatively low cost: does power typically transfer from uses of lower value to those of a higher value? Do democracies allocate power properly? In the models of Stigler (1971); Peltzman (1976); Olson (1982), democracy is inefficient because it is dominated by interest groups and the policies pursued by interest groups are inefficient and wasteful. In the rent-seeking framework, it is the contest itself which is inefficient and wasteful. These theories are the foundations of Barro's (1996) empirical work. However, Becker (1983) showed that under democracy the losses from inefficient policies enter into the workings of the political system and affect its allocation of power. The reason is that the larger the deadweight losses from a policy, the more opposition there will be to it from the groups which bear these losses. Alternatively the more inefficient a subsidy, the less the group which gains from the subsidy will exert pressure to obtain it. Consequently, even in a model such as Becker's, which focuses solely on interest group competition, the contest is not wasteful, and it tends to select efficient

over inefficient policies. To put it simply, if power ends up in the "wrong" hands the democratic political process takes account of this and tends to set it right.[11]

What about dictatorship? The basic difference between dictatorship and democracy is that dictators have the capacity to repress opposition to their policies (as outlined above). They can silence demonstrations, censor the media, ban opposition parties, put leaders of troublesome groups in jail, and, not uncommonly, torture or execute them. As a consequence, the repressed are, in effect, not allowed to spend resources to exert political pressure: instead they are silenced by the government. It follows that if the costs of public policies can be made to fall on those who are repressed, these costs do not enter into the competition among interest groups.[12]

Assume for a moment that this is the case, i.e., all of the costs of inefficient policies fall on those whose political demands are effectively repressed by the regime. Since the losses make the repressed worse off, this weakens the capacity of those who are opposed to the regime to resist it.[13] This is the strategy of "immiserization" practiced most notably, perhaps, by Papa Doc of Haiti.[14] Another nice illustration of this is the effect of sanctions against Saddam Hussein, discussed by Kaempfer et al. (2001). The sanctions generate rents, and these are appropriated by those who are close to Saddam. The losses from the sanctions are borne by those who are opposed to the regime, and this in turn weakens their capacity to oppose it, leading to his further entrenchment in power. To put it simply, the sanctions against Saddam Hussein don't necessarily weaken his hold on power at all.

On the other hand, to the extent that the repressed cannot be made to bear all of the costs of the public sector, some of these costs will fall on other groups—actual supporters, potential supporters and largely passive acceptors of the regime. To the extent that the costs of public expenditures and regulations fall on these groups, they would indeed enter into the competition among groups for subsidies and other rents from the regime under dictatorship, just as they do under democracy. However, even in this case, the mechanism does not work as well as under democracy. The reasons are: (1) The information problem deriving from the Dictator's Dilemma: In a democracy the different groups competing for redistributionist policies or public goods are free to debate openly, to criticize existing policies and to expose flaws in each others' proposals. Under dictatorship, any form of attack on policies

---

[11] Newer, dynamic models of democratic decision-making cast doubts on the efficiency of democracy in a dynamic context. The basic problem discussed there (e.g., in Besley and Coate, 1998) is the inability of a representative democracy to commit to future policy outcomes. The question from the point of view of this survey is of course, whether a dictatorship could be expected to do better in this respect. To my knowledge nothing has been written on this issue but it is worth noting the evidence in Przeworski et al. that the average life of a dictatorship is less than democracy.

[12] Thus in Becker's model, equation (13) would not hold for a repressed group since the group cannot spend resources to pressurize the government; neither would equation (14), in which each person maximizes his income from producing pressure.

[13] Contrast this proposition with Becker's point that under democracy larger deadweight losses increase pressure from the group experiencing them to lobby against the policies.

[14] For more details, see Wintrobe (1998a, b).

which have been or might be favored by the regime can be interpreted as a sign of disloyalty,[15] and for this reason people may not be eager to report problems to the autocrat. (2) What incentive is there for the dictator to correct bad policies? After all, among the fruits of dictatorship is "the quiet life"—freedom from competitive pressures as long as he is safe in office. (3) Finally, once decisions are made there may be no mechanism by which to correct them except by the overthrow of the dictator. Since there is no peaceful and regularized way to replace an autocrat, he may tend to oppose any attempt to change the policies, since any change may be threatening to his survival in office.

To sum up, the economic losses from inefficiency may or may not enter into the Dictator's political budget equation, depending on who experiences them. Let us take the two cases in turn: (1) Suppose that the losses are experienced by actual or potential supporters. The lack of political competition under dictatorship still implies that the economy may be allowed to deteriorate more (compared to a democracy) before some attempt is made to change the policies or replace the dictator; (2) On the other hand, suppose the economic losses are experienced primarily or wholly by those who are opposed to and repressed by the system. In this case the losses typically weaken rather than strengthen the capacity of those who are opposed to the regime to actually topple it, and this raises the attractiveness of inefficient policies from the point of view of the dictator.

A final issue is the relative influence of producer vs. consumer groups under dictatorship and under democracy. Ever since the work of Downs, it has been a standard proposition in the economics of politics that democracy favors producer groups over consumer groups (Downs, 1957; Stigler, 1971; Peltzman, 1976; Olson, 1982; Becker, 1983). The main reasons advanced are that since these groups are small, it is relatively easy for them to overcome the free rider problem, and since their *per capita* benefits from any subsidy would be large, they have a substantial interest in applying pressure to obtain it. On the other hand, consumer groups are large, and the *per capita* benefit from any subsidy would be small.

I have pointed out earlier that dictators cannot survive in office on the basis of repression alone but need support as well. Which groups can be expected to support dictators? Consumer groups, environmental groups and other groups with a large number of potential supporters, each of which has a small stake in issues like the prices of goods or the state of the environment have difficulty surviving or forming under autocracy. There are typically no laws protecting human rights under dictatorship. Without such laws, it is difficult for large groups—such as consumers—to organize. There is no free press to call attention to pricing or environmental or labor abuses and to aid in the formation of a mass membership and there are no independent courts in which to sue violators. And it is difficult for supporters of human rights, who have been crucial in generating the "rights revolution"

---

[15] As an illustration, Mao Tse Tung's personal physician, Li Zhisui, appeared to be afraid to criticize Mao even on the smallest of matters. See Zhisui (1994). Other illustrations of this point for various regimes can be found in Wintrobe (1998).

(Epp, 1998; Ignatieff, 2000) to mobilize support. In brief, the common weapons of mass organizations—publicity and the courts—are more easily countered by a dictator than a democratic politician.

On the other hand, the weapons of small producer groups such as cash donations actually thrive in the closed environment and tame courts of a dictatorship. In exchange, dictators obviously have much to offer producers for their support including tariffs, subsidies and other rents, fewer problems from labor unions, and the removal of unfavorable regulations. So the possibilities of trades of rents for support between the dictator and small, concentrated interest groups is actually *enhanced* under dictatorship, just as trades with representatives of broader public opinion are diminished. This implies that *producers typically have more power under dictatorship than in a democracy*.

This analysis also provides an alternative explanation for Barro's evidence cited above: that the rate of growth is slightly higher under dictatorship than democracy at low levels of dictatorship and lower at high levels of repression. Since producers especially benefit from economic growth, their greater political weight under dictatorship implies that dictators would emphasize this policy. Note, however, that this growth comes as the result of the greater influence of producer groups and is not necessarily a Pareto improvement. Thus the growth could arise to the detriment of the environment, the consumer, etc. Moreover, at high levels of repression, this positive effect on growth is increasingly overwhelmed by the information problems generated by the Dictator's Dilemma, which increasingly hamper growth and ultimately strangle it.

Finally it is worth pointing out that an extension of the theory of property rights used in this analysis provides a simple economic justification of human rights. Economic efficiency justifies the ownership of private property on the ground that property should be allocated to the party who is most highly motivated to maximize its value. Who can be counted on to manage or to take care of a piece of property best? The owner. Human rights give this privilege of "ownership" of the individual (if you like, of his human capital) to that individual himself or herself. Under dictatorship, it resides with the sovereign. But the dictator, as Sen (1993) suggests, tends to regard the people under his rule as "stock" and cannot be expected to care for their lives the way they would themselves. Perhaps this explains Przeworski et al.'s striking result that the average life span is systematically lower under dictatorship (Przeworski et al., 2000, Chap. 5).

Policy towards Dictatorship

Aid: A Single Standard

What policies should be followed towards dictatorships by democratic regimes interested in promoting freedom? Suppose, idealistically, that the only goal of Western policy is to reduce repression. The "weapons" in our arsenal are sanctions, trade

agreements, imposing human rights constraints, and aid packages. Take a classic example of a tinpot dictator like Ferdinand Marcos. Should we have given aid to his regime? Suppose Marcos' only goal was to consume as much as possible—in his case, this meant buying shoes for his wife Imelda. What limited his consumption? Why didn't he spend all of the GNP of the Philippines on shoes for her? The constraint is that he had to stay in office, so he could not allow his power to fall so low that he was in danger of being deposed. As a tinpot, the levels of both repression and loyalty under his regime were just high enough to stay in office.

Suppose first that the tinpot is safely in office, which, at one point, according to accounts of the regime, Marcos felt he was. Then there is no point in giving him aid, because all he will do with the money is to buy more shoes. A trade agreement would have the same effect. On the other hand, suppose he is in danger of being deposed. Then the aid simply props up the regime. So, *in neither case does aid reduce repression*. An alternative policy would be to insist on human rights observances as a condition of receiving aid. But if the levels of repression and loyalty were previously just sufficient to stay in office, Marcos will simply be deposed if he lowers repression. So he would have refused this offer, and the policy is ineffectual.

On the other hand, suppose that aid is tied to human rights observances in a particular way. In order to keep receiving the aid, repression must be steadily relaxed over time. Then the dictator has an incentive to use the aid to improve the welfare of his people. The reason is that, if their welfare improves, and he can claim credit for this, loyalty or support for him will tend to increase. As a result, he can afford to relax repression, and still buy the same number of shoes for Imelda as before.

Now look at totalitarian regimes or tyrannies, defined as regimes whose rulers are interested not in consumption, but in power. Should we aid them? Again, suppose that, as the result of either policy, economic growth improves. This gives the rulers an opportunity to accumulate more power, and since power is the only thing they care about, they take this opportunity, in the same way that a businessman who is already rich will grab an opportunity to make more money. So, for these regimes, aid which is untied to human rights observances is not merely wasted, but counterproductive—repression *increases* when the economy improves. This is what happened under Hitler and Stalin: the more popular they were, the more they took these opportunities to put the screws to all those elements of the population whose absolute loyalty was uncertain. In the same way, the enormous economic growth in China has resulted in not the slightest degree of relaxation in the level of repression there.

It might seem obvious that we should not aid these regimes, since the aid money would be spent on accumulating more power over the population, including repressing them. But, again, if aid is tied to a human rights constraint, which becomes progressively more stringent over time, the policy will work in the right direction. If the economy improves as a result, support increases, and the rulers can afford to relax repression and still have the same level of power as before. The human rights observances constraint is absolutely necessary if this is to lead to a fall in repression and not an increase.

So we have a very simple guide—a *single standard*—to the policies which should be pursued by foreign governments interested in reducing repression. This is to make human rights observance the cornerstone of Western policy. Aid to any type of regime can be expected to produce beneficial effects provided it is accompanied by a long term human rights constraint, one which becomes progressively more stringent over time. Without the human rights standard, the effects of aid will be ineffectual or perverse.

Trade

Another policy dilemma is whether to trade with dictatorships. Trade policy is a bit more complicated than aid. We can distinguish the following effects:

1. Trade may be expected to increase national income of the target regime, as productivity there will rise due to the availability of imported inputs at a lower price, and the demand for the target' exports increase. To the extent that the regime can successfully claim the credit for this improvement in welfare, loyalty to the regime may be expected to increase.
2. The rise in income will also increase tax revenues, giving the dictator more resources at his disposal. These may be used either for his own consumption, or to further his hold on power through increased expenditures on either repression or loyalty.
3. Since, the richer the people are, the more they tend to demand liberty, the increase in income tends to reduce loyalty to the dictatorship, as people increasingly demand their rights (Bilson, 1982; Londregan and Poole, 1996). However, note that the estimated size of this effect is very small. Thus, as Londregan and Poole conclude their analysis of this effect in non-European countries, "Those expecting income growth to promote the development of democratic institutions must be very patient indeed" (pp. 22–23).
4. The increase in trade creates further links to foreign businesses and among domestic producers, possibly resulting in the development of independent power bases within the target regime. This is particularly likely when the trade is not organized through the central government (as it is in Cuba, for example). Thus, in China, regional governments in particular have built up substantial connections with outsiders and with the private sector, and are much more independent of the central government for revenue than they were before Deng launched his "social market" revolution. To the extent that this happens, loyalty to the regime may fall. On the other hand, it has been argued that trade between different types of civilizations actually increases mistrust, as the increased intensity of contacts simply breeds hostility. For example, World War I occurred at precisely the last peak of the "openness" of the international system. In that case, while there may be a short-run fall in loyalty due to the initial increase in contacts, in the longer run, further contacts simply breed nationalism and possibly *increased* support for the dictatorship in the target regime.

To disentangle the implications for policy, suppose first that the net effect of these changes, as seems likely, is that support for the regime increases as the result of the trade agreement. Suppose also that the ruler is a tinpot. Then it can be argued that, with increased support, the tinpot will be himself motivated to relax repression (so that he can buy more shoes for his wife), and there is no need for a human rights constraint. But note that, even in this case, the human rights constraint does no harm; it simply asks the dictator to do what he would do in any case, and therefore it should be acceptable to him. On the other hand, if, on the balance, loyalty to the regime were to decrease, the tinpot would want to raise repression in order to stay in office, and the human rights constraint is absolutely necessary for the trade agreement to lower, not raise, repression.

Suppose now that we are dealing with a totalitarian dictator. Again, if loyalty were to increase, on the balance, as the result of the trade agreement, the dictator would tend to raise repression, and the binding human rights constraint is necessary to prevent a loss of freedom. The only case for a trade agreement with a totalitarian regime is where the opposite happens, and loyalty to the regime decreases from the trade agreement. In that case, repression falls as well. *This is the only case where trade with a totalitarian regime makes sense*. But note that, the totalitarian leader, in pursuing this trade agreement, cannot fail to be aware of the likely consequences of the trade agreement for his hold on power; namely, that his capacity for repression, the loyalty to him of the citizenry, and his power are all going to diminish as a result of his signing up. So, if this analysis were correct, it requires us to believe that the totalitarian is either unaware of, or is deliberately acting contrary to, his own interest in the long run.[16] It is noteworthy also that all the totalitarian regimes which have collapsed historically did so as the result of falling, not rising real income, and that the increase in income in China has resulted in not the slightest relaxation of repression there after almost two decades of reform and spectacular economic growth. The case for trade with totalitarian regimes, therefore is particularly weak.[17]

Finally, suppose that the human rights constraint cannot be implemented, either because the target regime is too powerful, or because no agreement can be reached among the countries involved in implementing the policy. Or alternatively, suppose the dictator promises to abide by the human rights constraint and then reneges. Then there is a difficult choice between a policy of sanctions, on the one hand, and trade agreements with no effective human rights constraint, on the other. Of course, the actual choices are never this stark, and the actual policies followed will be a mixture of trade and sanctions, but the basic principle involved in the choice remains one of engagement or isolation. In that case, the analysis here implies that the least harm is likely to come from a trade agreement with a tinpot regime, the most harm from trade with a totalitarian, with tyranny as an intermediate case.

---

[16] Note that the situation is very different for a tinpot, for whom the relaxation of repression following a trade agreement serves his interest, rather than acting contrary to it, as is the case for a totalitarian.

[17] The analysis of the effects of trade on tyranny is identical to that for totalitarian regimes; the only difference is that the magnitude of the change in the supply of loyalty is smaller.

Sanctions

Historically, the most important alternative to a policy of aid to motivate dictator-ships to behave is to use sanctions to punish those that do not. However, it is vital to realize that sanctions are not just the reverse of aid, and that policies like those pursued by the United States and the United Nations vis-à-vis regimes like Castro's Cuba, Saddam Hussein's Iraq or Milosevic's Serbia may superficially resemble those described here, but in fact they work very differently. In all these cases, the U.S. or the UN imposed sanctions, and offered to lift them as a reward for better behavior. Such policies are not necessarily wrongheaded, but they do not work in the manner of those advocated here. The reason is that the sequence is reversed: the regime has to liberalize *first*, i.e., before the sanctions are lifted, trade allowed to resume and aid to flow. This means that the regime has no chance to use the benefit of aid or trade to build loyalty prior to liberalization, as with the policies advocated here. So the dictator who agrees to liberalize puts himself in immediate danger of being deposed, and it is no surprise that dictators like Castro, Hussein and Milosevic were all reluctant to do so.

Kaempfer et al. (2001) extend Wintrobe's (1998a, b) model of dictatorship and combine it with the public choice analysis of sanctions. They note that 85% of the time that sanctions are imposed they are imposed on a non-democratic regime. They point out that damaging economic sanctions can be counterproductive, under-mining the political influence of the opposition. In the public choice approach, sanctions work through their impact on the relative power of interest groups in the target country. An important implication of this approach is that sanctions only work if there is a relatively well organized interest group whose political effective-ness can be enhanced as a consequence of the sanctions. For example, as the authors note, sanctions against Iraq have had a devastating effect on the country but have been ineffective in destabilizing the Hussein regime. The reason, they argue, is the fragmentation of the Iraqi opposition. At the other extreme, sanctions against South Africa were highly effective, because, in that case, there was a well organized opposition. The authors suggest that the effectiveness of the opposition is key to the effectiveness of sanctions and they try to show why this is true and to derive implications from this insight.

They also extend the model by adding two exogenous variables to it: s, the impact of sanctions on the terms of trade; and q, the level of opposition; moreover q depends on s, $q_s \cdot 0$, and by making the price of repression P_R (constant in Wint-robe's model) a variable which depends on q and s (in addition to their other effects on the model) as well as on the country's economic performance.

In their model, sanctions have two main and opposing effects on the Dictator's budget: (1) the budget of the dictator increases through the appropriation of sanc-tions' rents; (2) the budget falls due to the increase in opposition. There are two cases. In the first case, the opposition is significant enough that $q_s \cdot 0$. If, in addition, the second effect is large enough, the budget falls. If in addition, loyalty to the dictator falls due to the sanctions then sanctions are effective.

In the second case, there is no significant opposition. Then the net effect on the Dictator's budget of sanctions is that it *rises* due to the appropriation of the sanctions' rents. If in addition loyalty *rises* because those close to the dictator are happy about their increased capacity to appropriate these rents, then the sanctions are entirely counterproductive, and the budget of the regime, its power, and the level of repression all increase.

Conclusion

In recent years a small but now growing literature has looked at dictatorship from the point of view of public choice. While there is no consensus in the literature and it would be too soon to look for one, a number of ideas are attracting interest. The literature looks at (1) the objectives of dictators; (2) the constraints on their behavior; (3) their strategies for staying in office; (4) their incentives to provide public goods compared to that under democracy; (5) the economic efficiency of dictatorship compared to democracy; and (6) the policy towards dictatorships. On the *objectives* of dictators, some models simply assume that dictators maximize lifetime consumption, as in standard economic models. However although this is certainly true of small scale dictators, it hardly fits many of the most important dictatorships like those of Hitler or Stalin and other models explicitly posit a taste for power or, in common with many models of democracy, ideological objectives. Since dictators are by definition monopolistic, the case for including such other objectives is particularly strong. Wintrobe looks at the *strategies* used by dictators to stay in office and emphasizes the Dictator's Dilemma—the tradeoff between using repression and building support, noting that only the latter provides a firm foundation for an autocratic rule. Recent contributions extend this framework to consider dynamic models of repression, the issue of genocide and the efficacy of sanctions against dictatorships.

The *constraint* on autocratic maximization is sometimes specified as the maximum revenue available. Other models specify the so-called "irony of absolutism" as the chief limit to a Dictator's power. Wintrobe reasons that as long as more power is available there are ways to extract more revenue from the private sector. Similarly, as long as more revenue is available, it is possible to accumulate more power: For these reasons in his model the equilibrium of power and budget of the dictator are determined simultaneously.

The model of the dictator as "stationary bandit" introduced by Mancur Olson shows that even a dictator has an *incentive to provide public goods* in order to raise revenue though arguing that this incentive is less under dictatorship than under democracy.

There is, as always, disagreement about the *economic efficiency* of dictatorship vs. democracy; but the disagreement appears to be narrowing. On both theoretical and empirical grounds there appears to be a consensus that high levels of repression are inimical to economic efficiency. Empirical work by Barro and others provides some (very slim, as acknowledged) evidence that growth rates are higher under dictatorship at low levels of repression though even this is challenged in a major

empirical study by Przeworski and others. Theoretically, such a result can be explained by a reduction of rent seeking or redistributory pressures, as Barro does, or by the greater influence of producer groups under dictatorship, in which case the growth might come about at the expense of consumers or workers and need not signal greater efficiency. The "contest for power" framework, on the other hand, emphasizes that only democracy provides a mechanism for getting incompetent or corrupt rulers out and suggests that democracy might be more economically efficient than dictatorship when this factor is taken into account.

On policy the Wintrobe model provides a simple guide to the policies that should be pursued by foreign governments interested in reducing repression. This is to implement human rights observances, which become more stringent over time, as a condition for receiving aid or trade. One problem with this is that the dictator may promise to abide by human rights and then renege. The literature continues the standard skepticism of economists on the effectiveness of economic sanctions as a tool for getting dictators to lower repression.

References

Barro, R (1996a). "Democracy and growth." *Journal of Economic Growth*, 1, 1–27.

Barro, R. (1996b). *Getting It Right*. Boston: MIT Press.

Becker, G. (1983). "Theory of competition among interest groups for political influence." *Quarterly Journal of Economics*, 98, 371–400.

Besley, T. and S. Coate (1998). "Sources of inefficiency in a representative democracy: a dynamic analysis." *American Economic Review*, 88, 139–156.

Bernholz, P. (2001). "Ideocracy and totalitarianism: a formal analysis incorporating ideology." *Public Choice*, 108(1–2), 33–75.

Bilson, J. (1982). "Civil liberties—an econometric investigation." *Kyklos*, 35, 94–114.

Congleton, R. (2002). "From dictatorship to democracy without revolution," paper delivered at the American Economic Association meetings, Atlanta.

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Epp, C. (1998). *The Rights Revolution*. Chicago, IL: University of Chicago Press.

Friedrich, K. and Z. Brzezinski (1965). *Totalitarian Dictatorship and Autocracy*. Cambridge, MA: Harvard University Press.

Howitt, P. and R. Wintrobe (1995). "The political economy of inaction." *Journal of Public Economics*, 56, 329–353.

Ignatieff, M. (2000). *The Rights Revolution*. Toronto: Anansi Press for the CBC.

Islam, M. and S. L. Winer (2001). "Tinpots, totalitarians (and democrats): an empirical investigation of the effects of economic growth on civil liberties and political rights," ms., (Working Paper) Carleton University, Ottawa, Canada.

Kaempfer, W., A. Lowenberg, and W. Mertens (2001). "International economic sanctions against a dictator," paper delivered at the Public Choice Society Meetings, San Antonio, Texas.

Keefer, P., C. Clague, S. Knack, and M. Olson (1996). "Property and contract rights under democracy and dictatorship." *The Journal of Economic Growth*, 1(2), 243–276.

Londregan, J. and K. T. Poole (1996). "Does high income produce democracy"? *World Politics*, 49, 1–30.

Moselle, B. and B. Polak (2001). "A model of a predatory state." *Journal of Law, Economics and Organization*, 17, 1–33.

North, D. (1981). *Structure and Change in Economic History*. New York: Norton.

North, D. and B. Weingast (1989). "Constitutions and commitment: the evolution of institutions governing public choice in seventeenth century England." *Journal of Economic History*, 49, 808–832.

Olson, M. (1982). *The Rise and Decline of Nations*. New Haven: Yale University Press.

Olson, M. (2000). *Power and Prosperity: Outgrowing Communist and Capitalist Dictators*. New York: Basic Books.

Peltzman, S. (1976). "Towards a more general theory of regulation." *Journal of Law and Economics*, 19, 211–240.

Przeworski, A., M. E. Alvarez, J. A. Cheibub, and F. Limongi (2000). *Democracy and Development: Political Institutions and Well-Being in the World 1950–1990*. New York: Cambridge University Press.

Robinson, J. (1997). "When is a state predatory?", ms., (Working Paper) University of Southern California.

Root, H. (1994). *The Foundation of Privilege: Political Foundations of Markets in Old Regime France and England*. Berkeley: University of California Press.

Sen, A. (1993). "Political Rights and Economic Needs," The John M. Olin Lecture in Law and Economics at the University of Toronto Law School, October 1.

Spagat, M. (2001). "Political instability and growth in dictatorships," ms., (Working Paper) Royal Holloway College, England.

Stigler, G. J. (1971). "The Theory of Economic Regulation" *Bell Journal of Economics and Management Science* 2, 137–146.

Tullock, G. (1987). *Autocracy*. Dordrecht: Martinus Nijihoff.

Verwimp, P. (2001). "The political economy of coffee and dictatorship in Rwanda," ms., (Working Paper) Department of Economics, Leuven, Belgium.

Wintrobe, R. (1990). "The tinpot and the totalitarian: an economic theory of dictatorship." *American Political Science Review*, 84, 849–872.

Wintrobe, R. (1998a). *The Political Economy of Dictatorship*. New York: Cambridge University Press.

Wintrobe, R. (1998b). "Some lessons on the efficiency of democracy from a study of dictatorship," in S. Borner and M. Paldam (eds.) *The Political Dimension of Economic Growth*. McMillan: International Economic Association.

Wintrobe, R. (2002). "The contest for power: property rights, human rights, and economic efficiency," paper presented at the American Economic Association meetings, Atlanta.

Wittman, D. (1983). "Parties as utility maximizers." *American Political Science Review*, 77, 142–157.

Zhisui, L. (1994). *The Private Life of Chairman Mao*. New York: Random House.

Part V
Applications

Chapter 22
Public Finance[1]

Harvey S. Rosen

Scope of the Field

Public Finance is the branch of economics that studies the taxing and spending activities of government. The term is something of a misnomer, because the fundamental issues are not financial (that is, relating to money). Rather, the key problems relate to the use of real resources. For this reason, some practitioners prefer the label *public sector economics* or simply *public economics*. Public finance encompasses both *positive* and *normative analysis*. Positive analysis deals with issues of cause and effect, for example, "If the government cuts the tax rate on gasoline, what will be the effect on gasoline consumption?" Normative analysis deals with ethical issues, for example, "Is it fairer to tax income or consumption?"

Modern public finance focuses on the microeconomic functions of government, how the government does and should affect the allocation of resources and the distribution of income. For the most part, the macroeconomic functions of government—the use of taxing, spending, and monetary policies to affect the overall level of unemployment and the price level—are covered in other fields.

Methodological Basis

Mainstream economic theory provides the framework for public finance. Indeed, it would not be unreasonable to view public finance as just an area of applied microeconomics. As is the case in other fields of economics, the normative framework of public finance is provided by *welfare economics*, the branch of economic theory concerned with the social desirability of alternative economic states.[2] Much of welfare economics focuses on the conditions under which the allocation of resources in an economy is *Pareto-efficient*, defined as an allocation such that the

---

[1] This chapter is a reprint of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 252–262.

[2] Bator (1957) provides a classic exposition of welfare economics.

371

only way to make one person better off is to make another person worse off. Pareto efficiency seems a reasonable normative criterion—if the allocation of resources is not Pareto efficient, it is "wasteful" in the sense that it is possible to make someone better off without hurting anybody else. A stunning result of welfare economics is that if two assumptions are satisfied, then an economy will achieve a Pareto-efficient allocation of resources without any government intervention. The assumptions are: 1) All producers and consumers act as perfect competitors; that is, no one has any market power. 2) A market exists for each and every commodity. In a way, this result formalizes an old insight: When it comes to providing goods and services, free enterprise systems are amazingly productive.

Suppose for the moment that these two assumptions are satisfied. Does the government have any role to play in the economy? Only a very small government that protects property rights and provides law and order would seem appropriate. However, even if an allocation of resources is Pareto-efficient, it may not be socially desirable. A society may be willing to trade some efficiency in return for a fairer distribution of resources among its members (although "fairer" may be hard to define). Hence, even if the economy is Pareto efficient, government intervention may be necessary to achieve a fair distribution of real income.

Furthermore, real world economies may not satisfy the two assumptions required for Pareto efficiency. The first assumption is violated when firms have market power and raise their prices above competitive levels. Monopoly is an extreme example. The issues associated with market power are generally dealt with in the field of Industrial Organization, not Public Finance. The second assumption is violated when markets for certain commodities do not emerge. After all, if a market for a commodity does not exist, then we can hardly expect the market to allocate it efficiently. For example, there is no market for clean air. In effect, individuals can use up clean air (that is, pollute) at a zero price. That particular resource is not used efficiently.

Nonexistence of markets occurs in a variety of situations; each one opens potential opportunities for the government to intervene and improve welfare. In effect, then, the list of market failures provides the public finance agenda.

Public Expenditure

The theory of welfare economics focuses our attention on market failure and distributional considerations as reasons for considering governmental intervention. This section illustrates these issues.

Public Goods

A *public good* has two characteristics. First, once it is provided, the additional cost of another person consuming the good is zero—consumption is *nonrival*. Second, preventing anyone from consuming the good is either very expensive or impossible—

consumption is *nonexcludable*. A classic example of a public good is national defense. One person's consumption of the services provided by the army does nothing to diminish another person's consumption of the same services. Further, excluding any particular person from the benefits of national defense is all but impossible. In contrast, a private good (such as food) is both rival and excludable.

To see why the market may not provide public goods in efficient amounts, note that, for private goods, the market in effect forces each person to reveal what his true preferences are. If the value of the commodity to a person is greater than or equal to the market price, he buys it; otherwise not. There is no incentive to hide one's true preferences. In contrast, people have incentives to hide their true preferences for public goods. Each person knows that once national defense is provided, he can enjoy its services, whether he pays for them or not. Therefore, he may claim that defense means nothing to him, hoping that he can get a "free ride" after other people pay for it. Everyone has the same incentive, so that defense may not be funded, even though it is in fact beneficial. In short, the market cannot be relied upon to provide a public good in efficient amounts; some kind of collective decision making process may be better (Samuelson, 1954).

While important, this finding does not provide a firm set of guidelines for deciding when the government rather than the private sector should provide some commodity. The result depends in part on whether the public and private sectors pay different amounts for labor and materials, the extent to which the government can address the diversity of tastes for the commodity among the citizenry, and whether or not government provision will have a more favorable (somehow defined) impact on the distribution of real income. Whether public or private provision is better must be decided on a case by case basis. The fact that this can be difficult is reflected in the ongoing political debates in many countries about the merits of *privatization*—taking services that are supplied by the government and turning them over to the private sector.

Externalities

When the activity of one entity (a person or firm) directly affects the welfare of another, in a way that is outside the market mechanism, the effect is called an externality. The classic example is a polluter, who imposes losses on other individuals by degrading the environment. In general, efficiency requires that individuals pay a price for any commodity that reflects its value in alternative uses. But there is no market for (say) clean air. Individuals treat it as if its price is zero, and hence use it in inefficiently large amounts.

There are a number of ways in which government intervention can potentially enhance efficiency in the presence of an externality. 1) It can levy a tax on the externality producing activity. Basically, the tax makes up for the fact that the price being faced by the polluter is too low. 2) It can create a market for the right to pollute. Recall that the fundamental problem is that there is no market for the resource being polluted. In some cases, the government can create such a market.

The government announces it will sell permits to spew a given quantity of some pollutant into the environment. Firms bid for the rights to own these permissions to pollute, and the permissions go to the firms with the highest bids. Again, firms are forced to confront a cost for using up the resource. 3) It can simply order each polluter to reduce pollution by a certain amount. A major problem with such a command-and-control solution is that the reduction in pollution may be greater or less than the efficient amount. That is, the reduction that the government orders may not be the same reduction that would occur if the firm were facing the true price of the resource.

In general, most countries rely on command-and-control mechanisms for dealing with environmental problems. However, in recent years market-oriented approaches have made some inroads. In the United States, for example, there is now an active market in allowances to emit sulfur dioxide into the air. An important area for future research is to see if it is possible to expand the scope of such policies, and to determine whether the efficiency gains that theory predicts actually occur (Stavins, 2002).

Social Insurance

One way to obtain some protection against the uncertainties of life is to purchase insurance. In private insurance markets, people pay premiums to an insurance company, and receive benefits in the event of certain unlucky occurrences. In addition, a number of government programs also replace income losses that are consequences of events at least partly outside personal control. These programs, collectively referred to as *social insurance*, are among the largest components in the budgets of western governments.

Is there a rationale within conventional welfare economics for such substantial government involvement in insurance markets? There are reasons to believe that private insurance markets will fail to operate efficiently. To see why, note that we can expect an individual who knows he is especially likely to collect benefits to have an especially high demand for insurance, a phenomenon known as *adverse selection*. Due to adverse selection, in order to break even, the insurance company must charge a higher premium for individual coverage than it would if a random group of people were buying the insurance. However, these higher premiums exacerbate the adverse selection problem. Only individuals who know they are at great risk will pay the high prices. This, in turn, requires a further increase in premiums, and the pattern continues. The market fails to provide an efficient amount of insurance.[3] In essence, mandatory social insurance solves this problem by forcing everybody into one big group—the country.

Government retirement programs, which, in effect, provide insurance against the possibility that people will outlive the resources they have accumulated for retirement, are particularly important forms of social insurance. Typically, such programs

---

[3] For a more general treatment of this phenomenon, see Akerlof (1970).

have been funded on a *pay-as-you-go* basis, meaning that the benefits paid to current retirees come from payments made by those who are presently working. The problem is that in most countries, the ratio of retirees to workers will be increasing in coming years. Hence, other things being the same, it will be necessary either to increase the tax rate on current workers or reduce the benefits received by retirees. The best way to cope with this problem is the subject of a major academic and political controversy (Feldstein and Liebman, 2001). Considerable attention has been given to privatizing the systems. Under privatization, workers' contributions are earmarked for their own accounts. Workers then invest the funds in various financial assets, and finance their retirements out of the accumulations in the accounts. Major issues in privatization schemes include how to pay benefits to the current generation of retirees, and how to provide a socially acceptable living standard to individuals who are unable to accumulate enough wealth in their accounts during their working lives.

Other forms of social insurance are unemployment insurance and health insurance. Unemployment insurance provides benefits to workers who lose their jobs. The major problem is how to devise systems that provide protection but do not at the same time make unemployment too attractive (Meyer, 1995). One of the main issues in health insurance is the extent to which the government should directly provide insurance as opposed to providing people with incentives to purchase insurance from the private market. Various nations have come up with quite different solutions. In Canada, for example, health care services are produced by the private sector, with the reimbursements negotiated by the government. In the United Kingdom, health services are produced by the public sector through the National Health Service. In the United States, there is publicly provided insurance only for certain groups, basically the elderly (through Medicare) and for the poor (through Medicaid). A particularly contentious and important issue is the effect that the various systems have on people's health status (Fuchs, 1998).

Income Redistribution

As noted above, even in the absence of market failures, government intervention in the economy may be necessary to achieve a "fair" distribution of real income. A key question in this context is whether the government needs to intervene directly in the markets in order to enhance fairness. For example, should it impose ceilings on the prices of commodities consumed by the poor? The answer is no. Roughly speaking, it is a better policy for the government to redistribute income suitably and then let the markets work. Put another way, the issues of efficiency and distributional fairness can be separated. If society determines that the current distribution of resources is unfair, it need not interfere with market prices and impair efficiency. Of course, the government needs some way to reallocate resources, and problems arise if the only available mechanisms for doing so (such as taxes) themselves induce inefficiencies. These issues are discussed below.

This whole area is complicated by the fact that there is no consensus on what a fair income distribution looks like. Some believe that the government should engineer complete equality. Others believe that society should move toward equality, but take into account the losses in efficiency engendered by taxing high-income people and subsidizing low-income people. Still others believe that attention to the distribution of income at a given point in time is misguided; what matters is whether there is social mobility over time. The idea here is that even if people at the bottom of the income distribution are quite poor, it may not be a major social problem if the identities of these people change over time (Atkinson, 1983).

In many countries, income distribution programs rely primarily on *in-kind transfers*—payments from the government to individuals in the form of commodities or services rather than cash. In-kind transfers include medical care, food, housing, and energy consumption. A natural question is: 'Why do not governments simply give cash to the poor and let them spend the money as they want to?' One possibility is that policy makers care about the distribution of certain commodities rather than income *per se*. For example, they may want every family to consume housing of a given quality. In addition, in-kind transfers may help curb welfare fraud. In-kind transfers may discourage ineligible persons from applying because some well-off people may be willing to lie to receive cash, but be less willing to lie to obtain some commodity they do not really want (Nichols and Zeckhauser, 1982). Finally, in-kind transfers are attractive politically because they help not only the beneficiary but also the producers of the favored commodity. Thus, for example, agricultural interests can be expected to support programs for subsidizing food consumption by the poor.

One of the most contentious issues in this area is how income maintenance policies affect the behavior of the poor. Most attention has been focused on work effort—do beneficiaries reduce their work effort and if so, by how much? In the belief that welfare reduces work effort, several countries have introduced work requirements—in order to be eligible for welfare, recipients have to agree to accept work or job-training programs. The efficacy of such programs is not yet well understood. Another open question is whether income maintenance programs lead to the creation of a "welfare culture"—children brought up in households receiving welfare come to view it as a way of life and hence are unlikely to acquire the skills necessary to earn a living. It is indeed the case that a mother's participation in welfare increases the probability that her daughter eventually also ends up on welfare. However, it is not clear whether the exposure to welfare "causes" the daughter to go on welfare, or if other correlated aspects of the family environment are responsible for it (Blank, 1997).

A Caveat

We have discussed a number of situations in which the government can improve welfare by enhancing efficiency and fairness. However, the fact that the market-generated allocation of resources is imperfect does not necessarily mean the government is capable of doing better. For example, in certain cases the costs of setting up a government agency to deal with some market failure could be greater

than the cost of the market failure itself. Moreover, governments, like people, have only imperfect information, and hence can make mistakes. Finally, it is not clear that government decision-makers will have maximizing social welfare as their goal; we return to this theme at the end of this essay. Hence, it is best to think of welfare economics as helping us identify situations in which government intervention *may* enhance efficiency and fairness; whether it actually will needs to be evaluated on a case by case basis.

The Theory of Taxation

Taxes are the most important source of revenue for modern economies. The theory of taxation explores how taxes should be levied to enhance economic efficiency and to promote a "fair" distribution of income. Just as in the case of expenditures discussed in Section 3, welfare economics provides the underlying analytical framework. Various aspects of the theory are discussed in this section.

Tax Incidence

Policy debates about taxation are usually dominated by the question of whether its burden is distributed fairly. A discussion on this normative issue requires some understanding of the positive question of how taxes affect the distribution of income. A simple way to determine how taxes change the income distribution would be to conduct a survey in which each person is asked how many dollars he or she pays to the tax collector each year.

Although such an approach is convenient, it is quite likely to produce misleading answers. To see why, suppose that the government levies a tax of one dollar on the sellers of a certain commodity. Suppose that prior to the tax, the price of the commodity is $20, and that after the tax is levied, the price increases to $21. Clearly, the sellers receive as much per unit sold as they did before. The tax has not made them worse off. Consumers pay the entire tax in the form of higher prices. Suppose that instead, the price increases to $20.25. In this case, sellers are worse off by 75 cents per unit sold; consumers are worse off by 25 cents per unit sold. The burden of the tax is shared between the two groups. Yet another possibility is that after the tax is imposed, the price stays at $20. If so, the consumer is no worse off, while the seller bears the full burden of the tax.

The *statutory incidence* of a tax indicates who is legally responsible for the tax. All three cases above have exactly the same statutory incidence. But the situations differ drastically with respect to who really bears the burden. The *economic incidence* of a tax is the change in the distribution of private real income induced by the tax.

The example above suggests that the economic incidence problem is fundamentally one of determining how taxes change prices. In the conventional supply and demand model of price determination, the economic incidence of a tax depends on

how responsive supply and demand are to prices.[4] In general, the more responsive supply is to price, relative to demand, the greater the share of the tax that will be shifted to consumers. Intuitively, the more responsive demand is to price, the easier it is for consumers to turn to other products when the price goes up, and therefore more of the tax must be borne by suppliers. Conversely, if consumers purchase the same amount regardless of price, the whole burden can be shifted to them. In cases where the responses of supply and demand to price are well understood, then fairly reliable estimates of the economic incidence of a tax can be obtained. In some areas, the behavioral responses are not well understood, and then incidence analysis is on less firm ground. For example, there is still great controversy over the burden of taxes on corporations—to what extent are they borne by owners of capital, and to what extent by laborers? This is an important topic for research.

Excess Burden

Taxes impose a cost on the taxpayer. It is tempting to view the cost as simply the amount of money that he or she pays to the government. However, this is only part of the story. A tax distorts economic behavior—in general, consumers buy fewer taxed goods and more untaxed goods than otherwise would have been the case. Their decisions are not based entirely on the merits of the commodities themselves. In the same way, business owners make investments based in part on tax considerations, as opposed to economic fundamentals. Because a tax distorts economic activity, it creates a loss in welfare that actually exceeds the revenues collected. This is referred to as the *excess burden* of the tax.

In general, the more responsive behavior is to the tax, the greater the excess burden, other things being the same. Intuitively, because excess burdens arise because of distortions in behavior, the more that behavior is capable of being distorted, the greater the excess burden. Another important result is that the excess burden of a tax increases with the square of the tax rate—doubling a tax quadruples its excess burden, other things being the same. This means that, in general, it makes sense to spread taxes over as large a group of commodities as possible—a small tax on a number of commodities has a smaller excess burden than a very large tax on one commodity.

This discussion suggests that, just like the incidence problem discussed above, the excess burden of a tax depends on the behavioral response to the tax. Estimating such behavioral responses and computing excess burdens is an important role for public finance economists. Some estimates suggest that the excess burdens for real-world tax systems are quite high. One recent survey suggested that in the United States, the average excess burden per dollar of tax revenue is 18 cents. While any

---

[4] For a treatment of tax incidence in other models of price determination, see Fullerton and Metcalf (2002).

particular figure must be taken with a grain of salt, virtually all estimates suggest that the tax system is highly inefficient in the sense of generating large excess burdens (Jorgenson and Yun, 2001).

The fact that a tax generates an excess burden does not mean that the tax is bad. One hopes, after all, that it will be used to obtain something beneficial for society either in terms of enhanced efficiency or fairness. But to determine whether or not the supposed benefits are large enough to justify the costs, sensible policy requires that excess burden be included in the calculation as a cost to society.

Optimal Taxation

Public finance economists have devoted a great deal of attention to the problem of the design of optimal taxes. Of course, this is a normative issue, and it cannot be answered without a statement of ethical goals. To begin, suppose that the goal is to raise a given amount of money with the smallest amount of excess burden possible. There are a variety of ways to characterize the result. One of the most elegant is the rule that as long as goods are unrelated in consumption (that is, are neither substitutes nor complements), then the more responsive demand is to price, the lower should be the tax rate on that commodity. The intuition behind this rule is straightforward. Efficient taxes should distort decisions as little as possible. The potential for distortion is greater the more responsive the demand for the commodity is to its price. Therefore, efficient taxation requires that relatively high rates of taxation be levied on goods whose demands are relatively unresponsive to their price.

This result strikes many people as ethically unappealing. For example, the demand for food is relatively unresponsive to changes in its price. Is it really desirable to tax food at relatively high rates? Most people would argue that it is not desirable, because their ethical views indicate that a tax system should have *vertical equity*: It should distribute burdens fairly across people with different abilities to pay. Public finance economists have shown how to modify the efficiency rule to account for the distributional consequences of taxation. Suppose, for example, that the poor spend a greater proportion of their income on commodity X than do the rich, and vice versa for commodity Y. Then even if the demand for X is less responsive to price than the demand for Y, optimal taxation may require a higher rate of tax on Y than X. True, a high tax rate on Y creates a relatively large excess burden, but it also tends to redistribute income toward the poor. As in other areas of public finance, the optimal policy depends on the extent to which society is willing to tradeoff efficiency for fairness (Auerbach and Hines, 2002).

With its focus on efficiency and fairness issues, the theory of optimal taxation falls directly within the framework of conventional welfare economics. There are other criteria for tax design that are not reconciled so easily with welfare economics. The main one is *horizontal equity*, the notion that people in equal positions should pay equal amounts of taxes. One problem with implementing this principle is defining equal positions. The most common criterion is income, but wealth and consumption are also possible. A problem with all three measures, however, is that they are the

outcomes of people's decisions. Two individuals may have exactly the same wage rate, but one chooses to work 1,000 hours per year while another chooses to work 2,000 hours per year. Despite the fact that they have different incomes, in a meaningful sense they are in "equal positions" because their potential to earn income is the same.

Things are complicated further by the fact that adjustments in market prices may render some horizontal inequities more apparent than real. Suppose, for example, that in one type of job a large part of compensation consists of amenities that are not taxable—pleasant offices, access to a swimming pool, and so forth. In another occupation, compensation is exclusively monetary, all of which is subject to income taxation. This would appear to be a violation of horizontal equity, because the person in the job with a lot of amenities has too small a tax burden. But, if both arrangements coexist and individuals are free to choose, then the net after-tax rewards (including amenities) must be the same in both jobs. Otherwise, people would leave the job with the lower net after-tax rewards. In short, the fact that amenities are not taxed is not unfair, because the before- tax monetary compensation falls by just enough to offset this advantage. Put another way, introducing taxation for such amenities would *create* horizontal inequities (Feldstein, 1976).

We conclude that horizontal equity is a rather amorphous concept. Yet it has enormous appeal as a principle of tax design. Notions of fairness among equals, regardless of their vagueness, will continue to play an important role in the development of tax policy.

Revenue Raising Instruments

Public finance economists have used the theoretical framework discussed above to analyze the various revenue sources used by modern governments. This section discusses briefly some of the key issues associated with each kind of tax.

Income Tax

Taxes on income play a major role in the fiscal systems of all western countries. A starting point for the analysis and evaluation of real world income tax systems is a definition of income. Traditionally, public finance economists use the so-called *Haig-Simons* definition: Income is the money value of the net increase in an individual's power to consume during a period. This is equal to the actual amount during the period plus net additions to wealth. Net additions to wealth—savings—must be included in income because they represent an increase in potential consumption. Importantly, the Haig-Simons criterion requires the inclusion of *all* sources of potential increases in consumption, regardless of whether the actual consumption takes place, and regardless of the form in which the consumption occurs. While not uncontroversial, the Haig-Simons definition provides a useful guide.

The Haig-Simons definition encompasses those items ordinarily thought of as income: wages and salaries, business profits, rents, royalties, dividends, and interest. These forms of income are relatively easy to measure and to tax. However, in other contexts, implementing the Haig-Simons criterion can lead to major problems.[5] Some examples follow:

- Only the net income of business expenses increases the potential of consumption power. But distinguishing between consumption and costs of obtaining income can be difficult. To what extent is a desk bought for an office at home just furniture, and to what extent is it a business expense?
- A capital gain is the increase in the value of an asset—say, a share of stock—during a period of time. From a Haig-Simons point of view, a capital gain is income whether or not the stock is actually sold, because the capital gain represents an increase in potential to consume. However, capital gains and losses may be very difficult to measure, particularly when the assets are not sold. Indeed, in general, no attempts are made to tax capital gains of assets that have not actually been sold.
- 'In-kind' services are not easy to value. One important example is the income produced by people who do housework rather than participate in the market.

Such difficulties in implementing a Haig-Simons concept of income are of great practical significance. To the extent that income that comes in certain forms cannot be taxed, individuals' decisions are biased in the direction of taking their income in those forms. Thus, for example, there is a bias in favor of capital gains (which are taxed only when the asset is sold) as opposed to dividend income (which is taxed as it is earned). Such biases create efficiency losses to the economy. Further, complicated rules are often needed to determine whether a certain type of income falls in a category that is favored by the tax system. Capital gains again provide a good example; it is not always obvious whether the return that an individual receives from a company is a dividend or a capital gain. Such complexity leads to substantial compliance costs.

In additions, several forms of income that would be administratively relatively easy to tax are partially or altogether excluded from the income tax bases of most countries. An important example is the return on saving that is deposited in retirement accounts. Indeed, given the extent to which income that is saved in various forms is excluded from taxation, it is a misnomer to characterize these systems as income taxes. They are more a hybrid between income and consumption taxes.

Corporation Income Tax

Corporations are independent legal entities and as such are subject to taxes on their incomes. Most public finance economists believe that it makes little sense to levy a

---

[5] Bradford (1986) provides a careful discussion of issues relating to the implementation of an income tax.

special tax on corporations. Only real people can pay a tax; hence, it would make more sense to tax the incomes of corporation *owners* via the personal income tax. Again, this distinction is of more than academic importance. Treating the corporation as a freestanding entity for tax purposes leads to important distortions in economic activity. To see why, note that when a corporation earns income it is taxed once at the corporate level, and then again when it is paid out to shareholders in the form of dividends. In effect, then, corporate income that is paid out in the form of dividends is double taxed. This biases businesses against organizing in corporate form. Moreover, double taxation of corporate income effectively increases the tax rate on the return to corporate investments. This reduces the volume of investment undertaken by corporations, although there is substantial disagreement about the magnitude of this effect.

The incidence of corporate tax is highly controversial. In one highly influential model of Harberger (1962), the tax on corporate capital leads to a migration of capital from the corporate sector until after-tax rates of return are equal throughout the economy. In the process, the rate of return to capital in the non-corporate sector is depressed so that ultimately all owners of capital, not just those in the corporate sector, are affected. The reallocation of capital between the two sectors also affects the return to labor. Most public finance economists believe that the burden of the corporate tax is split between labor and capital, although there is significant disagreement about the exact division.

If corporate income was untaxed, individuals could avoid personal income taxes by accumulating income with corporations. Evidently, this would lead to serious equity and efficiency problems. The question is whether there is a way to integrate personal and corporate income taxes into a single system so as to avoid the distortions associated with double taxation. The most radical solution to this problem is called *full integration*. Under this approach, all earnings of the corporation during a given year, whether they are distributed or not, are attributed to stockholders just as if the corporation were a partnership. The corporate tax as a separate entity is eliminated. This approach has not been implemented in any country, in part because of administrative problems. The *dividend relief approach* is less extreme. With it, the corporation can deduct dividends paid to stockholders. Although this approach eliminates the double taxation of dividends, it still maintains the corporate tax as a separate entity. Variants of this approach are used in a number of European nations.

Consumption Taxes

The base of a consumption tax is the value (or quantity) of commodities sold to a person for *actual* consumption, as opposed to an income tax, whose base is the change in *potential* consumption. Consumption taxes take a variety of forms. A *retail sales tax* is levied on the purchase of a commodity. In the United States, retail sales taxes are not a significant component of revenue at the national level, but they are at the state level. However, even there, the rates generally do not exceed 7 percent or so.

In Europe, the most important type of consumption tax is a *value-added tax* (VAT). The *value-added* at each stage of production of a commodity is the difference between the firm's sales and the purchased material inputs used in production. If a firm pays $100 for its material inputs and sells its output for $150, then its value added is $50. A VAT is a percentage tax on value-added at each stage of production. For example, if the VAT rate were 10 percent, then the firm's tax liability would be $5. Note that the total value of a commodity when it is finally sold is equal to the sum of the value-added at each stage of production. Hence, a VAT of 10 percent applied to each stage is equivalent to a 10 percent tax on the final product. In Europe, VAT rates are as high as 25 percent. With rates at such levels, evasion is likely to be a problem for retail sales taxes; VATs are easier to administer, which accounts for their popularity.[6]

A distinguishing feature of both VATs and retail sales taxes is that the tax liability does not depend on the characteristics of the buyer. Whether one is rich or poor, the rate is the same. This prompts concerns over equity, which has been dealt with by applying lower rates to commodities such as food and medicine. But this may not be an effective way to deal with equity concerns. For example, even if it is true that food expenditures on average play an especially important role in the budgets of the poor, there are still many upper-income families whose food consumption is proportionately very high. In recent years, public finance economists have given a great deal of attention to the problem of designing *personal* consumption taxes. Such taxes require individuals to file tax returns and write checks to the government, allowing tax liabilities to depend on personal circumstances.

One example is a *cash-flow* tax. Each household files a return reporting its annual consumption expenditures during the year. Just as under the personal income tax, various exemptions and deductions can be taken to allow for special circumstances, and a progressive marginal rate schedule applied to taxable consumption. From an administrative viewpoint, the major question is how to compute annual consumption. Taxpayers would report their incomes, and then subtract all savings. To keep track of savings, qualified accountants are needed at various financial institutions. Whether a cash-flow tax is administratively feasible is very controversial.[7] Many analysts believe that its record-keeping requirements would make it very difficult or impossible administratively.

Wealth Taxes

Wealth is the value of the assets an individual has accumulated as of a given time. Wealth taxes do not play a major role in the fiscal systems of western countries. One justification for taxing wealth is that it is a good measure of an individual's ability to pay taxes. This is a controversial issue. Suppose that a miser has accumulated a

---

[6] See Cnossen (1998) for a discussion of issues relating to the implementation of VATs.

[7] The difficulties and advantages of this system are discussed in Pechman (1980).

huge hoard of gold that yields no income. Should she be taxed on the value of the hoard? Some believe that as long as the miser was subject to income tax while the hoard was accumulating, it should not be taxed again. Others would argue that the gold *per se* generates satisfaction and power for the individual, and should therefore be subject to tax. Perhaps the major problem with this argument is that many rich people have a substantial component of their wealth in *human* capital— their stock of education, skills, and so on. However, there is no way to value human capital except by reference to the income it yields. This logic points back to income as the appropriate basis for taxation.

Some nations levy taxes on wealth only when it is transferred at the time of the death of the owner. These are referred to as *estate taxes*. Estate tax proponents argue that it is a valuable tool for creating a more equal distribution of income. Further, many believe that ultimately, all property belongs to society as a whole. During an individual's life, society permits her to dispose of the property she has managed to accumulate as she wishes. But at death, the property reverts to society, which can dispose of it at will. Opponents argue that it is fundamentally wrong to argue that a person holds wealth only at the pleasure of "society," or that "society" ever has any valid claim on personal wealth.[8]

A controversial issue is the incentives created by an estate tax. Suppose that an individual is motivated to work hard during his lifetime to leave a large estate for his children. The presence of an estate tax might discourage his effort. On the other hand, with an estate tax, a greater amount of wealth has to be accumulated to leave a given after-tax bequest, so the tax might induce the individual to work harder to maintain the net value of his estate. Consequently, the effect of an estate tax on a donor's effort is logically indeterminate. Similarly, one cannot predict how the tax will affect the amount of savings. There is currently very little in the way of empirical evidence on these incentive issues.

To the extent that an estate tax reduces savings, it may actually *increase* inequality. If there is less saving, then, there is less capital investment. With less capital with which to work, the real wages of workers decrease and under certain circumstances, the share of income going to labor falls. To the extent that capital income is more unequally distributed than labor income, the effect increases inequality. This scenario is hypothetical. It simply emphasizes a point made above in a variety of different contexts—to understand the impact of a tax, one must take into account how taxpayers respond to it.

Deficit Finance

In addition to taxation, the government's other major source of revenue is borrowing. The *deficit* during a time period is the excess of spending over revenues. The *national debt* at a given time is the sum of all past budget deficits. That is, the

---

[8] See Gale and Slemrod (2000) for further details.

debt is the cumulative excess of past spending over past receipts. Future generations either have to retire the debt or else refinance it. It would appear, then, that future generations must bear the burden of the debt. But the theory of incidence tells us that this line of reasoning is questionable. Merely because the legal burden is on future generations, it does not mean that they bear a real burden. Just as in the case of tax incidence, the answer depends on the economic behavior.

Assume that the government borrows from its own citizens. One view is that such an internal debt creates no burden for the future generation. Members of the future generation simply owe it to each other. There is a transfer of income from those who do not hold bonds to the bondholders, but the generation as a whole is no worse off in the sense that its consumption level is the same as it would have been.

This story ignores the fact that economic decisions can be affected by government debt policy. According to the *neoclassical model of the debt*, when the government borrows, it competes for funds with individuals and firms who want the money for their own investment projects. Hence, debt finance leaves the future generation with a smaller capital stock, other things being the same. Its members therefore are less productive and have smaller real incomes than otherwise would have been the case. Thus, the debt imposes a burden on future generations, through its impact on capital formation. The key assumption in this argument is that public spending crowds out private investment. Whether crowding out actually occurs is a controversial issue; the empirical evidence is mixed (Elmendorf and Mankiw, 1999).

A further complication is introduced when we consider individuals' transfers across generations. Suppose that when the government borrows, people realize that their heirs will be made worse off. Suppose further that people care about the welfare of their descendants and do not want their descendants' consumption levels reduced. What can they do about this? They can save more to increase their bequests by an amount sufficient to pay the extra taxes that will be due in the future. The result is that nothing really changes. Each generation consumes exactly the same amount as before the government borrowed.

The striking conclusion is that private individuals undo the intergenerational effects of government debt policy so that tax and debt finance are essentially equivalent. This view is sometimes referred to as the *Ricardian model* because its antecedents appeared in the work of the 19th century economist David Ricardo. (However, Ricardo was skeptical about the theory that now bears his name.) Some public finance economists have challenged the plausibility of the *Ricardian model*. They believe that information on the implications of current deficits for future tax burdens is not easy to obtain. Another criticism is that people are not as farsighted and not as altruistic as supposed in the model. A number of statistical studies have examined the relationship between budget deficits and private saving. The evidence is rather mixed, and the Ricardian model has both critics and adherents among professional economists.

From time to time, events such as natural disaster and wars lead to temporary increases in the federal government expenditures. An old question in public finance is whether such expenditures should be financed with taxes or by borrowing.

Fiscal Federalism

The analysis so far has assumed that a nation has one government that sets tax and expenditure policies. In contrast, many countries have a federal system, which consists of different levels of government that provide public goods and services and have some scope for making decisions. The subject of *fiscal federalism* concerns the activities of the various levels of government and how they relate to each other. A key question is the optimal allocation of responsibilities among the different levels of government. Posed within the framework of welfare economics, the question is whether a centralized or decentralized system is more likely to enhance efficiency and equity (Oates, 1999).

Among the disadvantages of a decentralized system is that individual communities may ignore the externalities they create. Suppose, for example, that some jurisdiction provides excellent public education for its children. If some of the children eventually emigrate to other jurisdictions, the other communities benefit from having a higher quality of work force. But in deciding how much education to provide, the jurisdiction only considers its own welfare. Therefore, it may provide an inefficiently low amount of education. More generally, if each community cares only about its own members, then any positive or negative externalities it creates for other communities are overlooked. According to the standard arguments made above, resources are allocated inefficiently.

Another disadvantage of a decentralized system relates to the fact that for certain public services, the cost per person falls as the number of users increases. Suppose that the more the number of people who use a public library, the lower the cost per user. If each community sets up its own library, costs per user are higher than necessary. A central government, on the other hand, could build one library for the region, allowing people to benefit from scale economies. Of course, various activities are subject to different scale economies. The optimal scale for library services might differ from that for fire protection, and both surely differ from the optimal scale for national defense. This observation helps rationalize a system of overlapping jurisdictions—each jurisdiction can handle those services with scale economies that are appropriate for the jurisdiction's size.

Decentralized systems can also lead to inefficiencies with respect to raising revenues. Taxes levied by decentralized communities are unlikely to be efficient from a national standpoint. Instead, communities are likely to select taxes on the basis of whether they can be exported to outsiders. For example, jurisdictions that have a near-monopoly on certain natural resources such as coal may impose large taxes on these commodities, figuring that they will be shifted largely to coal users outside the community.

A major advantage to a decentralized system is that it allows communities to tailor their public services to the tastes of their residents. Taste for public services, just like the taste for all other commodities, varies across people. A centralized government tends to provide the same level of public services throughout the country, regardless of the fact that people's tastes differ. It is inefficient to provide

individuals with more or less of a public good than they desire if the quantity they receive can be more closely tailored to their preferences. Under a decentralized system, individuals with similar tastes for public goods group together, so communities are more likely to provide the types and quantities of public goods desired by their inhabitants.

Another advantage is that decentralized systems foster intergovernmental competition. If citizens can choose among communities, then substantial government mismanagement may cause citizens to choose to live elsewhere. This threat may create incentives to government managers to produce more efficiently and be more responsive to their residents.

Finally, a decentralized system may enhance experimentation and innovation in locally provided goods and services. For many policy questions, no one is certain what the right answer is, or even whether there is a single solution that is best in all situations. One way to find out is to let each community choose its own way, and then compare the results. For example, some jurisdictions might choose to provide innovative job-training programs for individuals who lose their jobs. If the innovations are successful, other jurisdictions can imitate them. If not, the costs to the country as a whole are small.

This discussion makes it clear that a purely decentralized system cannot be expected to maximize social welfare. Efficiency requires that those services that affect the entire country, such as national defense, be provided at the national level. On the other hand, it seems appropriate for goods that affect only the members of a particular jurisdiction to be provided locally. This leaves us with the in-between case of community activities that create externalities that are not national in scope. While one solution would be to create a single regional government, a larger jurisdiction carries the cost of less responsiveness to local differences in tastes. An alternative method is a system of taxes and subsidies. The central government can subsidize activities that create positive externalities. In some countries, central governments give grants to communities that roughly follow this model.

Public Finance and Public Choice

Traditionally, the field of public finance has tended to convey a rather rosy view of government. With a tax here and an expenditure there, the state readily corrects all market imperfections, seeing to it meanwhile that incomes are distributed in an ethically desirable way. The implicit assumption is that the government is a neutral and benign force. In contrast, the field of public choice assumes that individuals view government as a mechanism for maximizing their self interest. Such a viewpoint can lead to rather different conclusions from those of conventional public finance.

A good example is provided by optimal tax theory. Suppose that in a certain society, there are three commodities, X, Y, and leisure. Labor is totally fixed in supply, and therefore, income is fixed. Note that a proportional tax at the same rate on X and Y is equivalent to a tax on income. Now, suppose that currently, this

society levies a tax on X, but its constitution forbids taxing Y. Viewing this situation, a student of optimal tax theory might say something like, "You are running an inefficient tax system. You could eliminate excess burden if you taxed X and Y at equal rates — an income tax. I recommend that you lower the tax on X and impose a tax at the same rate on Y. Set the rates so that the same amount of revenue is collected as before."

Suppose, however, that the citizens suspect that if they allow taxation of Y, their politicians will not lower the tax rate on X. Rather, they will simply take advantage of the opportunity to tax something new to make tax revenues as large as possible. Therefore, by constitutionally precluding the taxation of Y, the citizens may be rationally protecting themselves against an inefficiently large public sector. In other words, if government does not necessarily act in the interest of its citizens, then what looks inefficient from the point of view of optimal tax theory may be efficient in a public choice setting.[9]

In recent years, public choice has had substantial influence on the field of public finance. In both theoretical and empirical work, public finance economists study the incentives facing government decision-makers and how these incentives affect policy outcomes. In making their own policy recommendations, there is a heightened awareness that a policy that emerges from the legislative process may look quite different from the original proposal, and one should take this into effect in formulating recommendations. In the future, one can expect both Public Finance and Public Choice to continue to enjoy the benefits of intellectual cross-fertilization.

References

Akerlof, G. (1970). "The market for lemons: quality uncertainty and the market mechanism." *Quarterly Journal of Economics*, 84, 488–500.

Atkinson, A. B. (1983). *The Economics of Inequality*. Oxford: Oxford University Press.

Bator, F. M. (1957). "The simple analytics of welfare maximization." *American Economic Review*, 47, 22–59.

Blank, R. (1997). *It Takes a Nation — A New Agenda for Fighting Poverty*. New York: Russell Sage Foundation.

Bradford, D. (1986). *Untangling the Income Tax*. Cambridge, MA: Harvard University Press.

Cnossen, S. (1998). "Global trends and issues in value-added taxation." *International Tax and Public Finance*, 5, 399–428.

Elmendorf, D. W. and N. G. Mankiw, (1999). "Government debt," in J. B. Taylor and M. Woodford (eds.) *Handbook of Macroeconomics*. Amsterdam: North-Holland.

Feldstein, M. (1976). "On the theory of tax reform." *Journal of Public Economics*, 6, 77–104.

Feldstein, M. and J. B. Liebman (2001). "Social security." Working Paper No. 8541, National Bureau of Economic Research, September.

Fuchs, V. (1998). "Health, government, and Irving Fisher." Working Paper No. 8490, National Bureau of Economic Research, August.

---

[9] Holcombe (1998) provides further comparisons between the optimal tax theory and a public choice approach.

Gale, W. G. and J. B. Slemrod (2000) "A matter of life and death: reassesing the estate and gift tax." *Tax Notes*, August 14, 927–932.

Hargberber, A. C. (1962). "The incidence of the corporation income tax." *Journal of Political Economy*, 70, 215–240.

Holcombe, R. (1998). "Tax policy from a public choice perspective." *National Tax Journal*, 51, 359–371.

Jorgenson, D. W. and K. Y. Yun (2001). *Investment Volume 3: Lifting the Burden*. Cambridge: The MIT Press.

Meyer, B. D. (1995). "Lessons from the US unemployment insurance experiments." *Journal of Economic Literature*, 33, 91–131.

Nichols, A. L and R. J. Zeckhauser (1982). "Targeting transfers through restrictions on recipients." *American Economic Review Papers and Proceedings*, 72, 372–377.

Oates, W. E. (1999). "An essay on fiscal federalism." *Journal of Economic Literature*, 37, 1120–1149.

Pechman, J. A. (ed.) (1980). *What Should Be Taxed: Income or Expenditure?* Washington: The Brookings Institution.

Samuelson, P. A. (1954). "The pure theory of public expenditure." *Review of Economics and Statistics*, 36, 387–389.

Stavins, R. N. (2002). "Experience with market-based environmental policy instruments," in K.-G. Maler and J. Vincent (eds.) *The Handbook of Environmental Economics*. Amsterdam: North-Holland/Elsevier Science.

Chapter 23
Taxation[1]

Stanley L. Winer and Walter Hettich

Introduction

There are many reasons for studying taxation. As the ancient Roman writer Cicero pointed out so succinctly, when he called it the sinews of the state, taxation is central to the existence and functioning of a nation, as well as to the functioning of its lower levels of government. Taxing citizens is a vital method of financing the most essential public sector activities, such as the courts, the legal system, national defense, and police protection. In addition, it provides the means for producing social programs, such as public health services, education, and welfare. Finally, taxation is one of the most important ways in which a community's distributional goals may be attained.

The study of collective choice is an essential part of any comprehensive analysis of taxation. The activities of communities differ in nature from activities carried out by the private sector. The provision of most publicly provided goods or services cannot be accomplished and organized through markets. Instead, collective choice procedures are needed to allocate the required resources and to decide the level and extent of public provision. In democratic countries, governmental expenditures, and the ways of raising the necessary revenues, are usually determined by some type of majority rule, although such rule may be limited or attenuated by constitutional provisions and constrained by the operation of a competitive system of political parties. Taxation thus represents an essential tool for decision makers who want to command scarce resources for use in the public sector as part of the democratic process.

In this paper we review both the positive and normative aspects of taxation. We examine how to study why taxes and revenue structures have taken their present form and why they are used in a particular way as part of the democratic process.

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume 1, 173–190.

In addition, we also consider the classic normative questions, namely what makes a good tax system and how to assess the efficiency of taxation. In dealing with both aspects of tax literature, we attempt to set out a plan for a more complete and comprehensive analysis of taxation in the face of collective choice than is attempted in most of the available literature on fiscal issues.

Since our emphasis is on the positive and normative study of tax structure and tax systems, we pay only limited attention to the political economy of redistribution. Some branches of fiscal literature make the link between progressive taxation and redistribution their main focus, while trying to analyze how income tax rates are determined as part of the political struggle over a society's income shares. While we shall consider studies of this nature, we conceive of fiscal analysis as an enterprise of broader scope than is implied by the approach adopted in this work. In our view, questions of efficiency as well as of redistribution are involved in studying political equilibria, and fiscal policy encompasses the use of many different kinds of taxation as well as of other policy instruments such as regulation that affect the drawing of resources into the public sector.

Although most dictatorial regimes also make use of taxation, we deal primarily with fiscal choices in democratic states in this essay. The reader with a special interest in the analysis of authoritarian regimes may want to consult the recent work by Wintrobe (1996), where a theory of such states is developed. Comparative international analysis relevant to this topic can be found in Musgrave (1969) and Mulligan and Sala-i-Martin (2004), and in Kenny and Winer (2006) where tax systems in a large sample of countries are examined in an econometric framework.

In view of space constraints, references to the literature will be illustrative rather than exhaustive. The essay emphasizes theoretical ideas and empirical issues relevant to the study of tax systems and tax structure. Further bibliographic material can be found in Hettich and Winer (1999), in Boadway and Keen (2000) concerning the political economy of redistribution, in Gould and Baker (2002) and Kirchgaessner (2002) with respect to taxation and political institutions, and in Webber and Wildavsky (1986) concerning fiscal history.[2]

Basic Issues

Two Major Approaches

There are two broad approaches to the study of taxation, both with an extensive and well-developed literature. The first one is associated with the work of Wicksell (1896) and Lindahl (1919), two of its most important early proponents, as well as

---

[2] There are also several excellent studies of the political and economic aspects of specific taxes or episodes in tax history of particular countries. This literature notably includes Witte (1985) on the income tax system of the United States and Gillespie (1991) on the tax structure of Canada.

the work of Buchanan (e.g., 1968, 1976). Here, taxation is seen as part of an exchange—albeit an imperfect one—between citizens and their government. Tax payments are made to obtain public goods and services, and to some extent, to participate in collectively determined redistribution. The emphasis is on taxation as a price for public output consumed by voters, and on institutions or methods designed to link the fiscal and the expenditure sides of the budget.

A second approach sees taxation as a set of policies that are linked only indirectly to the expenditure side via the government budget restraint. Taxation is analyzed as the coercive taking of resources to finance largely unspecified government activities. The emphasis is on ways to minimize the efficiency costs of taxation through the policy choices of a social planner. Such a planner may also take account of distributional aims in achieving his or her objectives, by including distributional weights in the design of fiscal policy. Such weights will be derived from an exogenously given welfare function, rather than being the outcome of a political process. The second approach has its origins in the work of Edgeworth (1925), Ramsey (1927), and Pigou (1951) and has been developed with great analytical sophistication by Mirrlees (1971) and others in the recent literature on optimal taxation.[3]

Although the two approaches to taxation are quite distinct in emphasis and in the results that they reach, both must contend with the same central problem, namely the separation of taxing and spending. Governments provide goods and services that are different in nature from those provided through private markets. So-called public goods, such as defense, are consumed equally by all members of the collectivity, and it is not possible to ration such goods according to price, as is done in markets for private goods. The same is true for goods that are mixed in nature, having both public and private characteristics. The difficulty of excluding those who do not pay voluntarily from enjoying the benefits of public output gives rise to the problems of preference revelation and free-riding. In response, most collectivities use coercive taxation to finance public output, creating tax systems in which there is only a diffuse and distant link between additional consumption of publicly provided goods and increases in tax liability.

The separation of taxing and spending gives rise to welfare losses over and above the losses due to the tax payment itself, a primary focus of the tax literature.[4] Individual taxpayers will respond to tax rates by adjusting their activities so as to reduce their tax liability, with such adjustments being quite unrelated to the consumption of publicly provided goods. If an income tax is levied, for example, taxpayers may reduce work effort and consume more leisure, in order to maximize utility in the face of such taxation. This results in a reduction of economic

---

[3] For a review of Optimal Taxation, see, for example, Atkinson and Stiglitz (1980), Stiglitz (1987), and Salanie (2003).

[4] See Creedy (1998) for an extended discussion of the meaning and measurement of excess burden.

welfare in comparison to a situation where payment for the same public output elicits no such trade-off or evasive adjustment. The size of this loss—the excess burden or deadweight cost of taxation—is used in the literature as a measure of the inefficiency created by a particular tax. The same type of analysis can also be used to compare the efficiency (or inefficiency) of different available tax instruments.

Separation of taxing and spending also has implications for redistribution. Since markets cannot be used to allocate public goods and their costs among users, and to determine what level of such goods should be produced, other mechanisms must be employed to reach such decisions.[5] All available collective decision processes create their own incentives for redistribution between those in the majority and those in the minority with regard to a particular fiscal issue. In addition, the separation of the two sides of the budget makes it more difficult to understand the distributional implications of various ways of providing and financing of public programs. This opaqueness may be exploited by those who are in a position to use public resources for their own purposes.[6]

Although both basic approaches to taxation must confront the separation of taxing and spending, they deal with its implications in quite different ways. The first approach focuses on collective decision processes and fiscal structures designed to create a closer link between taxing and spending, or on institutional and fiscal constraints that would have the effect of limiting coercion. Wicksell was the first to suggest ways to reduce coercion, and thus separation, by proposing unanimity, or qualified unanimity, as a budgetary decision criterion. Increases in budgets, as well as the ways of financing them, could only be adopted if they passed according to this criterion. Lindahl further formalized the analysis by providing a theoretical process in which tax shares and the output of public goods were jointly decided in bargaining among the affected decision makers.[7]

The second basic approach adopts a rather different perspective. Decision processes are taken as exogenous, and their effects are not examined as part of the analysis. This is exemplified by the assumption of a social planner who makes decisions on behalf of the collectivity according to an exogenously given welfare function. In this literature, the emphasis is shifted to the identification and measurement of welfare losses, and to the design of tax systems that maximize social welfare, given the assumed analytical framework.

---

[5] This problem has given rise to several normative approaches to the distribution of the tax burden as alternatives to taxation according to benefits received, including most notably the principle of taxation according to ability to pay. See, for example, Musgrave (1959, Ch. 5).

[6] Wagner (1976), Buchanan and Wagner (1977), West and Winer (1980) and others have considered the role of fiscal illusion in the political manipulation of taxation and public debt.

[7] On the concept of a Lindahl equilibrium, see also Foley (1977), Johansen (1965) and Head (1974).

A Comprehensive Approach with Collective Choice

In describing the two approaches, we have emphasized what may be called *normative questions*. Tax analysis has an additional dimension, however. Although it is interesting to ask how efficiency in taxation should be defined, and how it can be measured, it is of equal importance to examine the nature of existing fiscal systems. In this context, we ask why tax systems have the characteristics that we observe and what may explain the variations in revenue systems among different jurisdictions. This type of examination is usually called *positive analysis*.

The study of actual revenue systems reveals that there is a similar underlying structure despite the many variations that are observed (Hettich and Winer, 1999). Taxes are imposed on several major bases, such as personal income, corporate profits, or property. Each tax has a rate structure, which may be simple or more complex. In addition, there are special provisions that affect the definition of the base and that may specify separate rates for particular components of the base, such as capital gains. In practice, tax policy is a manipulation of some aspect, or a combination of characteristics, of this *tax skeleton*. If we follow the methodological approach that underlies all economic theory, we can interpret observed tax systems as equilibrium outcomes of economic and political processes and forces.

A comprehensive or complete approach to taxation will include both positive and normative analysis. It will allow us to analyze observed tax systems, as well as guide us in asking questions about possible improvements in the many existing features of such systems. To achieve this, we need models that encompass theoretical analysis of both positive and normative questions and that allow statistical testing of hypotheses derived from them (Hettich and Winer, 2002).

If we briefly return to consider the two approaches to taxation reviewed earlier, we see that only the first one provides a suitable starting point for the development of a comprehensive analysis. Observed tax systems arise from decisions made through collective choice processes. To explain them, we must start by modeling such processes and by linking actual revenue systems to the predictions of our models. The assumption of a social planner does not provide a starting point for meaningful theoretical or empirical research of this nature. Although we can derive a sophisticated normative analysis with the planner framework, we cannot fully link it to the results of positive analysis as required by a comprehensive approach.[8]

Creation of an inclusive fiscal policy analysis is not an easy task. Although the literature contains many of the necessary elements, they have not been assembled as yet into a fully integrated theoretical system. Figure 23.1 gives a schematic presentation of the different elements in a complete analysis and shows their interrelation.

---

[8] In addition to bypassing the essential role of collective choice, a social planning approach imputes motives to public decision makers that differ from those of their self-interested private counterparts included in the same framework. Brennan and Buchanan (1980), Kau and Rubin (1981), Levi (1988), Wilson (1989), Edwards and Keen (1996) and others have drawn attention to the importance of motivation by public officials in the analysis of taxation.
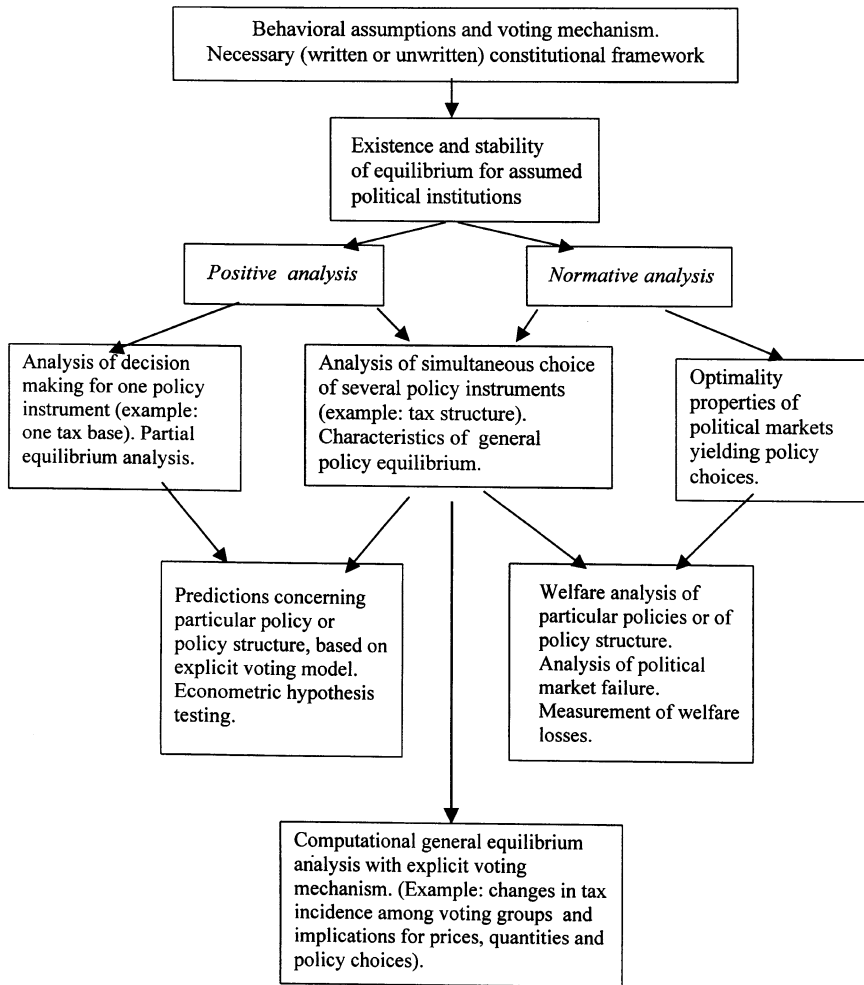
Fig. 23.1 A comprehensive analysis of taxation when collective choice matters

As in traditional microeconomics, we start with the behavioral assumptions of essential decision makers. In democratic countries, decision makers include voters, who have additional roles as taxpayers and consumers of public goods, and who also participate in the private economy. In most models, they are assumed to maximize their utility. We also have politicians who propose policies or platforms and whose goal is to be elected. The interactions of voters and politicians takes place in a given constitutional framework (written or unwritten), a postulate that parallels the assumption of a set of existing property rights in the study of private markets.

To be useful, any proposed model must yield stable equilibria and must be accompanied by proofs of their existence. Otherwise, it is not possible to carry out comparative static analysis of the kind common in economics. The model should yield predictions or hypotheses useful for positive analysis, whether based on partial equilibrium analysis or on a more general framework, so that they can be tested with accepted statistical techniques. A comprehensive approach can also be used for computational general equilibrium analysis that includes an explicit voting mechanism in the modeling of the public sector. This will allow investigation of how exogenous shocks or changes in policy affect the economic welfare of different voting groups, and how these changes in welfare feed back to determine the choice of tax and other policy instruments via the collective choice mechanism represented in the model.

It is desirable that the same framework can also be used for normative analysis. This requires that, under specified conditions, political equilibria satisfy certain characteristics, such as Pareto optimality. If this demonstration can be accomplished, the framework can also be used for work described in the boxes on the right-hand side in Fig. 23.1. One should note the similarity between the proposed scheme and the approach common in economics applied to the private sector, where positive and normative analysis are both based on the same model of competitive markets, where the First Theorem of Welfare Economics (the "invisible hand" theorem) links positive and normative analysis, and where the study of market failure is used as an aid to restore the operation of decentralized market forces. However, any examination of optimality must now refer to *political* markets, not to their private counterpart. The same is true when we turn to the study of deviations from optimality. We now deal with political market failure, rather than with the malfunction of private markets.

Political Equilibria and the Partial and General Equilibrium Study of Tax Systems

A comprehensive approach to the political economy of taxation begins with positive theory, represented schematically by the left side of Fig. 23.1. In proceeding down the left side of the figure, we confine the discussion by and large to two frameworks that have been most widely used for positive theoretical and empirical work. These are the *median voter* and *probabilistic voting* models. Subsequently, we will turn to the use of these models in normative theorizing. While it will become apparent that neither of them is entirely satisfactory from the perspective presented here, there is much that has been learned about the political economy of taxation from exploration of these approaches.

We begin the discussion of each model with a brief description of its constitutional structure and key behavioral assumptions, and then turn to the question of the existence of equilibrium. We shall see that the manner in which this question is dealt with has a determining influence on the development of the theory.

The Median Voter Model and Its Extensions

One of the first models of the public sector to explicitly incorporate a collective choice mechanism was based on the *median voter theorem* of Duncan Black (1958). This is a model of direct democracy where alternatives to the status quo may be proposed without cost, and in which the institutions of representative democracy do not play an explicit role. Behavioral assumptions for individuals are straightforward: citizens vote sincerely for the policy platform that maximizes their welfare, given the structure of the private economy, and there is no uncertainty about how any voter will behave at the polls.

The model focuses on the problem of coercion made possible by the use of majority rule and aggravated by the separation of spending and taxing.[9] It is successful in explaining coercive redistribution, despite the tendency for redistributive voting games to lack equilibria (see McKelvey, 1976; Schofield, 1978), because the analysis is carried out in a carefully limited framework. With some exceptions, mentioned below, application of the median voter theorem to describe a political equilibrium requires that the fiscal system be reducible to one independent parameter over which (indirect) preferences are single-peaked. Even if preferences are well behaved in more than one dimension, endless cycling over alternative proposals rather than an equilibrium tends to occur, and the model is then of little help in understanding observed tax policy, which exhibits considerable stability over time and place.

A standard model involves a single rate of tax (usually in a linear progressive system) that is chosen by majority rule, and a uniform subsidy or one pure public good on which all revenue is expended. Voting over the average tax rate, or equivalently over the size of the subsidy provided, continues until one rate emerges— the Condorcet winner—that cannot be defeated by any other proposal in a pairwise majority vote. This tax rate and the implied subsidy level maximize the welfare of the median voter—the voter whose preferred tax rate is at the median of those most desired by each voter (see, for example, Romer, 1975; Meltzer and Richard, 1981). If the median voter's income is below the average, the median voter demands and receives a positive tax rate and corresponding subsidy.

The extent of redistribution toward the median voter and the corresponding size of government are limited by behavioral responses to taxation. In the linear income tax case, the equilibrium tax rate, and thus the degree of average tax progressivity, depends on the elasticity of labor supply. As this elasticity increases, more substitution from work to leisure occurs at any rate of tax. This in turn causes a reduction in the aggregate size of the tax base and in the fiscal surplus (the difference between benefits received and the full cost including excess burden of the taxes paid) that

---

[9] Even if there are no public goods, the use of majority rule allows coercion to exist. Separation of spending and taxing, which is necessarily present when public goods are provided, opens up additional routes by which coercion may be exercised.

can be enjoyed by the median voter. In equilibrium, the tax rate demanded by the decisive voter therefore declines.[10]

Extension of the median voter model to a nonlinear income tax system is possible, such as when a second parameter that controls the degree of marginal rate progressivity is added to the average rate of tax. In this case, establishing existence of an equilibrium requires either that further restrictions be placed on the nature of voter preferences (see, Roberts, 1977; Meltzer and Richard, 1981; Gans and Smart, 1996), or it must be assumed that each fiscal parameter is decided by majority rule in a separate "committee" of a legislature or in a separate election in which the median voter is decisive (as in Meltzer and Richard, 1985).[11,12] Then, in addition to the skewness of income as indicated by mean relative to median income, the variance of income also affects the equilibrium size of government. With a higher variance, incomes at the high end of the income scale are even bigger, and this can lead to even more redistribution being demanded by the median voter whose income is below the average (Cukierman and Meltzer, 1991).

A further application of the median voter model allows for private supplementation of publicly provided private goods, such as with healthcare (Gouveia, 1997; Epple and Romano, 1996a, b). When equilibria in the model extended in this way can be established—a difficult issue in this more complex policy setting—an intriguing "ends against the middle" result can be established. It may be that middle income earners who favor a large public sector are opposed by the poor who want lower taxes because they do not value the publicly supplied good highly

---

[10] One longstanding and as yet unanswered question that may be raised at this point is why the extension of the franchise in the nineteenth century to those with lower incomes did not lead to the expropriation of capital through the fiscal system. A possible answer provided by the median voter, as well as by the alternative framework considered later, lies in the negative implications for wealth and income of high taxes on the rich, although this remains a conjecture. On this point, see for example Winer and Rutherford (1992), who explore the argument in a computable equilibrium model calibrated to the U.K. economy in the nineteeth century.

Roemer (2001) constructs an interesting model in which equilibrium is established because of the need of every political party to construct a coalition of members with various interests, all of whom prefer its policies to that of the opposition. The difficulty of maintaining this coalition severely constrains the ability of political entrepreneurs to engineer a winning coalition, regardless of what the opposition proposes. This may be considered an alternative general way of modeling political equilibrium. In such a framework, redistribution can be limited by a party's need to appeal to particular groups of voters—such as the poor and also religious voters, some of whom may be rich.

[11] The restriction on preferences is related to the Mirrlees–Spence single crossing property, so that incomes and abilities of all voters are monotonically related. The application of this kind of restriction to allow another dimension of policy in the median voter model is reviewed in Boadway and Keen (2000). It appears that such restrictions cannot be used to allow a median or decisive voter model to extend to the analysis of the tax skeleton as a whole.

[12] It should be noted that when we assume that decisions on different tax parameters are made sequentially in different elections or committees, each policy parameter must have an independent relationship to welfare in the minds of those involved. Such a procedure requires indirect preferences to be Euclidean (represented by concentric circles), so that an optimal choice for any voter in a given dimension is independent of the choice of policies in other dimensions (Ordeshook, 1992, pp. 283–285).

relative to private consumption, and by the rich who want lower taxes so they can finance even higher levels of the publicly supplied private good than will ever be forthcoming from the public sector. In the equilibrium, the middle income group may win out at the expense of both the poor and the rich.[13]

Before turning to the probabilistic voting framework, one may note that the splitting of dimensions that has sometimes been used to justify extension of the median voter model to multidimensional fiscal systems is an alternative way of establishing a political equilibrium (Shepsle, 1979; Shepsle and Weingast, 1981; Shepsle et al., 1981). Those who adopt this approach relate particular legislative rules and procedures, or norms of behavior (called the "structure"), to the nature and stability of policy outcomes in institutional settings where a vote cycle would otherwise occur.

The approach usually takes one beyond the median voter framework, where a single decisive voter gets what he or she wants. A norm of behavior among politicians on a specific committee of a legislature, for example, may survive because members receive a return to co-operating, inducing them not to vote according to their narrow self-interest. They will do this because they recognize that voting according to broader criteria will eliminate the uncertainty that would result from cycling over alternatives.

A major challenge in using the structure-induced framework lies in identifying the specific institutional arrangements that result in a particular feature of the observed tax skeleton. This difficulty limits the application of the approach to taxation, including applications to nonlinear tax systems. An exception is the work of Inman and Fitts (1990), who use the approach to focus on universalism and reciprocity ("you scratch my back and I will scratch yours") as a norm of behavior that emerges to overcome political instability in legislatures, with each legislator agreeing to support the allocations most preferred by every other member. As long as the benefits of public spending are concentrated within particular electoral constituencies, while the costs are spread by the tax system over the country as a whole, the norm leads legislators to agree to larger budgets and greater use of special tax provisions than would occur if benefits and costs were matched more closely within each district.[14]

Probabilistic Voting

The application of the median voter model points to the role of the skewness of income in determining the extent of coercive redistribution through the fiscal system, as well as to behavioral responses to taxation and the resulting welfare

---

[13] The triumph of the middle class in such a context is often referred to as *Director's Law* (Stigler, 1970).

[14] There are also a few explorations of U.S. tax reform in the structure-induced equilibrium tradition, including Stewart (1991) and McCubbins (1991), who concentrate on the implications of a divided Congress for the politically feasible set of tax proposals.

losses as factors limiting the overall size of government. Because of the manner in which equilibrium is established, the model has little to say about the tax skeleton or tax structure as a whole.

The probabilistic voting model provides a basis for analyzing the tax skeleton as a whole, and it does so by using an approach to the problem of establishing equilibrium that does not require the number of tax instruments to be severely restricted. This is a model of representative democracy in which political parties are forced to compete for votes in order to win the struggle for power. Just as in the median voter framework, the specific institutions that maintain and shape electoral competition are not formally represented in the model.

In addition to differences in constitutional setting, the probabilistic voting model also adopts an alternative approach to political behavior. While economic behavior and the structure of private markets are essentially the same in both frameworks, individual voting, while still sincere, is no longer deterministic, a fact suggested by the model's name. In a probabilistic setting, political parties do not know with certainty how any voter will behave at the polls. This is the key assumption allowing for the possibility that an equilibrium may exist, even if the tax system is multidimensional (Hinich, 1977; Coughlin and Nitzan, 1981; Chen 2000; McKelvey and Patty, 2006).

When voting is strictly deterministic, as in the median voter model, each voter will abruptly switch support from the incumbent to the opposition (or vice versa) if promised a sufficiently favorable policy outcome. The points at which voters switch their support from one party to another become the objects of a bidding war between the parties, leading almost inevitably to vote cycling over alternative platforms. However, if voting behavior is probabilistic, a small change in a policy platform directed at any voter will lead at most to a small change in the probability of support from that voter, not to a total loss of his or her support. If, in addition to the probabilistic nature of voting the objective functions of the parties—total expected votes or expected plurality—are also concave in policy choices for each platform of the opposition, a Nash equilibrium in the electoral game may exist despite the complexity of the fiscal system being decided upon.[15] In this setting, each party is forced by competition to maximize its total expected vote defined across all citizens, and the equilibrium in the model represents a balancing of the heterogeneous and conflicting economic interests of all citizens. Here, every voter, and not just the median voter, has some direct political influence on the equilibrium fiscal system.

It should be noted that if the policies of opposing parties become very polarized, the probability that some radical voters will support the party at the other end of the spectrum may fall to zero. If this happens, the expected vote functions of both parties may not be sufficiently concave over the entire policy space, and a vote cycle may reemerge (Usher, 1994). Thus the instability of majority rule continues to

---

[15] See Enelow and Hinich (1989) and Lin et al. (1999) for a discussion of the conditions underlying the concavity of expected vote functions.

cast a shadow, even in this framework. For this reason, use of the probabilistic voting model implies the important assumption that issues that would lead to extreme polarization of the electorate are kept out of the political arena, thereby limiting the domain over which policy instruments can be defined.

Since it is reasonable to assume that expected support for any party will rise with an increase in expected welfare for any voter, every party has an incentive to adjust the tax mix so as to make the aggregate excess burden of taxation as small as possible, although increases in the welfare cost of taxation will be tolerated if this allows for greater satisfaction of particular, politically sensitive or influential groups. For this reason, the probabilistic voting model is well suited to the study of how the full costs of taxation, including excess burden, are taken into account in determining the nature of the tax skeleton.

Each tax instrument, such as a particular tax base or special provision, will have a different loss of expected votes or political cost associated with it, reflecting factors such as the costs of organizing political opposition and the welfare losses resulting from the economic adjustments to the use of the instrument. Governments that are forced by competition to maximize expected support will therefore aim for a tax skeleton that equalizes the marginal political costs of raising another dollar of revenues across tax sources. This logic will be familiar to those who adopt the optimal tax approach to fiscal design, although such logic must be substantially adapted in the present context. First we must allow for the difference between social welfare and expected political support, and second, we must acknowledge that the task is to characterize a political equilibrium that may or may not be efficient. (We consider the efficiency of political equilibria in the models we are discussing in the section **Normative Analysis**.)

In the probabilistic voting framework it is possible to understand how tax policy instruments may arise endogenously, if we acknowledge that systems that are costly to administer reduce the level of services and subsidies and hence the level of political support that can be obtained with a given total revenue (Hettich and Winer, 1988, 1999; Warskett et al., 1998). Tax bases, rate structures, and special provisions can be explained in this manner. To economize on the administrative costs of actually operating a tax system, governments must group related activities into composite tax bases to lower transaction costs for themselves—the costs of becoming informed about taxpayers, of designing tax structures, and of enforcing tax laws. In a similar manner, they combine taxpayers into rate bands, rather than taxing each individual at a unique rate. However, such grouping creates a loss in expected support, since differentiated treatment of heterogeneous taxpayers would maximize expected political support in a frictionless world. Governments must balance this loss against the gain in support from spending fewer resources for administrative activities and more resources for the provision of public goods.

By extension, similar arguments can also be used to explain the existence of special provisions. If there is a group that offers effective opposition to the inclusion of a specific economic activity in a particular base, it may be cheaper to placate it with a special provision, rather than with the creation of a separate base for the disputed item. Thus, capital gains may become part of a fairly broadly defined

income tax, while being taxed at a rate that differs from the rate applied to other types of income. It should be noted that in this framework, special provisions are a rational response by governments that expect to compete with opposition parties in future elections. They cannot be interpreted as deviations from some ideal tax base designed to satisfy particular normative criteria, which in actuality may have limited support among voters. Nor are they introduced primarily as a hidden substitute for direct subsidies, as is so often argued in the tax expenditure literature.[16] Special tax provisions would exist even in a world where no attempt is made to give direct subsidies to encourage particular activities.

Moreover, since revenue structures are equilibrium outcomes, they should be expected to adjust whenever a significant exogenous shock occurs, such as when some exogenous factor alters the size of a potential tax base and thus changes the economic and political consequences of relying on that tax source.[17] We should therefore expect tax systems to change frequently, although this will not be a sign of political instability or of "tax reform."

The focus on the equilibrium mix of policies also has other important implications for positive tax analysis. For example, it casts doubt on the separate treatment of particular sources of revenue, such as tariffs, debt, or seigniorage, which have often been studied without reference to the rest of the fiscal system. Tariffs are an instrument of protection, but they were also a major source of revenue of the advanced democracies in the nineteenth century and are still important revenue producers in many less developed countries today. In the probabilistic voting framework, the setting of tariffs will involve trade-offs between protection and revenue, as well as trade-offs between tariffs and other sources of revenue (Winer and Hettich, 1991). Similar arguments will also apply to the study of debt or to seigniorage, or for that matter, to the study of other single-revenue sources.

At a more abstract level, the issue raised here concerns the difference between partial and general equilibrium analysis of tax instruments, a distinction made on the left side of Fig. 23.1. Analysis of a part of the whole tax system is often a productive way to proceed, just as limiting the analysis to one private market allows for greater focus and detail. But at the same time, such an analysis must be carried out while remaining cognizant of the broader equilibrium setting.

A further illustration of general equilibrium analysis applied to the tax skeleton that is made possible by probabilistic voting concerns the so-called "flat" tax, which we will interpret for argument's sake as a uniform proportional tax on a single base with only limited exemptions. If special provisions are, indeed, a means

---

[16] The tax-expenditure literature is derived from the work of Henry Simons (1938) who argued for a tax on comprehensively defined income as the mainstay of the tax system. A tax expenditure is a deviation of actual tax payments from tax liabilities that would apply if taxation was levied on this ideal base. We consider Simons' approach to taxation further in the section **Normative Analysis**.

[17] If the size of a potential tax base expands, we may expect the marginal excess burden of relying more heavily on that source to fall relative to the excess burden from using other bases. Political opposition to increasing reliance on the growing base may also decline because the fixed costs of organizing opposition are rising relative to individual excess burdens.

of making the fiscal system politically more efficient, helping to adapt taxation to the characteristics of voters in an administratively effective manner, as was suggested above, it will be unlikely that a policy can succeed that removes this type of policy instrument completely. We may therefore expect democratic tax systems to be complex. While "reforms" can occur that simplify tax laws to some extent, if this becomes a politically popular aim, the result will probably be a fiscal system that retains considerable complexity.

Finally, one should note the implications of taxation for the use of other policy instruments. Any constraint on the use of a particular fiscal instrument, such as imposition of a "flat" tax on income, may lead to the introduction of more special provisions in other tax bases, or to the increased use of policy instruments such as regulation, which can have similar economic effects on voters. Forced simplicity in taxation may thus lead to additional, and perhaps more obscured, complexity in other places.

In summary, the probabilistic framework predicts stable equilibrium outcomes for choices in multiple dimensions. It emphasizes the incentives that governments have to deal with the full costs of taxation, while taking the relative political influence of various groups of taxpayers into account and making it possible to show how the tax skeleton arises endogenously. On the other hand, the model has not been used extensively to study coercive redistribution (although Dixit and Londregan (1996, 1998) and Tridimas and Winer (2005) among others, attempt to deal with this issue), and it lacks specific institutional features and detailed references to actual governing arrangements, a limitation that also affects the median voter model.

Statistical Research

A complete program of work on the political economy of taxation will include statistical modeling and testing of hypotheses in addition to theoretical work. It may also involve the construction and use of computable general equilibrium models. In this and the next section, we complete the coverage of the elements of a comprehensive approach depicted on the left side of Fig. 23.1 by considering how the two approaches have been used to inform empirical research.

Statistical analysis using the median voter and probabilistic voting models can be compared by imposing restrictions on the following system of semireduced form equations:

$$sk = sk\,(s1,\,s2,\,\ldots,\,sk-1,\,\ldots,\,sk+1,\,\ldots,\,sK,\,G,\,x);\; sk \geq 0; k = 1,\,2\ldots,K \tag{23.1a}$$

$$G = G\,(s1, s2,\ldots,sK, x);\; G > 0. \tag{23.1b}$$

Here time subscripts and error terms are omitted, revenue structure $s = \{s1, s2, \ldots, sK\}$ includes all tax bases, rate structures, and special provisions that define the tax skeleton; G is the level of public expenditure; and x is a vector of conditioning

variables including exogenous or predetermined factors relevant to decisions by economic and political agents.

These equations are consistent with a wide variety of models of political equilibrium. They acknowledge that in an equilibrium the use of any policy instrument depends in general on the setting of all other instruments. For example, tax structure depends on how much revenue in total is to be raised, and the reliance on any particular type of tax depends on how much revenue is raised in other ways.

While the equations are quite general, they still omit many aspects of fiscal structure. In particular, the formation of the tax instruments themselves is suppressed as is the structure of public expenditures. Relationships between fiscal instruments and other policies such as regulations and laws are ignored. Moreover, fiscal institutions are not explicitly represented, although their effects will be embedded in the coefficients of the estimating equations and might be included to some extent in the vector of exogenous factors.

To our knowledge, no one has yet estimated such a system to explain a complete tax structure consisting of bases, rates, and special provisions. The problems of doing empirical research with such general systems resemble the difficulties associated with empirical work in any general equilibrium context. In fact, the problems are more acute here than is the case in the study of the private economy since the equilibrium framework must take account of the interaction between the private economy and the political system. In such a setting, it is often useful to proceed by simplifying further, while justifying why some particular part of the larger fiscal system is deserving of special attention.

Researchers who base their work on the median voter model have focused on the implications of coercive redistribution for the overall level of taxation. In this case, the number of fiscal instruments is usually reduced to two (i.e., usually $K = 1$), such as a single proportional tax rate in addition to the overall level of public expenditure. Specific estimating equations are derived by maximization of the median voter's utility subject to the relevant constraints. The vector $\times$ of exogenous variables reflects the median voter's characteristics, such as his or her income, and the factors that determine behavioral responses to taxation.

When K is equal to 1, the government budget restraint will determine one of the two policy variables, and estimation of only one equation has to be carried out (see, for example, Meltzer and Richard, 1983). It should be noted that modeling the average tax rate on a particular base, such as income, rather than modeling an overall average rate, is not a proper empirical application of the median voter model. This must be so unless one believes that coercive redistribution is only exercised via income taxation.[18]

---

[18] Borge and Rattsoe (2004) provide an interesting test of a key implication of the Meltzer/Richard framework—that a more unequal income distribution (as measured by the ratio of mean to median income) will lead to more redistribution and a larger public sector—in the context of a study of local tax structure. They show that more income inequality led to a shift in the Norwegian local public sector from poll taxes towards greater reliance on property taxation, with a consequent rise in the degree of redistribution.

Implementation of the estimating equation requires that the researcher first figure out who the median voter is, and this usually involves additional assumptions so that the median voter can be identified as the person with median income. The ratio of mean to median income is a critical explanatory variable resulting from application of the model, with skewed distributions hypothesized to lead to larger public sectors, and with more elastic behavioral responses expected to offset this tendency.

Another approach to applying the median voter model starts with an assumption that a complex tax structure is fixed independently of public spending. The median voter model is then solved for the level of public expenditure most desired by the median voter as a function of exogenously given tax shares and other factors (Borcherding and Deacon 1972; Bergstrom and Goodman, 1973).

There is an extensive body of empirical work, which we cannot review here, in which the median voter model is used to explain the overall size of government for different political jurisdictions. It is fair to say that over the last three decades, this approach has dominated empirical public choice.

More recently, the probabilistic voting model (or a framework that is informally but closely related) has been applied to model tax systems where K in (1a) and (1b) is equal to 2 or more. In these applications, the instrument set s is usually interpreted as the shares of total revenue coming from several sources such as profits, personal income, consumption, trade, property, seigniorage, and public debt. Instead of including characteristics of the median voter, the vector $\times$ now consists of exogenous factors that determine the marginal political costs associated with relying on each revenue source. These include the factors determining the full economic costs of each tax source, such as the size of potential tax bases, and the factors determining how the full costs of raising revenue in each way are translated into political opposition.[19]

Empirical work of this nature has been conducted by Pommerehne and Schneider (1983) who model the revenue structure of Australian national governments, by Winer and Hettich (1991) for the government of Canada in the nineteenth century, and by Kenny and Winer (2006) for a sample of 100 countries. Some research in this vein considers just one or two parts of the larger equilibrium system. Moomau and Morton (1992), for example, limit themselves to the property tax, Winer and Hettich (2002) look at the relationship between state income taxation and special provisions for the property tax, while Kenny and Toma (1997) examine the choice between income taxation and the inflation tax in the United States. Chernick and Reschovsky (1996) use a partial approach to study determinants of tax progressivity among U.S. states and Sjoquist et al. (2005) consider the relationship between local sales and property taxation. Aidt and Jensen (2007) study the effects of the extension of the franchise on the choice between indirect and direct taxation among European states between 1860 and 1938 and stress the importance of the interaction between tax collection costs and the allocation of voting rights.

---

[19] See Hettich and Winer (1999, Ch. 8) and Kenny and Winer (2006) for further details.

Some Evidence Concerning the Role of Institutions

The role of political institutions is only implicit in the empirical work described above, as it is in the theoretical models that underlie these applications. By estimating reduced form equations across electoral systems, or by doing analogous case studies, interesting stylized facts about the role of institutions can be generated.

There is a growing body of work of this kind, much of it dealing with the consequences of alternative electoral systems for the overall level and composition of spending (see, Kirchgaessner, 2002; Gould and Baker, 2002 for reviews; and Aidt et al., 2006; Blume et al., 2007; Persson and Tabellini, 2001, 2003; Milesi-Ferretti et al., 2002 for recent contributions). As yet, few studies relate electoral systems or other aspects of governance to specific features of the tax system except at an aggregate level. However, existing research points to future directions for work applied more directly to taxation.

Of particular interest is a branch of the literature that investigates the relationship between electoral systems or structural characteristics of government and overall fiscal discipline. Persson and Tabellini (2003), for example, show that government spending is lower in countries with majoritarian as opposed to proportional electoral systems, a result confirmed by Blume et al. (2007).

Using data on OECD countries, Ashworth and Heyndels (2002) investigate how volatility in tax systems is affected by the degree to which government is fragmented, while Volkerink and de Haan (2000) ask similar questions with regard to reliance on deficit financing. Fragmentation is measured by the number of decision makers involved in fiscal decisions or by the number of parties in a governing coalition. Their studies show that fragmented governments tend to have tax systems that exhibit more persistence in the face of exogenous shocks, and larger deficits.

The effects of legislative rules and laws for ensuring that, at an aggregate level at least, spending is kept in line with revenues have also been investigated. Many of these studies, reviewed in Kirchgaessner (2002) and Poterba (1997), use data from U.S. states. It appears that balanced-budget rules and other types of limitations have to some extent been successful in linking spending to available revenues and in inducing somewhat more rapid fiscal adjustments.

There is also some statistical evidence concerning the role of specific policy processes in linking spending and taxing. In an analysis of European Union countries, von Hagen (1992) finds that overall fiscal discipline is stronger where there is a top-down budgetary process run by a strong prime minister, and where parliament has limited powers of amendment.[20]

---

[20] But see also Bohn and Inman (1996). Breton (1996) models the power of a prime minister in a parliamentary system to control spending and compares this power to that of the president in a U.S. congressional system of checks and balances. He suggests that a strong prime minister backing a strong minister of finance coupled with traditions of budgetary secrecy and cabinet solidarity combine to offer distinct advantages for maintaining the overall balance of spending and taxing.

Finally, there is some exploratory work on institutions, which bears on the nature of the tax skeleton. Hettich and Winer (1999, Ch. 11) use descriptive statistics to show that the Canadian tax system is less complicated and involves more frequent major reforms than that of the United States, a result they explain with the greater transactions costs facing politicians in the congressional system, characterized by checks and balances, than in the Canadian parliamentary setting. And, Steinmo (1993) uses the case study method in an interesting attempt to relate stylized differences in tax structures among Sweden, the United Kingdom, and the United States to differences in their political systems. He argues, for example, that the fragmented political authority inherent in the U.S. system of checks and balances has led to a complex and inefficient tax structure there, while Sweden's corporatist model fostered a regressive tax system that is stable and efficient, one which delivers heavy social spending to working class voters in exchange for heavy taxation.

This is an evolving literature. It would be of much interest if research of this nature could be grounded in the application of formal structural models in which the tax skeleton is represented. To accomplish this is particularly difficult for electoral systems based on proportional representation because of the well-recognized problem of modeling the post-election bargaining among prospective coalition members that affects final equilibrium policy outcomes.[21]

Computable Equilibrium Modeling and the Representation Theorem

Another way to further our understanding of taxation is by constructing an applied or computable equilibrium model that can be used for simulation. Rather than being estimated econometrically, these models are calibrated either to synthetic or real data sets for specific jurisdictions at a point in time.

In applying the median voter framework, one must specify how the private economy depends on the tax instrument or size of government that is determined in political equilibrium. (The structure must be such as to ensure that voters' preferences are single-peaked over the relevant policy instrument.) Public policy is chosen so that, given the relationship between the median voter's well-being and the private economy, the median voter's welfare is maximized. When a computable model of a federal system is constructed and the median voter model is applied at

---

[21] Austen-Smith (2000) builds on the median voter model, and on models of agenda control by Romer and Rosenthal (1978) to formally compare the average rate of tax in an electoral system with a first-past-the-post (FPTP) electoral system to that in a three-party system with proportional representation (PR). Austen-Smith's study is motivated by the desire to model the observation that average income tax rates appear to be higher and post tax distributions of income flatter in countries with PR than in countries with FPTP. The key to his explanation appears to lie in differences between electoral systems regarding the decisive voter and the incentives created for taxpayers to choose among available occupations.

each level, it is necessary to assume that voters make their decisions about whom to support in each election without considering the consequences for policy at other levels of government.

Nechyba (1997, 1999) explores various issues, including the setting of property tax rates and the effects of school vouchers, in a large-scale median voter model of the relationship between state and local governments. The model allows for tax competition between cities and interjurisdictional migration and is calibrated to data for New Jersey. Voting decisions at each level of government are assumed to be independent in the minds of the voters, and at each level the median voter is decisive. Holtz-Eakin (1992) has constructed a synthetic political economy in order to compare the results of various experiments based on the median voter theorem to results when a probabilistic voting approach is used with the same data.

In a computable version of a probabilistic voting model, what is optimized by the choice of (several) policy instruments is a political support function defined across all voters, rather than the median voter's utility. This technique is illustrated at some length below. Work of this sort includes Rutherford and Winer (1999, 2002) and Hotte and Winer (2001), who use the model to work out the effective political influence that must be imputed to each of several groups of voters so that the model replicates a benchmark fiscal system (the U.S. rate of tax on labor, capital, and the size of government in 1973 and 1983). These weights are then used to construct counterfactuals that allow changes in the benchmark system over time to be decomposed into a part due to changing economic structure and a part due to changes in relative political influence. Since it will be useful in the next section where normative issues are considered, we illustrate the derivation of the optimization problem referred to above that can be used to compute an equilibrium in a probabilistic model. This derivation is based on the work of Coughlin and Nitzan (1981).

To simplify, we limit the discussion to a situation with two political parties, two tax bases, two tax rates, and one public good. To acknowledge tax administration and information costs implicitly, we assume that the number of tax rates is less than the number of voters and that taxation is proportional rather than lump sum. Indirect utility for voter h is $vh(t1, t2, G)$ and, after substitution of the general equilibrium structure of the private economy, the government budget restraint can be written as

$$G = R1(t1, t2, G) + R2(t1, t2, G).$$

Each party chooses tax rates and the size of public expenditure to maximize its total expected vote. The probability that voter h supports the incumbent as perceived by the party, fhi, depends on the difference in the voter's evaluation of his or her welfare under the incumbent's policies (i) and those of the opposition (o): $fhi = fh(vhi - vho)$. The expected vote for the incumbent government then is $EVi = 3hfh(vhi - vho)$, and the vote for the opposition is defined analogously. In addition, we assume that knowledge of the probability density functions describing voting

behavior and of the structure of the private economy is common to the competing parties.

Given the platform of the opposition, first-order conditions for the choice of tax rates that maximize EVi subject to the budget restraint are of the form

$$\frac{\sum \partial f_h/\partial v_h \cdot \partial v_h/\partial t_1}{\partial(R_1 + R_2)/\partial t_1} = \frac{\sum \partial f_h/\partial v_h \cdot \partial v_h/\partial t_2}{\partial(R_1 + R_2)/\partial t_2}. \tag{23.2}$$

From (23.2) it can be seen that the platform chosen by the incumbent equalizes the marginal effect of tax policies on expected votes per dollar of revenue across tax sources. The condition illustrates the equalization of "marginal political costs" across tax instruments referred to earlier. A Nash equilibrium, if it exists, is a simultaneous solution to such first order conditions for both incumbent and opposition parties.[22]

After substitution of equilibrium values of the partial derivatives in (1), the resulting condition can also be used to characterize the tax system that emerges in a Nash equilibrium. Let $\theta h = Mf_h/Mv_h$ be the particular values at a Nash equilibrium of the sensitivities of voting to a change in welfare, and let the other partial derivatives also be evaluated at the equilibrium. Then the first-order conditions characterizing optimal equilibrium strategies take the form:

$$\frac{\sum \theta_h \cdot \partial v_h/\partial t_1}{\partial(R_1 + R_2)/\partial t_1} = \frac{\sum \theta_h \cdot \partial v_h/\partial t_2}{\partial(R_1 + R_2)/\partial t_2}. \tag{23.3}$$

Now it can be seen that this equilibrium condition may be replicated by solving a particular optimization problem. It is straightforward to show that maximization of the following "political support function" (S) by choice of the same policy instruments, subject to the same government budget constraint, leads to the identical condition:[23]

$$S = 3h\theta hvh. \tag{23.4}$$

The use of this optimization problem to compute a political equilibrium constitutes what we shall call the *Representation Theorem*.[24] Note that since S is maximized in a political equilibrium, it makes sense to think of the weights θh in the support

---

[22] Neither the existence of an equilibrium nor the convergence of platforms that occurs in this version of the model is guaranteed under all conditions.

[23] Second-order conditions sufficient to ensure the existence of a constrained maximum must also be satisfied. For further details, see Coughlin (1992) and Hettich and Winer (1999).

[24] Note also that the support function S is not a social welfare function. The weights in S are determined *within* the model by voting behavior, and different types of voting behavior or different political institutions will give rise to different support functions. On the role here of political institutions, see Chen (2000).

function as measures of the effective influence exerted by different voters on equilibrium policy outcomes.

As well as permitting the probabilistic voting model to be operationalized, the Representation Theorem has important implications for the normative evaluation of tax systems.

Normative Analysis

In our initial discussion of the elements depicted in Fig. 23.1, we pointed out that a fully general approach would have a normative as well as a positive dimension. Although there is an extensive literature using collective choice models as a basis for positive theoretical and empirical research, there is only a limited body of work on how to explicitly link them to normative questions. Filling in the boxes on the right side of Fig. 23.1 remains a challenging task. In this section we consider some of the issues involved in using the median voter and probabilistic voting models to do so. We also briefly consider some other contributions to the normative literature in the light of the comprehensive approach to political economy.

In normative analysis, we evaluate imperfect situations by comparing them to a state that has defined optimal properties. Three steps are needed in this kind of work. To start with, one must define a counterfactual or standard of reference representing an optimal allocation of resources. The underlying theoretical analysis must prove that this allocation exists and that it is a stable equilibrium outcome of a relevant or acceptable collective choice process. (It should be recalled that public goods and the corresponding taxes cannot be allocated or distributed without recourse to a collective choice process.)

Once this has been accomplished, a second step becomes possible. Imperfect situations can be contrasted with the socially optimal, democratically arrived at allocation. Finally, the loss in welfare resulting from the imperfect operation of the decision process is measured in monetary terms.

The three steps are well known from neoclassical welfare economics relating to competitive markets, where the first theorem serves to define the ideal counterfactual or standard of reference, and where the second step is represented by the analysis of market failure. In a final step, the implications of imperfect markets are then measured by quantifying the resulting welfare losses.

Although the same sequence of steps must be followed in a normative analysis of taxation that includes collective choice as a significant component, there are important differences of interpretation. Since the relevant equilibria must now refer to a political process, the counterfactual, as well as the analysis of imperfections, must refer to the working of political mechanisms rather than to the operation of private markets. Thus, we are interested in the identification and measurement of the consequences of *political* market failure. This involves identifying the sources of such failures, and then relating such failures to specific identifiable parts of tax

structure that are undesirable as a result. Such a political market failure analysis of tax policy remains to be accomplished.[25]

The importance of establishing a normative analysis that includes collective choice in such a systematic manner can be better understood if we use it to evaluate a well-known result derived from the social planner model where politics plays no role. The latter approach has been widely used to argue for tax policies that minimize excess burdens measured in relation to lump sum taxation. To achieve such minimization, it is necessary to adjust the tax system so as to equalize the marginal excess burden created by raising an additional dollar of revenue across all different tax sources.

The limitations of this sort of policy recommendation can be seen clearly if we ask the same questions in a framework based on a collective choice model such as probabilistic voting. In a probabilistic voting model, political competition tends to force parties to adopt Pareto-efficient policies. Otherwise, the possibility remains that the opposition can propose a Pareto-improving policy platform and thereby increase its expected electoral support. This tendency is readily apparent from the Representation Theorem stated in the previous section, which shows that under certain conditions, the equilibrium can be replicated by maximizing a particular weighted sum of utilities subject to the general equilibrium structure of the economy.[26]

This does not imply, however, that marginal excess burdens per dollar, or marginal efficiency costs, will be equalized across tax sources, thereby minimizing total excess burden. The reason is that voters differ in their effective political influence even when the franchise is universal. Hence, in directing resources towards voters who the governing party thinks are especially influential, the incumbent party will accept an increase in the marginal efficiency cost of a particular tax source above that of other taxes if it thinks this will improve the chances for reelection. We can clarify this point by continuing with the development of the model introduced in the section **Computable Equilibrium Modeling and the Representation Theorem**. Using condition (23.3) in the special case where the equilibrium political influence weights (the θh's) for all voters are

---

[25] Political failures may occur as a result of lobbying and the use of advertising to sway voters, or in a dynamic context where the problem is to ensure the consistency of policies over time. Contributions on the first aspect of political economy are reviewed in Grossman and Helpman (2001) and on the second in Drazen (2000, Ch. 4–6). The link to specific features of tax systems in this literature remains to be more fully developed. In this regard, see also footnote 25 below. Political market failure may also result from unregulated tax competition between jurisdictions: the relevant literature is reviewed by Wilson (1999).

[26] We have implicitly assumed that political competition is perfect, in the sense that parties must continually maximize expected votes, and that no one can systematically influence voters with advertising misrepresenting how policies will affect individual welfare. If these conditions are not met, the equilibrium will not be efficient, and the optimization problem that is used to replicate the equilibrium will be different from what has been stated above. See Hettich and Winer (1999, Ch. 6). Other situations may also lead to political market failure.

equal, we can substitute the change in aggregate welfare defined by $\Delta Wk = 3h\mathrm{M}vh/\mathrm{M}tk$ into (23.3), subtract 1 from each side, and rearrange to get

$$\frac{\Delta W_1 - \partial(R_1 + R_2)/\partial t_1}{\partial(R_1 + R_2)/\partial t_1} = \frac{\Delta W_2 - \partial(R_1 + R_2)/\partial t_2}{\partial(R_1 + R_2)/\partial t_2}. \qquad (23.5)$$

Here the numerator on each side of the equation is the marginal excess burden of the corresponding tax change—the change in welfare over and above the change in revenue—while the quotient on each side of (23.5) represents the marginal efficiency cost of each tax source.

Thus if the θ's are all equal, the tax system equalizes the marginal costs per dollar across tax sources and hence minimizes the total excess burden of taxation. On the other hand, if political influence is distributed unequally as in (23.3) and (23.4), unweighted marginal welfare losses for different tax sources may vary significantly as parties trade off the welfare of and support from different voters, even though Pareto efficiency is being achieved.

In other words, by weighting welfare changes for different people equally, traditional normative analysis imputes all observed inequality of marginal efficiency costs to the inefficiency of tax policy. This is an extreme view, given the existence of vigorous political competition for the support of rational economic agents.[27] Even if we allow for the existence of political market failures, which we have not done here, at least some part of the inequality of marginal efficiency costs in equilibrium will still be due to the pursuit of support from voters who differ in their effective political influence. (What part of the inequality is actually due to political market failure is not known, and little studied.) Moreover, proceeding as if the marginal efficiency costs should be equalized when political influence actually differs across groups of voters may lead to reforms that only serve to move society along or possibly even inside of the Pareto-efficiency frontier.[28]

---

[27] An important general lesson here is that normative analysis that is not informed by a model of political equilibrium is likely to be misleading. Another interesting example of this is provided by the literature on the time consistency of public policy. A policy is not time consistent if it requires a course of action today (about today and tomorrow) that will subsequently become undesirable. It is often argued that the inability of governments to commit to consistent policy over time will result in a loss of social welfare compared to a situation where governments are prevented from adopting discretionary policies based on period by period political optimization (see, for example, Fischer, 1980). The problem with this and similar arguments is that it does not allow for the constraints on misuses of discretionary power that exist in a democracy (Hettich and Winer, 1985; Marceau and Smart, 2002). People in democratic societies are not powerless in opposing unwanted government actions using normal political channels. Moreover, we may also find the legal system being altered to make it difficult for governments to unilaterally expropriate private property. As a result, it is not obvious that further restrictions on the ability of governments to respond to changing events, which must have social costs as well as benefits, are warranted to counteract possible time inconsistency.

[28] For related but different arguments concerning the problems of doing welfare analysis without taking political equilibrium into account, see Coate (2000) and Besley and Coate (2003).

Cost-benefit analysts have long recognized the problem of determining the proper direction of reform when the weights attached to various groups of people are not equal.[29] They have tried to infer such *distributional weights* (as they are called in this literature) from existing data and to use them in aggregating losses and gains for different groups.[30] Whether weights derived from existing political equilibria, which may be imperfect, are appropriate for normative analysis is unclear.

There is as yet no consensus on what institutional characteristics of the voting process would be required to yield an ideal outcome, or on what weights would be embedded in the equilibria arising in such a system. As a result, definition of a counterfactual ideal and measurement of losses as a consequence of political market imperfections remain unsolved analytical problems in the approach based on probabilistic voting, and in related or similar approaches.

What is the nature of normative analysis in work based on the median voter framework? While the probabilistic voting model emphasizes the problems of reconciling conflicting, and heterogeneous interests, the median voter model draws our attention primarily to the consequences of coercion under majority rule.

There is an analog to the role of the Representation Theorem in normative work based on the median voter model. It involves the demonstration that total revenue, or the aggregate tax rate, is efficient in equilibrium under certain special circumstances. The question of what the required conditions are has been extensively explored, with rather discouraging results (Bowen, 1943 and Bergstrom, 1979). Efficiency of the public sector in this world will occur only in the special and rather unlikely case where preferences are symmetrically distributed around those of the median voter, so that the consequences of coercion for the welfare of voters on either side of the median just balance out.

Individual preferences are usually taken as given and inviolable. So it would be understandable if a policy analyst in search of efficiency, who based his analysis on the median voter model, were led to propose changes in the basic voting rule, rather than in particular policies, as a way of improving the allocation of resources.

Limiting Majority Rule

Proposals for reform of the basic voting rule have a long history in the literature on taxation. There have also been several proposals to constrain the use of policy instruments as a way of indirectly limiting the exercise of coercion. Such proposals are not usually associated with either of the two formal models we have been analyzing. Nonetheless, to complete the discussion of normative tax analysis, it is of interest to briefly consider some of this work. The discussion will also point to

---

[29] See, for example, the text by Boardman et al. (1996, Ch. 2).

[30] Rutherford and Winer approached this issue by calibrating the weights so that maximization of the support function replicated the benchmark equilibrium.

the difficulties of normative theorizing without the use of a comprehensive framework.

As we noted earlier, Wicksell (1896) advocated the adoption of a qualified or approximate unanimity rule to limit coercive redistribution through the public sector. Of course, he did not use the median voter model as a basis for his proposal. But he clearly understood the essential link between collective choice and the allocation of resources, and realized the dangers that are inherent in majority rule. He proposed an institutional solution in his perceptive analysis of what would be required to generate a more efficient political equilibrium in a democratic society.

Wicksell's analysis is an example of a "process-oriented" approach to reform. His analysis does not include a blueprint for tax structure, and is confined to reform of the policy process. A concern with the coercive power of government also lies behind more recent process-oriented proposals. These involve detailed tax blueprints, the purpose of which is to make it difficult for democratic governments to engage in coercive actions while still permitting them to finance needed public services.

Simons (1938) was a successful advocate of a process-oriented approach to restricting the power of government to coerce private citizens. He was not primarily concerned with coercive redistribution between rich and poor, and was content to leave the determination of vertical equity to the political process. He argued, instead, for a tax levied on comprehensively defined income as a way of limiting the ability of governments to interfere in private markets (or, as he put it, to "dip deeply into great incomes with a sieve"). Buchanan and Congleton (1998) have recently proposed a flat tax without exemptions, based on concerns similar to those expressed by Wicksell and Simons.

Normative tax theory after 1945 was dominated by discussion of the comprehensive income tax system advocated by Simons, until Optimal Taxation replaced his approach in the early 1970s. Simons' work also stimulated several important tax commissions during the period. This occurred even though the political foundations of Simons' argument for the comprehensive income tax were not generally appreciated.[31]

While not clearly connected to a formal model of political equilibrium, the arguments of Simons and Buchanan and Congleton carry with them a statement of what the ideal tax system should look like. As a result, they allow identification of the parts of existing tax systems that are undesirable, and measurement of departures from the ideal then becomes possible. These are key steps in a comprehensive normative analysis of taxation.

However, there is a serious flaw in the design of the tax blueprints advocated by those concerned with the coercive power of government. These proposals are at odds with the understanding of political equilibrium developed using the probabi-

---

[31] See Hettich and Winer (1985, 1999) for review of the relevant literature on this point. It is of interest to note that measurement of deviations from the broadly based personal income tax, following Surrey (1973) and others, is the basis of tax-expenditure budgets, which have even been enshrined into law in some countries.

listic voting approach. In this framework, political competition creates pressures for any government to implement a tax system that is, to some extent, adapted to deal with excess burdens. As we have already seen, competition in such a political system pushes the government to implement a complicated tax skeleton which is unlikely to resemble the fiscal structures advocated by Simons or by Buchanan and Congleton.

Conclusion

There has been much work in the past two decades that approaches public sector problems from a political economy perspective. This is true for issues relating to taxation as well as for topics touching on other aspects of the public economy. Most of this research has not been part of a broad, comprehensive framework of the sort outlined in Fig. 23.1, however. Authors have mostly focused on one specific aspect or problem, and have used a particular collective choice model to deal with a question or topic of limited scope.

We have shown in this chapter that public sector analysis can be carried out as part of a comprehensive theoretical framework. Although the discussion is concerned primarily with taxation, it has implications for all research on the public economy that acknowledges the necessity for collective choice. A truly general framework will allow the analysis of positive as well as of normative questions and will link the two areas of inquiry in a meaningful fashion. While most individual studies will continue to focus on some particular aspect of the government sector, their implications for related questions can be better understood when they are evaluated against the background of an inclusive approach.

Taxation is a crucial topic in public finance. It touches directly on the need for a collective choice mechanism, and it involves analysis of coercive redistribution arising from the use of majority rule. In addition, it requires an understanding of how tax systems are structured to deal with the welfare losses stemming from the separation of spending and taxing, a separation that arises from the very nature of public goods.

While using collective choice models to examine these issues, we must confront the theoretical problems related to existence and stability of equilibrium. Otherwise, predicted policy outcomes may be only transitory phenomena, unsuitable for comparative static analysis, the method of research that has provided the logical underpinning of most work in economics. Moreover, equilibrium must now include political as well as economic forces.

In this chapter, we have focused on the two main collective choice models that have been used to examine taxation, namely the median voter model and probabilistic voting. In each case, we examined the nature of equilibrium analysis, along with the contributions of the model to the understanding of major fiscal issues. Although both approaches have given rise to extensive literatures from which many useful insights can be derived, our review shows that probabilistic voting is able to

encompass a wider range of questions. In particular, this model allows for the examination of both positive and normative questions, while the median voter model has little to say on the efficiency of taxation. Regarding theoretical and empirical research, median voter analysis has provided a strong focus for the examination of coercive redistribution, but it has not proved suitable for the study of tax structure and tax design owing to its limited success in dealing with multidimensional issues. Probabilistic voting provides an appropriate basis for studying the nature of observed tax systems, and it can, at least in principle, also be used to shed light on coercive redistribution.

Both models still fall short of integrating into the analysis the wealth of existing fiscal institutions within which the exchange between citizens and governments occurs. While there is work on fiscal institutions, it is largely limited to linking them to the aggregate level of spending or to attempts to control budget deficits. Research in this area only rarely deals with specific features of observed tax systems or fiscal structure. A framework that encompasses taxation as an instrument of coercive redistribution, that can explain the tax skeleton and its relationship to excess burden, that accounts for the role of administration costs, and that assigns an explicit role to fiscal institutions, remains to be constructed.

Whatever approach is chosen in future work, the nature of equilibrium remains crucial. If it is ignored, analytical results may be doomed to irrelevance or disregard in the political arena, a fate that has befallen a large number of proposals for a comprehensive income or consumption tax or a generalized flat tax. Advocates of such taxes have never demonstrated that they represent equilibrium outcomes of an acceptable and democratic collective choice process. Research based on probabilistic voting strongly suggests that democratic regimes will inevitably create complex tax systems with multiple bases, varied rate structures, and a myriad of special provisions.

Although a comprehensive approach remains to be fully developed, consideration of existing work against the background of a generalized framework helps in seeing the strengths and weaknesses of available models and is useful in guiding the researcher in future work. It also makes clear that much has already been accomplished, and that the collective choice literature devoted to taxation is a rich and valuable source of analytical and policy-relevant insights.

References

Aidt, T. S., J. Dutta, and E. Loukoianova. (2006). "Democracy comes to Europe: franchise expansion and fiscal outcomes 1830–1938." *European Economic Review*, 50(2), 249–283.

Aidt, T. S. and P. S. Jensen. (2007). "Tax structure, size of government and the extension of the voting franchise in Western Europe, 1860–1938." Cambridge Working Papers in Economics 0715, Faculty of Economics, University of Cambridge.

Ashworth, J. and B. Heyndels. (2002). "Political fragmentation and the evolution of national tax structures in the OECD." *International Tax and Public Finance*, 8(3), 377–393.

Atkinson, A. B. and J. E. Stiglitz. (1980). *Lectures on Public Economics*. New York, USA: McGraw-Hill.

Austen-Smith, D. (2000). "Redistributing income under proportional representation." *Journal of Political Economy*, 108(6), 1235–1269.

Bergstrom, T. C. (1979). "When does majority rule supply public goods efficiently?" *Scandinavian Journal of Economics*, 81, 216–226.

Bergstrom, T. C. and R. P. Goodman. (1973). "Private demands for public goods." *American Economic Review*, 63(3), 280–296.

Besley, T. and S. Coate. (2003). "On the public choice critique of welfare economics."*Public Choice*, 114, 253–273.

Black, D. (1958). *The Theory of Committees and Elections*. Cambridge, UK: Cambridge University Press.

Blume, L., J. Mueller, S. Voigt and C. Wolf. (2007). "The economic effects of constitutions: replicating -and extending- Persson and Tabellini." CESifo Working Paper 2017, June.

Boadway, R. and M. Keen. (2000). "Redistribution," in A.B. Atkinson and F. Bourguignon (eds.) *Handbook of Income Distribution*, Vol. 1. North-Holland: Elsevier, pp. 677–790.

Boardman, A. E., A. R. Vining, and D. L. Weimer. (1996). *Cost Benefit Analysis: Concepts and Practice*. New Jersey, USA: Prentice-Hall.

Bohn, H. and R. P. Inman. (1996). "Balanced-budget rules and public deficits: evidence from the U.S. States." *Carnegie-Rochester Conference Series on Public Policy*, 45, 13–76.

Borcherding, T. F. and R. T. Deacon. (1972). "The demand for the services of non-federal governments." *American Economic Review*, 62, 891–901.

Borge, L. E. and J. Jorn Rattsoe. (2004). "Income redistribution and tax structure: empirical test of the Meltzer-Richard hypothesis." *European Economic Review*, 48(4), 805–826.

Bowen, H. R.. (1943). "The interpretation of voting in the allocation of economic resources." *Quarterly Journal of Economics*, 58, 27–48.

Brennan, G. and J. Buchanan. (1980). *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. New York: Cambridge University Press.

Breton, A. (1996). *Competitive Governments: An Economic Theory of Politics and Public Finance*. New York: Cambridge University Press.

Buchanan, J. M. (1968). *The Demand and Supply of Public Goods*. New York: Rand McNally.

Buchanan, J. M. (1976). "Taxation in fiscal exchange." *Journal of Public Economics*, 6, 17–29.

Buchanan, J. M. and R. D. Congleton. (1998). *Politics by Principle, Not Interest: Towards a Nondiscriminatory Democracy*. New York: Cambridge University Press.

Buchanan, J. M. and R. E. Wagner. (1977). *Democracy in Deficit: The Political Legacy of Lord Keynes*. New York: Academic Press.

Chen, Y. (2000). "Electoral systems, legislative process, and income taxation." *Journal of Public Economic Theory*, 2(1), 71–100

Chernick, H. and A. Reschovsky. (1996). "The political economy of state and local tax structure," in G. Pola, G. France, and G. Levaggi (eds.) *Developments in Local Government Finance*. Cheltenham, UK: Edward Elgar, pp. 253–272.

Coate, S. (2000). "An efficiency approach to the evaluation of policy changes." *Economic Journal*, 110(463), 437–455

Coughlin, P. (1992). *Probabilistic Voting Theory*. New York: Cambridge University Press.

Coughlin, P. and S. Nitzan. (1981). "Electoral outcomes with probabilistic voting and Nash social welfare maxima." *Journal of Public Economics*, 15, 113–121.

Creedy, J. (1998). *Measuring Welfare Changes and Tax Burdens*. Cheltenham, UK: Edward Elgar.

Cukierman, A. and A. H. Meltzer. (1989). "A political theory of government debt and deficits in a neo-Ricardian framework." *American Economic Review*, 79(4), 13–732.

Dixit, A. and J. Londregan. (1998). "Ideology, tactics, and efficiency in redistributive politics." *Quarterly Journal of Economics*, 113, 497–529.

Dixit, A. and J. Londregan. (1996). "The determinants of success of special interests in redistributive politics." *Journal of Politics*, 58(4), 1132–1155

Drazen, A. (2000). *Political Economy in Macroeconomics*. Princeton, NJ: Princeton University Press.

Edgeworth, F. Y. (1925). *Papers Relating to Political Economy*. London: Macmillan.

Edwards, J. and M. Keen. (1996). "Tax competition and leviathan." *European Economic Review*, 40, 113–134.

Enelow, J. and M. Hinich. (1989). "A general probabilistic spatial theory of elections." *Public Choice*, 61(2), 101–113.

Epple, D. and R. E. Romano. (1996a). "Ends against the middle: determining public service provision when there are private alternatives." *Journal of Public Economics*, 62, 297–325.

Epple, D. and R. E. Romano. (1996b). "Public provision of private goods." *Journal of Political Economy*, 104(1), 57–84.

Fischer, S. (1980). "Dynamic inconsistency, cooperation and the benevolent dissembling government." *Journal of Economic Dynamics and Control*, 2, 93–107.

Foley, D. (1977). "Resource allocation and the public sector." *Yale Economic Essays*, 7(1), 45–98.

Gans, J. S. and M. Smart. (1996). "Majority voting with single-crossing preferences." *Journal of Public Economics*, 59, 219–237.

Gillespie, W. I. (1991). *Tax, Borrow and Spend: Financing Federal Spending in Canada, 1867–1990*. Ottawa: Carleton University Press.

Gould, A. C. and P. J. Baker. (2002). "Democracy and taxation." *Annual Review of Political Science*, 5, 87–110.

Gouveia, M. (1997). "Majority rule and public provision of a private good." *Public Choice*, 93, 221–224.

Grossman, G. M. and E. Helpman. (2001). *Special Interest Politics*. Cambridge, MA: MIT Press.

Head, J. (1974). *Public Goods and Public Welfare*. Durham: Duke University Press.

Hettich, W. and S. L. Winer. (1985). "Blueprints and pathways: the shifting foundations of tax reform." *National Tax Journal*, 38(4), 423–445.

Hettich, W. and S. L. Winer. (1988). "Economic and political foundations of tax structure." *American Economic Review*, 78, 701–713.

Hettich, W. and S. L. Winer. (1999). *Democratic Choice and Taxation: A Theoretical and Empirical Analysis*. New York: Cambridge University Press.

Hettich, W. and S. L. Winer. (2002). "Rules, politics and the normative analysis of taxation," in R. Wagner and J. Backhaus (eds). *Handbook of Public Finance*. Dordrecht: Kluwer Academic Publishers.

Hinich, M. (1977). "Equilibrium in spatial voting: the median voter result is an artifact." *Journal of Economic Theory*, 16, 208–219.

Holtz-Eakin, D. (1992). "Elections and aggregation: interpreting econometric analyses of local governments." *Public Choice*, 74(1), 17–42.

Hotte, L. and S. L. Winer. (2001). "Political influence, economic interests and endogenous tax structure in a computable equilibrium framework: with application to the United States, 1973 and 1983." *Public Choice*, 109(1), 66–99.

Inman, R. P. and M. A. Fitts. (1990). "Political institutions and fiscal policy: evidence from the U. S. historical record." *Journal of Law, Economics and Organization*, 6(Special Issue), 79–132.

Johansen, L. (1965). *Public Economics*. Amsterdam: North Holland Press.

Kau, J. B. and P. H. Rubin. (1981). "The size of government." *Public Choice*, 37(2), 261–274.

Kenny, L. W. and M. Toma. (1997). "The role of tax bases and collection costs in the determination of income tax rates, seigniorage and inflation." *Public Choice*, 92(1), 75–90.

Kenny, L. W. and S. L. Winer. (2006). "Tax systems in the world: an empirical investigation into the importance of tax bases, collection costs, and political regime." *International Tax and Public Finance*, 13(2/3), 181–215.

Kirchgaessner, G. (2002). "The effects of fiscal institutions on public finance: a survey of the empirical evidence," in S. Winer and H. Shibata (eds.) *Political Economy and Public Finance: The Role of Political Economy in the Theory and Practice of Public Economics*. Cheltenham, UK: Edward Elgar.

Levi, M. (1988). *Of Rule and Revenue*. Los Angeles: University of California Press.

Lin, T. M., J. M. Enelow, and H. Dorussen. (1999). "Equilibrium in multicandidate probabilistic spatial voting." *Public Choice*, 98, 59–82.

Lindahl, E. (1919). "Just taxation: a positive solution" (translation by Elizabeth Henderson), in R.A. Musgrave and A.T. Peacock (eds.) *Classics in the Theory of Public Finance*. London: Macmillan (1958), pp. 168–176.

Marceau, N. and M. Smart. (2002). "Corporate lobbying and commitment failure in capital taxation." CESifo Working Paper 676, February.

McCubbins, M. D. (1991). "Party politics, divided government, and budget deficits," in S. Kernell (ed.) *Parallel Politics: Economic Policy Making in the United States and Japan*. Washington DC: The Brookings Institution, pp. 83–118.

McKelvey, R. D. (1976). "Intransitivities in multidimensional voting models and some implications for agenda control." *Journal of Economic Theory*, 12(3), 472–482.

McKelvey, R. D. and J. W. Patty. (2006). "A theory of voting in large elections." *Games and Economic Behavior*, 57(1), 155–180.

Meltzer, A. H. and S. F. Richard. (1981). "A rational theory of the size of government." *Journal of Political Economy*, 89, 914–927.

Meltzer, A. H. and S. F. Richard. (1983). "Tests of a rational theory of the size of government." *Public Choice*, 41(3), 403–418.

Meltzer, A. H. and S. F. Richard. (1985). "A positive theory of in-kind transfers and the negative income tax." *Public Choice*, 47, 231–265.

Milesi-Ferretti, Gian Maria, R. Perotti, and M. Rostagno. (2002). "Electoral systems and public spending." *Quarterly Journal of Economics*, 117(2), 609–657.

Mirrlees, J. A. (1971). "An exploration in the theory of optimum income taxation." *Review of Economic Studies*, 38, 175–208.

Moomau, P. H. and R. B. Morton. (1992). "Revealed preferences for property taxes: an empirical study of perceived tax incidence." *Review of Economics and Statistics*, 74(1), 176–179.

Mueller, D. (1989). *Public Choice II*. Cambridge: Cambridge University Press.

Mulligan, C. B., R. Gil, and X. Sala-i-Martin. (2004). "Do democracies have different public policies than nondemocracies?" *Journal of Economic Perspectives*, 18, 51–74.

Musgrave, R. (1959). *The Theory of Public Finance*. New York: McGraw-Hill.

Musgrave, R. (1969). *Fiscal Systems*. New Haven: Yale University Press.

Nechyba, T. J. (1997). "Local property and state income taxes: the role of interjurisdictional competition and collusion." *Journal of Political Economy*, 105(2), 351–384.

Nechyba, T. J. (1999). "School finance induced migration and stratification patterns: the impact of private school vouchers." *Journal of Public Economic Theory*, 1(1), 5–50.

Ordeshook, P. (1992). *A Political Theory Primer*. New York: Routledge.

Persson, T. and G. Tabellini. (2003). *The Economic Effects of Constitutions*. Cambridge Mass: MIT Press.

Persson, T. and G. Tabellini. (2001). "Political institutions and policy outcomes: what are the stylized facts?" Center for Economic Studies and IFO Institute for Economic Research, Munich Germany, Working Paper 459, April.

Pigou, A. C. (1951). *A Study in Public Finance*, Third Edition. London: Macmillan.

Pommerehne, W. W. and Schneider, F. (1983). "Does government in a representative democracy follow a majority of voters' preferences? — an empirical examination," in H. Hanusch (ed.) *Anatomy of Government Deficiencies*. Berlin Heidelberg New York: Springer, pp. 61–84.

Poterba, J. M. (1997). "Do budget rules work?," in A. Auerbach (ed.) *Fiscal Policy: Lessons From Economic Research*. Cambridge, MA: MIT Press, pp. 329–336.

Ramsey, F. P. (1927). "A contribution to the theory of taxation." *Economic Journal*, 37, 47–61.

Roberts, K. W. S. (1977). "Voting over income tax schedules." *Journal of Public Economics*, 8, 329–340.

Roemer, J. (2001). *Political Competition: Theory and Applications*. Cambridge, MA: Harvard University Press.

Romer, T. (1975). "Individual welfare, majority voting and the properties of a linear income tax." *Journal of Public Economics*, 4, 163–168.

Romer, T. and H. Rosenthal. (1978). "Political resource allocation, controlled agendas and the status quo." *Public Choice*, 33, 27–43.

Rutherford, T. and S. L. Winer. (1999). "Endogenous policy in a computational general equilibrium framework, with application to the United States 1973 and 1983," in W. Hettich and S. Winer (eds.) *Democratic Choice and Taxation*, Chap. 7. Cambridge: Cambridge University Press.

Rutherford, T. and S. L. Winer. (2002). "Endogenous policy in a computational general equilibrium framework." Research Report 9007, University of Western Ontario, June 1990, in S. Winer (ed.) *Political Economy in Federal States: Selected Essays of Stanley L. Winer*, Chap. 12. Cheltenham UK: Edward Elgar.

Salanie, B. (2003). *The Economics of Taxation*. Cambridge Mass: MIT Press.

Schofield, N. (1978). "Instability of simple dynamic games." *Review of Economic Studies*, 45, 575–594.

Schumpeter, J. A. (1981). "The crisis of the tax state," Translated (1954) in A. Peacock et al. (eds.) *International Economic Papers 4*. London: MacMillan, pp. 29–68.

Shepsle, K. A. (1979). "Institutional arrangements and equilibrium in multi-dimensional voting models." *American Journal of Political Science*, 23, 27–59.

Simons, H. (1938). *Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy*. Chicago: University of Chicago Press.

Sjoquist, D. L., Mary Beth Walker, and S. Wallace. (2005). "A mixture model analysis of local sales taxes, property taxes and expenditures." *Public Finance Review*, 33(1), 36–61.

Steinmo, S. (1993). *Taxation and Democracy: Swedish, British and American Approaches to Financing the Modern State*. New Haven: Yale University Press.

Stewart, C. H., Jr. (1991). "The politics of tax reform in the 1980s," in A. Alesina and G. Carliner (eds.) *Politics and Economics in the Eighties*. Chicago: University of Chicago Press.

Stigler, G. J. (1970). "Director's law of public income redistribution." *Journal of Law and Economics*, 13, 1–10.

Stiglitz, J. E. (1987). "Pareto efficient and optimal taxation and the new new welfare economics," in A. Auerbach and M. Feldstein (eds.) *Handbook of Public Economics*, Volume II. North-Holland: Elsevier, pp. 992–1042.

Surrey, S. S. (1973). *Pathways to Tax Reform: The Concept of Tax Expenditures*. Cambridge, MA: Harvard University Press.

Tridimas, G. and S. L. Winer (2005). "The political economy of government size." *European Journal of Political Economy*, 21, 643–666.

Usher, D. (1994). "The significance of the probabilistic voting theorem." *Canadian Journal of Economics*, 27(2), 433–445.

Volkerink, B. and J. de Haan (2000). "Fragmented government effects on fiscal policy: new evidence." Unpublished Paper, University of Groningen.

von Hagen, J. (1992). "Budgetary procedures and fiscal performance in the European communities." Commission of the European Communities, Economic Papers No. 96, October.

Wagner, R. E. (1976). "Revenue structure, fiscal illusion, and budgetary choice." Public Choice, 25, 45–61.

Warskett, G., S. L. Winer, and W. Hettich (1998). "The complexity of tax structure in competitive political systems." *International Tax and Public Finance*, 5, 127–155.

Webber, C. and A. B. Wildavsky. (1986). *A History of Taxation and Expenditure in the Western World*. New York: Simon and Schuster.

Weingast, B. R. (1979). "A rational choice perspective on congressional norms." *American Journal of Political Science*, 23, 245–262.

West, E. G. and S. L. Winer. (1980). "Optimal fiscal illusion and the size of government." *Public Choice*, 35(5), 607–622.

Wicksell, K. (1896). "A new principle of just taxation," in R. Musgrave and A. Peacock (eds.) *Classics in the Theory of Public Finance*. New York: Macmillan (1958), pp. 72–118.

Wilson, J. D. (1989). "An optimal tax treatment of leviathan." *Economics and Politics*, 1(2), 97–118.

Wilson, J. D. (1999). "Theories of tax competition." *National Tax Journal*, 52(2), 269–304.

Winer, S. L. and W. Hettich. (1991). "Debt and tariffs, an empirical investigation of the evolution of revenue systems." *Journal of Public Economics*, 45, 215–242.

Winer, S. L. and T. Rutherford. (1992). "Coercive redistribution and the franchise: a preliminary investigation using computable general equilibrium modeling," in A. Breton et al. (eds.) *Preferences and the Demand for Public Goods*. Dordrecht: Kluwer, pp. 351–375.

Winer, S. L. and W. Hettich. (2002). "Explaining the use of related instruments," in S. Winer (ed.) *Political Economy in Federal States: Selected Essays of Stanley L. Winer*, Chap. 3. Cheltenham, UK: Edward Elgar.

Wintrobe, R. (1996). *The Political Economy of Dictatorship*. New York: Cambridge University Press.

Witte, J. F. (1985). *The Politics and Development of the Federal Income Tax*. Madison, WI: University of Wisconsin Press.

Chapter 24
The Political Economy of Monetary Policy Conduct and Central Bank Design[1]

Manfred Gärtner

Introduction

There are few areas in which public choice has had as much success in making inroads into mainstream economics and, in particular, in influencing real-life developments as in the design of monetary institutions and the day-to-day conduct of monetary policy. This survey tracks these developments, from the humble beginnings in the 1970s related to Nordhaus' (1975) account of the opportunistic political business cycle to the widespread academic and political discussion on monetary policy rules and targets of today.[2] The next section contains a compact review of the two classical ideas in political macroeconomics, the political business cycle and the inflation bias. We then move on to more modern stochastic models, in which the desire for undistorted stabilization of supply shocks calls for refined remedies to the time-inconsistency problem, such as performance contracts and inflation targets for central banks. The following section moves on to a discussion of current developments that focus on instrument and targeting rules for monetary policy. Finally, we briefly assess these developments.

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice*, Volume I, 159–172

[2] This survey puts particular emphasis on recent, policy-related developments. Older developments are selectively discussed in order to bring out the public choice roots of many current developments and provide a theoretical background for current discussions. For more detail on these earlier developments, readers may consult two previous surveys of mine that focus on political business cycles and the first-generation discussion of time-inconsistency (Gärtner, 1994) and the second-generation discussion of tune inconsistency including a refined macroeconomic framework with persistence and the interaction with fiscal policy (Gärtner, 2000).

How it Started: Political Business Cycles and All That

Today's academic discussion and recent developments in monetary policy and institutions rest on three main pillars: The traditional theory of economic policy in the spirit of Theil (1961) and Tinbergen (1952), the endogenization of economic policy, the groundwork for which was laid in many classical writings in public choice, although the main influence stems from the compact and compelling formalizations by Nordhaus (1975) and MacRae (1977); and, the rules-versus-discretion debate that came in the wake of the rational expectations revolution, with implications for endogenous policy making that were initially formalized by Kydland and Prescott (1977) but worked out and popularized by Barro and Gordon (1983). We will focus here on the public-choice-related roots of modern monetary policy conduct and design.

The birth of New Political Macroeconomics, as it would be called decades later, and, hence, also of positive analyses of monetary policy, was Nordhaus' concise formal demonstration of what opportunistic governments might do to an economy. In strong contrast to Theil–Tinbergen-type benevolent policymakers, opportunism takes the form of vote maximization at periodically held elections. Voters derive instantaneous or period utility from the state of the economy, as represented by inflation π and the logarithm of income y (or, alternatively, unemployment):

$$u = -0.5\pi^2 + \xi y \qquad (24.1)$$

Votes cast on election day then reflect total utility and, hence, the course of the economy during the incumbent government's recent term in office, with more distant periods receiving less weight due to voter forgetfulness.

Operating within a natural-rate aggregate-supply framework in which income (or, again, unemployment) depends on inflation surprises,[3]

$$y = \pi - E_{-1}\pi \qquad (24.2)$$

and inflation expectations are adaptive, governments maximize reelection prospects by resorting to expansionary policies, fiscal or monetary, in the run-up to an election, while deliberately driving the economy into a recession once the election is over, thereby creating election-related swings in economic activity known as the *political business cycle*.

From the perspective of mainstream macroeconomics, the Nordhaus model (and its cousin, the *partisan theory* proposed by Hibbs (1977), which suggested that election-related swings were due to ideologically motivated differences between the preferences of party constituencies) was almost dead on arrival. Despite the extraordinary interest it drew from public choice scholars, its key building blocks were at that time being discarded by macroeconomists: a non-vertical, long-run

---

[3] We do not make a distinction between a Phillips curve and an aggregate-supply curve. So, to simplify notation, we usually normalize the log of potential income, y^*, to zero and give the aggregate-supply curve unity slope.

Phillips curve (which was not essential to the political business cycle, however), adaptive inflation expectations, and backward-looking voters. A number of authors[4] quickly pointed out that little in terms of added rationality in inflation expectations formation was required in order to eliminate the political business cycle.

While efforts by Alesina (1987), Persson and Tabellini (1990), and others gave the study of election-related macroeconomic cycles a vigorous second life under the labels of *Rational partisan theory* and *Rational political business cycles*, political business cycles do not feature prominently on today's research agenda any longer.[5]

Instead, research interest has shifted towards the rational-expectations equilibrium implications of endogenous policy making, with a particular emphasis on monetary policy. The starting point for this work, overlooked by most early critics, is the insight that while rational inflation expectations do indeed eliminate the political business cycle, they do leave the economy and policy trapped in a suboptimal, inefficient equilibrium. If monetary policy is driven by preferences such as (24.1), either because it caters to the electorate, or because this describes the government's or the central bank's very own preferences, the model's discretionary rational-expectations solution in the context of a one-shot game between the government and the economy is

$$\pi = \xi \tag{24.3}$$

Thus, despite the desire for full price stability inherent in (24.1), discretionary monetary policy cannot deliver.[6] The reason is the time inconsistency of price stability. Once it is achieved with income being at its potential level, the central bank can always raise its own utility, or public support, by generating some inflation and a lot of income gains. While this mechanism and insight had already been described by Kydland and Prescott (1977), it attracted little attention until it was restated and popularized by Barro and Gordon (1983). The latter work triggered a still ongoing discussion of what institutional arrangements would lead to the best macroeconomic outcomes, in particular, a reduction of the inflation bias. Initially, Barro and Gordon (1983) had suggested that reputational forces may take care of the inflation bias. However, because such forces are strongly weakened when the government's horizon does not extend to infinity, Rogoff's (1985) suggestion to put monetary policy into the hands of a conservative central bank, characterized by total oblevity towards income developments, received the most attention. In the above context, an arch conservative monetary policy guided by preferences $\xi = 0$ delivers full price stability without any detrimental effects on income. To achieve such policy, the governing body of the central bank must have preferences

---

[4] See, for example, Frey and Ramser (1976) and McCallum (1977).

[5] This, by any means, should not be read to mean that political business cycles are dead. See, for example, the contribution by Drazen (2000b).

[6] The inferiority of this result obtained under discretion is usually demonstrated by comparing it with the optimal inflation rate $\pi = 0$ that obtains when the central bank has to *commit* to an inflation rate *before* expectations are being formed.

$\xi = 0$, and the central bank needs to be made completely independent of the government (which political competition forces to attend to the preferences of voters represented by ξ).[7] Condoned by the apparent empirical support for this proposition in the form of significant negative correlations between long-run inflation and measures of central bank independence,[8] the long ruling orthodoxy was that central banks must be completely independent and as conservative (meaning inflation averse) as possible.[9]

Enter the Stabilization Bias

Two innovations rekindled interest in the basic Nordhaus scenario and kept the discussion alive and vigorous up to the present.

The first was a modification of the utility function that gave inflation and income *symmetric* treatment. Nordhaus (1975), Kydland and Prescott (1977), Barro and Gordon (1983), and virtually hundreds of papers since had employed an asymmetric functional form, assuming that utility depended nonlinearly on inflation, but linearly on income (or unemployment). This did help simplify the math, yet still sufficed to derive the political business cycle under adaptive inflation expectations and the inflation bias when expectations were rational.

The second innovation was to conduct the analysis in a more realistic stochastic context in which the economy was subject to supply shocks, and the potential need for stabilization entered the picture.

The Trade between Price Stability and Shock Stabilization

In order to demonstrate the implications of these two innovations, let us proceed from a hybrid utility function that comprises both the original asymmetric treatment (for $\alpha = 0$) and the later symmetric treatment (for $\alpha = 1$):

$$u = -0.5\pi^2 - \alpha\xi(y - k)^2 + (1 - \alpha)\xi y \qquad (24.4)$$

$k > 0$ is society's income target, which is assumed to exceed potential income (which has been normalized to zero) because the latter is inefficiently low (also carrying involuntary unemployment) due to distortive taxes, monopolistic trade unions, legal constraints, and other imperfections in goods and labor markets.

---

[7] See Eijffinger and Hoeberichts (1998).

[8] See Alesina and Summers (1993).

[9] See, however, Forder's (1998a, b) illuminating and sobering account of the validity of empirical evidence on central bank independence and inflation.

Aggregate supply is subject to surprise inflation plus supply shocks, ε, which are white noise with zero mean and variance σ_ε^2:

$$y = \pi - E_{-1}\pi + \varepsilon \tag{24.5}$$

Maximizing equation (24.4) subject to (24.5) yields the following rational-expectations solutions for inflation and income:

$$\pi = (1 - \alpha)\xi + \alpha\xi k - \frac{\alpha\xi}{1 + \alpha\xi}\varepsilon\alpha \tag{24.6}$$

$$y = \frac{1}{1 + \alpha\xi}\varepsilon \tag{24.7}$$

Equations (24.6) and (24.7) convey three important insights:

1. The first two terms on the right-hand side of (24.6) constitute the inflation bias that monetary policy cannot get rid of, even in the absence of shocks. If utility is linear in y ($\alpha = 0$), this bias equals ξ. If utility is nonlinear in y, with decreasing marginal utility ($\alpha = 1$), this bias amounts to ξk. It is positive if k exceeds potential income. Then the marginal utility of income is positive at the no-surprise equilibrium level, and inflation must be positive in order to generate a marginal disutility of inflation large enough to counterbalance the net temptation to raise income. At full price this does not apply because the marginal disutility of inflation is zero.
2. The coefficients in the stochastic terms of both (24.6) and (24.7) indicate how supply shocks are split into inflation and income responses. Note that the absolute values of the two coefficients sum up to unity.[10] So, only $1/(1 + \alpha\xi)$ percent of any given adverse supply shock is actually permitted to drive income down, while the remaining $\alpha\xi/(1 + \alpha\xi)$ percent materializes in increased inflation.
3. When utility is linear in y ($\alpha = 0$), the solutions simplify to $\pi = \xi$ and $y = \varepsilon$. Inflation is always constant at a level reflecting the conservativeness of monetary policy. Supply shocks are never permitted to affect inflation, independently of the conservativeness of monetary policy.

The third insight states the specific conditions under which the famous monetary-policy conservativeness result holds: In order to achieve second-best outcomes, that is, full price stability and the exact extent of shock stabilization society requests, monetary policy needs to be as conservative as possible in the sense that it should only look at the goal of price stability while ignoring movements of income altogether.[11]

---

[10] This is because the aggregate-supply curve has been given a slope 1. In the general case, the slope coefficient would also feature in the stabilization terms.

[11] For a result to be classified as first best, income also would have to be as required. Since this is considered to be beyond the reach of monetary policy, optimal monetary policy is only judged by whether it achieves second-best results.

If, however, more realistically, the utility function is symmetric ($\alpha = 1$), a dilemma pops up. To see this, note that the solutions for inflation and income now become

$$\pi = \xi k - \frac{\xi}{1 + \xi}\varepsilon \tag{24.8}$$

$$y = \frac{1}{1 + \xi}\varepsilon \tag{24.9}$$

The key insight here is that in a stochastic context with decreasing marginal utility from income gains, delegating monetary discretion to an arch conservative central bank (characterized by $\hat{\hat{\xi}} = 0$) constitutes a fourth-best solution only. All it ensures is that we achieve price stability. The price to be paid is distorted responses to supply shocks. The variance of inflation is minimized at var $\pi = 0$ $\sigma_\varepsilon^2 = 0$, but this goes at the cost of maximum variance of income at var $(y) = \sigma_\varepsilon^2$. Society would prefer an intermediate solution, namely var $(\pi) = [\xi/(1 + \xi)]^2\sigma_\varepsilon^2$ and var $(y) = [1/(1 + \xi)]^2\sigma_\varepsilon^2$. In the face of this trade-off between inflation bias and stabilization bias, a superior outcome, a third-best result, is achieved if society picks a more moderately conservative central bank, one that is more conservative than society, but not arch conservative ($\xi > \hat{\xi} > 0$) (Rogoff 1985).

Figure 24.1 may help clarify the trade-offs involved and serve as a background for issues addressed later on. The convex line constitutes the trade-off between income variability and inflation variability implied by the model. Society's preferences, represented by concave indifference curves, determine the desired split between income and inflation variability. We can move down along this line from the point on the ordinate (which obtains for $\hat{\hat{\xi}} = 0$) towards the point on the abscissa (which obtains for $\hat{\hat{\xi}} \to \infty$). The dilemma is that in order to remove the inflation
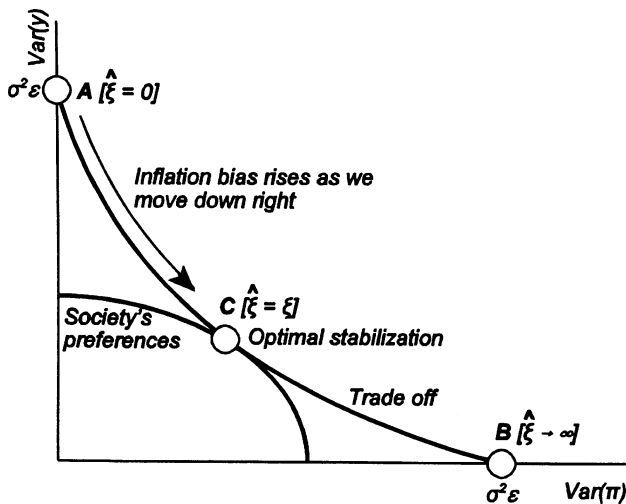
Fig. 24.1 The trade-off between income variability and inflation variability

bias we need $\hat{\bar{\xi}} = 0$. This would put us into point A on the variance trade-off line with the variance of income being at a maximum at σ_ε^2 and the variance of inflation being at a minimum at zero. Society would prefer B. As we move from A towards B by raising $\hat{\bar{\xi}}$, we reduce the stabilization bias, but pay by increasing the inflation bias. All we can achieve is a third-best optimum in a point such as C, where society's net marginal benefit from increasing $\hat{\bar{\xi}}$ is zero.

The 1990s brought an avalanche of research on how to move beyond the third-best outcome generated by a moderately conservative central bank. This quest for second-best outcomes in a stochastic macroeconomic framework focused on two main suggestions: to equip central bank chiefs with a *performance contract*, or to commit them to an *inflation target*.[12]

Performance Contracts

Equipped with a linear performance contract of the form $s = -\lambda\pi$, where s is a variable component of the central bank's governing body's salary, which depends on inflation, the central bank's derived utility function changes into

$$u = -0.5\pi^2 - 0.5\hat{\bar{\xi}}(y - k)^2 - \lambda\pi \qquad (24.10)$$

Now optimal policy under discretion leads to the following behavior of inflation and income:

$$\pi = \hat{\bar{\xi}}k - \lambda - \frac{\hat{\bar{\xi}}}{1 + \hat{\bar{\xi}}}\varepsilon \qquad (24.11)$$

$$y = \frac{1}{1 + \hat{\bar{\xi}}}\varepsilon \qquad (24.12)$$

These results show that a properly designed linear performance contract can indeed lead to second-best results. The inflation bias, comprising the first two terms on the right-hand side of (24.11), is removed if $\lambda = \hat{\bar{\xi}}\pi$. And, shock stabilization is prevented when the central bank's preferences are representative of society's ($\hat{\bar{\xi}} = \xi$). This actually is ensured best if the central bank is not independent of the government. Whatever tendencies towards a higher inflation bias this may carry can easily be taken care of by setting the punishment coefficient in the performance contract appropriately.[13]

---

[12] The discussion on performance contracts was initiated by Persson and Tabellini (1993), Waller (1995), and Walsh (1995). Major contributors to the early academic discussion of inflation targets in the current context were Herrendorf and Lockwood (1997), Muscatelli (1995), and Svensson (1997). See also Bernanke and Mishkin (1997), and Walsh (1998), Ch. 8.

[13] A linear contract focusing on the performance of aggregate income could be tailored to achieve the same second-best result, of course. The literature emphasizes inflation performance contracts, however.

Inflation Targets

Inflation targets have been very popular in academic research as a probably more realistic and viable alternative to performance contracts. Inflation targets also do provide a natural link from the literature discussed here to the recent intensive discussion of general monetary policy rules and targets at which we will look below. The general idea is that society (via the government) can communicate an inflation target π^T to the central bank. The questions to be answered are: what this target should be, how target misses are to be punished, and what preferences the central bank should have.

After adding the inflation-target term to the central bank's utility function, the derived utility function reads

$$u = -0.5\pi^2 - 0.5\hat{\xi}(y-k)^2 - 0.5\lambda(\pi - \pi^T)^2 \tag{24.13}$$

Under discretion, the inflation rate follows

$$\pi = \frac{1}{1+\lambda}\hat{\xi}k - \frac{\lambda}{1+\lambda}\pi^T - \frac{\hat{\xi}}{1+\lambda+\hat{\xi}}\varepsilon \tag{24.14}$$

while income is determined by

$$\frac{1+\lambda}{1+\lambda+\hat{\xi}}\varepsilon \tag{24.15}$$

Again, a second-best optimum can be achieved. The condition for the inflation bias to disappear is $\pi^T = \hat{\xi}k/\lambda$. This is an awkward result, however: not only because the central bank must be told to deflate, but even more so because the central bank systematically misses the assigned target. In the aspired zero-inflation equilibrium, the deviation from the inflation target must be large enough to offset any temptation to inflate that results from the central bank's own preferences.[14]

The condition for avoiding a stabilization bias is that the shock's coefficient in, say, (24.15), which describes the central bank's response, must be the same as the shock's coefficient in (24.9), which states society's desired response. This is accomplished if $\hat{\xi} = (1+\lambda)\xi$, meaning that now the government must pick a central banker who is *less conservative* and less inflation-averse than society.

Table 24.1 summarizes the consolidated knowledge about central bank independence and conservativeness in this section's macroeconomic environment.

---

[14] Svensson (1997) proposes that the central bank can simply be assigned a utility function that completely overrides any generic preferences which the central bank actually has. This formally solves the problem of a negative inflation target that is never met, but is arbitrary and unconvincing, not only from a public choice perspective.

Table 24.1 How conservative should the central bank be?

Macroeconomic and monetary policy framework	Optimal degree of central bank conservatism
• Deterministic macroeconomic framework	
– Baseline model (perfect discretion)	– arch conservative ($\xi = 0$)
• Stochastic macroeconomic framework	
– Baseline model (perfect discretion), third best	– moderately conservative ($\xi > \underline{\xi} > 0$)
– Added performance contract; second best	– as conservative as society ($\xi = \underline{\xi}$)
– Added inflation target; best	– less conservative than society ($\xi > \underline{\xi}$)

The important point it does highlight is that little scientific support remains for the quest for the most independent, most conservative central bank that did and still does seem to shape the design and development of institutions in many of the world's countries and regions.[15]

A Macroeconomic Framework with Income Persistence

The above findings do not change dramatically if, more in line with our empirical knowledge about the time series properties of income and other macroeconomic variables, we let shocks have lasting effects on income due to some degree of persistence, as in equation (24.16).

$$y = \beta y_{-1} + \pi - E_{-1}\pi + \varepsilon \tag{24.16}$$

Because inflation surprises and shocks now affect all future incomes, policy choices are being made so as to maximize the expected present value $E_{t-1}U_t$ of current and future period utilities:

$$E_{t-1}U_t = \sum_{i=0}^{\infty} \delta^i E_{t-1}u_{t+i} \tag{24.17}$$

Under discretion, there is still an inflation bias, which now takes the form

$$\pi = \frac{\xi k}{1 - (\alpha + b)\delta} - cy_{-1} - d\varepsilon \tag{24.18}$$

where b, c, and d are coefficients composed of the structural equations parameters that we do not need to spell out here.[16] This bias features a constant part, which is

---

[15] The most prominent example is probably the European Central Bank, the blueprint for which does not seem to take account of the trade-offs and refined results emerging in a stochastic macroeconomic context.

[16] For details, see Sevensson (1997).

similar to the bias in the natural rate framework discussed above. In addition to the familiar dependence on preferences ξ, this bias also depends on the degree of persistence β. The straightforward explanation is that the more persistent the income is, the longer the income gains generated by current inflation surprises last. But then the temptation to inflate is larger, and because this is anticipated by the labor market, we end up with a higher inflation bias.

The second term defining the inflation bias is endogenous and time dependent. It states that this inflation bias is the higher, the lower the income was last period. The mechanism at play here is that the marginal utility of income is higher when an adverse supply shock hit income last period and persistence will therefore tend to keep income below potential income this period also. The central bank will therefore be prepared to inject a larger inflation hike into the economy in the hope of income gains. But since again the labor market anticipates this, these income gains do not really accrue, and all we are left with is an inflation bias above average.

State-dependent Performance Contracts

While the math to demonstrate this is laborious, it is intuitively clear that a linear inflation performance contract cannot do away with this type of *variable* inflation bias. The required extension of the optimal contract is straightforward, though. Since the inflation bias is variable, dependent on last period's income, the performance contract must also be state dependent of the form

$$s_t = -(\lambda_1 - \lambda_2 y_{t-1})\pi_t \tag{24.19}$$

This contract may specify λ_1 so as to eliminate the constant inflation bias, as in the natural-rate framework discussed above. And it may specify λ_2 such as to counterbalance the added incentive to inflate after income fell, thereby removing the state-dependent inflation bias. Once the performance contract is designed optimally, central bank preferences should be identical to society's in order not to bias stabilization. This mimics the result obtained in the natural-rate context.

State-dependent Inflation Target

In the presence of income persistence, inflation targets must be path dependent, comprising a constant term to take care of the fixed inflation bias and a term that follows lagged income to take care of the variable inflation bias: $\pi_t^T = \beta_0 + \beta_1 y_{t-1}$. As Svensson (1997) shows, however, even a state-dependent inflation target cannot get rid of both types of inflation bias, and keep stabilization undistorted. It must be combined with the appropriate central bank preferences that compensate for the stabilization bias introduced by the inflation target.

Current Developments

Current research on monetary policy and central banks is looking for answers to three important questions:

1. How can the stabilization options be improved? Rather than discussing how different targets within a given family can be optimized so as to achieve second-best solution within a given trade-off, researchers turn to completely different target variables and how they may affect the trade-off options. We will exemplify this by comparing inflation targets as discussed above to price level targets.
2. Is the consolidated knowledge as surveyed in the section **Enter the Stabilization Bias** reasonably robust to changes in the macroeconomic environment within which monetary policy operates? A key role in this discussion is being played by the so-called New Keynesian aggregate-supply curve which, in line with recent methodological changes in macroeconomics, is being derived from solid micro-foundations and features forward-looking inflation expectations.
3. How can some of the more abstract theoretical insights of political macroeconomics be brought to bear on the actual conduct of monetary policy? This question is being discussed in a separate strand of research focusing on policy rules, which has close ties to the topics discussed so far.

The Choice of Targets and their Effects on Trade-offs

The question which macroeconomic variable monetary policy should target is not a trivial one. To demonstrate how that choice of target variables affects the variability of macroeconomic variables, as well as the implied trade-offs between these variabilities, let us compare inflation targets with price level targets. In order to focus on the issue at hand, assume, as much of the literature does, that society can assign a target to the central bank in the strict sense that the target *overrides* any pertinent preferences the central bank itself may have (rather than *adding* it to the central bank's preferences, as assumed previously). Equipped with such an assigned inflation target, the central bank utility function reads

$$u = -0.5\left(\pi - \pi^T\right)^2 - 0.5\hat{\xi}(y - k)^2 \tag{24.20}$$

The discretionary optima for inflation and income that follow are

$$\pi = \hat{\xi}k - \pi^T - \frac{\hat{\xi}}{1 + \hat{\xi}}\varepsilon \tag{24.21}$$

and

$$y = \frac{1}{1 + \hat{\xi}}\varepsilon \tag{24.22}$$

The volatility trade-off from which society may choose by selecting ξ is characterized by $\mathrm{var}(\pi) = \left[\hat{\xi}/\left(1+\hat{\xi}\right)\right]^2 \sigma_\varepsilon^2$ and $\mathrm{var}(y) = \left[1/\left(1+\hat{\xi}\right)\right]^2 \sigma_\varepsilon^2$, and depicted as the lower convex line in Fig. 24.2. The specific target value π^T neither affects the trade-off, nor where we end up on it. Actually, the depicted curve is an efficiency frontier. As long as monetary policy is governed by preferences coming from this very family of utility functions, comprising inflation and income as arguments that enter in quadratic form, we end up somewhere on this line. All society can do is move up or down this curve into its preferred point by picking $\hat{\xi}$.

Now, other families of utility functions exist, comprising different variables or functional forms. An argument that is often advanced against inflation targets for monetary policy is that it makes the variance of the price level go towards infinity as we increase the time horizon, making it difficult for individuals and firms to form expectations. In an attempt to remedy this, the government may assign a price level target to the central bank instead, even though society's preferences are still as given in (24.4) with $\alpha = 0$. The central bank's utility function then reads

$$u = -0.5\left(p - p^T\right)^2 - 0.5\hat{\xi}(y - k)^2 \tag{24.23}$$

where p is the log of the price level. Note that the aggregate-supply function (24.5) may be rewritten as

$$y = p - E_{-1}p + \varepsilon \tag{24.24}$$

since inflation is the first difference in the log of the price level. Maximizing (24.23) subject to (24.24) mimics the maximization of (24.20) subject to (24.5), except that the price level p has taken the place of inflation π. Hence, the solution for the price level is equal to the solution we previously derived for inflation,

$$p = \hat{\xi}k - p^T - \frac{\hat{\xi}}{1 + \hat{\xi}}\varepsilon \tag{24.25}$$

Whether we assign an inflation target or a price level target has no effect on income, which again follows

$$y = \frac{1}{1 + \hat{\xi}}\varepsilon \tag{24.26}$$

Since $\pi \equiv p - p_{-1}$, the behavior of inflation is directly derived from (24.25):

$$\pi = -\frac{\hat{\xi}}{1 + \hat{\xi}}(\varepsilon - \varepsilon_{-1}) \tag{24.27}$$

This implies an inflation variance of $\mathrm{var}(\pi) = 2\left[\hat{\xi}/\left(1+\hat{\xi}\right)\right]^2 \sigma_\varepsilon^2$, which is twice as large as when the central bank pursued an inflation target. As Fig. 24.2 illustrates, this dramatically worsens the options for stabilization policy and is likely to affect
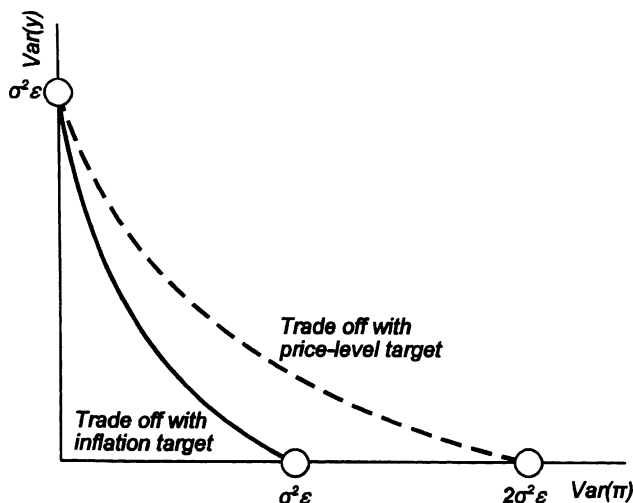
Fig. 24.2 The choice of targets and the trade-off options

society's pick of central bank conservativeness. In fact, a second-best optimum cannot even be achieved because $[\xi/(1+\xi)]^2 = 2\left[\hat{\xi}/\left(1+\hat{\xi}\right)\right]^2$—which would provide the right inflation variability—and $1/(1+\xi) = \left[1/\left(1+\hat{\xi}\right)\right]$—which would provide the desired variability of income—cannot be met at the same time. Independently of society's preferences, which we may not know, we can state that assigning a price-level target is inefficient. Switching to an inflation target permits lowering the variance of inflation (income) without raising the variance of income (inflation).[17]

The example used here goes to show that the choice of a target variable, or variables, is a delicate one with obvious welfare implications. The inefficiency of price level targeting relative to inflation targeting is not robust, however, to changes in the macroeconomic framework. This is not really surprising since the trade-off is generated by the complete model, comprising both the macroeconomic structure and the incentives governing monetary policy. Svensson (1999b) demonstrates that, when faced with an economy with a sufficient degree of income persistence, society may be well advised to assign price-level targeting even though it possesses preferences cast in terms of an optimal inflation rate, because it results in lower inflation variability. Dittmer and Gavin (2000) show that in a model with a New Keynesian Phillips curve, as discussed in the following section, price-level

---

[17] Note that a price-level target here is only inefficient from the partial perspective of shock stabilization. Things are more complex when we bring the inflation bias into the picture, because, as we move down from the no-bias point on the ordinate, the bias increases as we move along the inflation-target trade-off, but not as we move along the price-level-target trade-off.

targeting always generates a more favorable trade-off between income and inflation variability, even if income is not persistent.

The New Keynesian Aggregate-supply or Phillips Curve

Roberts (1995) uses the Calvo (1983) model (in which prices are sticky because during any given period a firm has a fixed probability, strictly smaller than 1, that it may adjust prices) to show that a loglinear approximation about the steady state of the aggregated pricing decisions of individual firms reads

$$y = \beta y_{-1} + \pi - E\pi_{+1} + \varepsilon \qquad (24.28)$$

While this aggregate-supply curve looks very similar to the neoclassical supply curve with persistence that we used above, the inclusion of tomorrow's expected rate of inflation rather than today's has important implications.[18] One is that any movement in inflation, and particularly when it is rationally anticipated, affects income.[19] Clarida et al. (1999) look at how this bears on the issues discussed in the preceding sections of this paper. Major findings are:

1. There is an inflation bias if the central bank has an income target that exceeds potential income. This is most easily rationalized if we think of monetary policy as a series of one-shot games in which policymakers take next period's expected inflation as given. It also holds in a more general setting, however, when the central bank has a longer horizon.
2. The inflation bias is negatively correlated with central bank conservativeness, that is, with the weight that the income target has in its utility function. An inflation nutter ($\hat{\xi} = 0$), as an arch conservative central bank is sometimes referred to, would entirely eliminate the inflation bias.
3. As a final analog to results obtained within the Neoclassical framework, only a moderately conservative central bank would strike the right balance between the desires to reduce the inflation bias and to keep shock stabilization as undistorted as possible.

The framework used by Clarida et al. (1999), being somewhat richer than the one reported here, with shocks on the supply side and on the demand side, permits a host of other insights not directly comparable to the consolidated knowledge acquired within the neoclassical framework. A key issue that has been raised within this

---

[18] For a detailed discussion of this and related New Keynesian aggregate-supply curves and their implications for monetary policy see Clarida et al. (1999).

[19] Incidentally, a vote-maximizing government facing a backward-looking electorate and a New Keynesian aggregate-supply curve with $\beta = 0$ would create a political business cycle with some of the same features as the Nordhaus cycle. In fact, in a two-period setting it would be the very same cycle that a government creates when aggregate supply is neoclassical and inflation expectations are of the simplest adaptive mould ($E_{-1}\,\pi = \pi_{-1}$).

context, however, is whether preferences do indeed feature an income target which exceeds potential income, thus generating a problem of time inconsistency. This is an important question, because if there was no inflation bias, or if it had different causes than presumed since spelled out by Kydland and Prescott (1977) and Barro and Gordon (1983), there might not exist the dilemma of choosing between inflation-bias reduction and less distortion of stabilization policy, making things much simpler for monetary policy.

As Cukierman (2002) has demonstrated, though, an income target exceeding normal or potential income is not necessary for an inflation bias to occur. All that is needed is an asymmetry in the central bank's utility function. Suppose that preferences are such that the central bank wants income to rise, but only until it reaches potential income. It does not want to push it beyond that level, but, if it exceeds potential income owing to a favorable supply shock, it refrains from trying to drive it down. As a consequence, whenever a positive shock hits and income is above potential income, inflation remains at zero. Whenever a negative shock drives income below normal levels, monetary policy cushions that fall by creating inflation. As a result, average and expected inflation are strictly greater than zero. We have an inflation bias in equilibrium. In this context, much of the same remedies and policy recommendations would apply, with the math being a bit more cumbersome because of the employed piecewise utility functions.

The Quest for Monetary Policy Rules

This is probably the most active topic on today's research agenda on monetary policy. The field is still in a flux, and there are several perspectives from which to look at it. In order to understand the current discussion, we need to introduce some definitions.[20]

From a simplifying perspective there are two kinds of monetary policy rules. The first category comprises *instrument rules*. These specify how some instrument of monetary policy, typically an interest rate or monetary aggregate, responds to a set of macroeconomic variables. If these variables are predetermined at the time the instrument is being set, we speak of an *explicit instrument rule*. An *implicit instrument rule* specifies the instrument as a function of forward-looking variables that are not predetermined, of course. Owing to this simultaneity between instrument and determining variables, this must be considered an *equilibrium condition* rather than a rule.

The second group of monetary policy rules comprises *targeting rules*. Characteristic for a targeting rule is the "assignment" of a loss function to the central bank. If this loss function features only one target variable, say inflation, we have a *strict* targeting rule. If additional variables are included, say income, we speak of a *flexible* targeting rule. To the extent that the right or best target variables are

---

[20] We follow Svensson (1999a, b), who is one of the most active contributors to this discussion.

difficult to control or to observe, the use of loss functions with *intermediate targets* is sometimes proposed. These targets should be highly correlated with the true goal, but easier to control and to observe.

Current research on monetary policy rules is related to the work reported in the section **Enter the Stabilization Bias**. But it also differs in a few major aspects.

1. There is a deliberate shift from a predominantly analytical towards a sophisticated yet practical monetary policy analysis, with strong doses of pragmatism and a quest for quantitative results. As consequence, research interests of academics and central banks have begun to meet in this area.[21]

2. Employed models have been stripped of time inconsistency. So there is no more inflation bias and no more potential for conflict between price stability and stabilization policy. Stabilizing inflation and income around their desired values remains the only challenge. The discarding of the inflation bias appears to come as a response to criticism by a group of central bank notables and academics that the underlying story (Barro and Gordon 1983) was unconvincing and empirically inaccurate. As a result, models are being employed in which a loss function or rule is imposed on the central bank that features an income target coinciding with potential income.[22]

3. While the New Classical or Lucas aggregate-supply curve, more recently with added persistence, had completely dominated the literature discussed in the section **Enter the Stabilization Bias**, there is no such consensus in the rules discussion. By contrast, this discussion accepts that no consensus regarding the right model of the macroeconomy has emerged yet, and emphasizes that this calls for thorough checks as to whether any derived rules are robust in the sense that they still function reasonably well within alternative macroeconomic models. These models cover a wide range of possibilities. Some reduce to a single equation. Some comprise up to 100 equations. Some are derived from intertemporal optimizing behavior of representative agents. Some are made up of equations purported to mimic the dynamic relationships we see in empirical VARs.

4. A final innovation characteristic of the rules discussion is the use of analytical and empirical methods that have become standard in real business cycle and dynamic general equilibrium analyses. This includes the calibration of models, stochastic simulations, and judgment of the empirical validity by means of comparing distributions of and correlations between simulated time series to those encountered in reality.

---

[21] There has been a host of conferences with "monetary policy rules" in the title, sponsored or hosted by central banks. A first example is the conference jointly sponsored by the Sveriges Riksbank and the Institute for International Economic Studies at Stockholm University, held June 12–13, 1998, in Stockholm.

[22] Among those who have criticized the premise of central banks pursuing income targets that exceed potential income from the background of their hands-on experience with monetary policy making is Blinder (1995). Academic criticism of this idea has come, among others, from McCallum (1997) and Taylor (1983).

The Taylor Rule and Other Instrument Rules

Instrument rules for monetary and fiscal policy have a long tradition in economics. In the past, the most famous such rule was the Friedman rule, proposing that the money supply should grow at a fairly constant rate equal to the trend growth rate of income. Such a rule is an explicit, if not an exogenous rule, since it hardly allows for any feedback from current economic variables into monetary policy, certainly not in the short run.[23]

Among the recent crop of more sophisticated monetary policy rules, which includes McCallum's (1988) rule for the monetary base, the Henderson and Mc Kibbin (1993) rule for the federal funds rate, and dozens of other rules, the rule that has swept the field is the one proposed by Taylor (1993). The Taylor rule states that the central bank has a real interest rate target, from which it deviates if inflation and/or income are off target. Solving this for the nominal interest rate yields

$$i = r^T + \pi + 0.5(\pi - \pi^T) + 0.5(y - y^*) \qquad (24.29)$$

When following the Taylor rule the central bank sets its instrument, the federal funds rate, at r^T when inflation and income are at their optimal levels. An increase in inflation makes the central bank raise the nominal interest rate by a factor of 1.5. This raises the real interest rate, dampening the effect of aggregate demand. While it does not include any forward-looking variables, the Taylor rule can nevertheless call for preemptive strikes against future inflation. This is the case if rising income, which also drives up the real interest rate, drives up inflation with a lag.

Initially proposed as a descriptive and expository devise that can be used to account for the general flavor of monetary policy in the United States and explain the Fed's policy shift during the Volcker era, the Taylor rule has become much more. And meanwhile, the quite voluminous amount of empirical research suggests that Taylor's rule is indeed a quite reasonable description of policy behavior of many central banks, including the Bundesbank, which is usually considered the most extreme inflation nutter in recent history.[24] The rule also has come to fame in financial circles, where it is now a common tool for forecasting changes in the interest rate.

Inflation Targeting and Other Targeting Rules

As mentioned, a targeting rule is characterized by the assignment of a loss or utility function to the central bank. In the section **Enter the Stabilization Bias** we showed this in a parsimonious framework for inflation and income. Many possible targets

---

[23] For an account of how the Friedman rule fared in practice, see Hafer and Wheelock (2001).

[24] See, for example, Clarida et al. (1998), who estimate policy reaction functions for the G3 (Germany, Japan, and the United States) and the E3 (United Kingdom, France, and Italy) countries, and Peersman and Smets (1998), who explore the Taylor rule as a benchmark for analyzing monetary policy in the euro area.

are being discussed in the literature, such as the price-level, inflation, nominal GDP, or nominal GDP growth, with inflation targeting drawing the most academic interest and being the most successful among central banks.[25] The term inflation targeting is a misnomer, however, because only *strict* inflation targeting refers to a utility function of the form $u = -0.5(\pi - \pi^T)^2$. If additional target variables enter the utility function, this is being referred to as *flexible* inflation targeting. An example is the familiar utility function

$$u = -0.5(\pi - \pi^T)^2 - 0.5\xi(y - y^*)^2 \tag{24.30}$$

which, for obvious reasons, might just as well be referred to as flexible *income targeting*.

In an effort to facilitate practical implementation or monitoring, a target rule is often expressed as a set of equations the target variables must fulfill. In the case of (24.30), if there is perfect control over the target variables and there is no trade-off, we obtain these equations from the first-order conditions for the unrestricted maximum of the utility function as $\pi = \pi^T$ and $y = y^*$. If control is imperfect, the expected values must equal the targets. Things do become much more complicated, however, when, as is always the case, we have trade-offs between macroeconomic variables, be it within periods or intertemporarily. While first-order conditions usually still exist, they may be too complicated for practical purposes. It may then be advisable to switch to intermediate target variables which, ideally, should be "highly correlated with the goal, easier to control than the goal, easier to observe than the goal, and transparent" (Svensson, 1999a, p. 619). In terms of how to pursue the target, Svensson (1999a) further reports that the target variable included in the loss function is usually not the best indicator for the instrument to respond to.

Comparing Instrument and Targeting Rules

From a purely technical viewpoint, instrument and targeting rules are simply two sides of the same coin. Maximization of any utility function or target subject to a macroeconomic model leads to an optimal instrument rule. For example, maximization of (24.30) with respect to the instrument π, subject to (24.5), gives the instrument rule $\pi = \pi^T - \varepsilon\xi/(1 + \xi)$. This may be rewritten. After solving (24.5) and substituting the result, we obtain

$$\pi = \pi^T - \xi(E_{-1}\pi = \pi^T) - \xi(y - y^*) \tag{24.31}$$

---

[25] It is generally believed that quite a number of central banks, including those of Australia, Canada, New Zealand, Sweden and the United Kingdom, have adopted some form of inflation targeting during the last 10 to 15 years.

The result is a Taylor-like instrument rule in which the instrument π depends on the income gap and on the expected deviation of inflation from the inflation target. We may note here that (24.29) is not an explicit instrument rule. The endogeneity of income on the right-hand side makes this rule implicit, an equilibrium condition.[26]

Just as the optimization of a given utility or loss function generates an explicit or implicit instrument rule, any given instrument rule can be traced back to a utility function that is being optimized. This mapping from preferences to instrument rule or back does, of course, crucially depend on the macroeconomic model to which it is attached, and it may not be unique.

The competition between instrument and target rules therefore boils down to the question of which one is more practical. A starting point for this discussion must be that realistic models of the macroeconomy are much more complex than the models we looked at. While this need not affect the utility function to be assigned to the central bank, it leads to immensely complicated optimal instrument rules, which will also be very difficult to monitor. On the other hand, it will also bear heavily on how a central bank pursues its assigned targets. It is flexible in doing so, however, and free to incorporate any progress the science of economics may make. As our view of how the economy functions change, the target(s) need not be adjusted. An instrument rule, by contrast, would have to be adjusted continuously, which may lead to credibility problems. But this is where the conceded uncertainty about the proper macroeconomic model comes into play. Robustness studies of instrument rules produced two interesting results.[27]

First, complex, optimal instrument rules derived from one specific model perform poorly when plugged into a different model. So, using such a rule would be

---

[26] The Taylor rule maximizes a utility function such as (24.30) only then as a strict instrument rule, if sufficient lags make inflation and income predetermined when the interest rate is being set. A pragmatic macroeconomic structure, purported to parsimoniously represent results from typical VARs that interest rates affect income after one year and inflation after two years, that serves this purpose comprises a dynamic *IS* curve,

$$y = -\beta r_{-1} + \gamma y_{-1} + \varepsilon$$

and an accelerationist aggregate-supply curve (without expectations

$$\pi = \pi_{-1} + \delta y_{-1} + \eta$$

where all variables are measured as deviations from their targets.

Minimization of the loss function $\mathrm{var}(\pi) + \xi\,\mathrm{var}(y)$, which directly relates to (24.29), yields an explicit interest rate rule:

$$r = \hat{\alpha}\pi + \hat{\beta}y$$

where the coefficients depend on the model's structural coefficients. See Ball (1999).

[27] See the conference volume edited by Taylor (1999), which focuses on the issue of how robust various policy rules perform in a variety of different macroeconomic frameworks.

very risky if we have serious doubts about the true nature of macroeconomic transmission channels and interaction.

Second, simple instrument rules, taken from the same family as the Taylor rule, do not perform much worse than the complex optimal rule.

Third, and this is actually implied in the second result, the near-optimal performance of simple rules is rather robust across a wide spectrum of models.

Ball (1999) addresses the issue of preference uncertainty, referring to ξ, the relative weight of income stabilization in society's utility function. In terms of Fig. 24.2, he proceeds from the assumption that we do not know society's indifference curves. Then, the best we can do is focus on efficiency and identify those rules or targeting variables that generate lower trade-off lines, so that society can be made better off, no matter what the weight parameter is in its utility function. Employing a calibrated version of the macroeconomic model shown in note 25, Ball compared inflation targeting, nominal GDP growth targeting, and the Taylor rule. In this framework, inflation targeting is efficient and nominal GDP growth targeting is inefficient. The verdict for the Taylor rule is mixed. In its original form reported as (24.29), i.e., endowed with the coefficients of 0.5 advocated by Taylor, the rule is inefficient. In order to make a rule with the same structure as the Taylor rule efficient, the interest rate response to output gaps would have to be about twice as high. This rule can be derived from the efficient inflation target.

Assessment

Monetary policy is an exciting field to work in these days, both for its intellectual and methodological challenges and for its close interaction with policy makers and institutions. From a public choice perspective, nevertheless, and despite the enormous progress that is being achieved, recent developments may cause mixed feelings. In a way one may wonder whether, on an undisputedly higher level of theoretical and methodological sophistication, we are not coming back full circle to fostering and refining the seemingly extinct art of optimal economic policy making as envisaged by Tinbergen and Theil. The resurgence of a more technocratic approach becomes obvious when we interpret recent developments against the political macroeconomics approach that was dominating the discussion until a few years ago and that we traced in the first half of this survey. Its main structure is sketched in Fig. 24.3.

The political-macroeconomics approach has three building blocks: the preferences of society (or voters), the preferences of the policymaker (here the central bank), and a macroeconomic model (usually degenerated into an aggregate-supply equation). Monetary policy conducted within the stochastic macroeconomic model generates economic outcomes that are then evaluated by society on the basis of its preferences. The key result is that monetary policy governed by society's preferences produces an inefficient outcome featuring an undesired, high level of price instability. Society can improve on this suboptimal outcome in a number of ways.
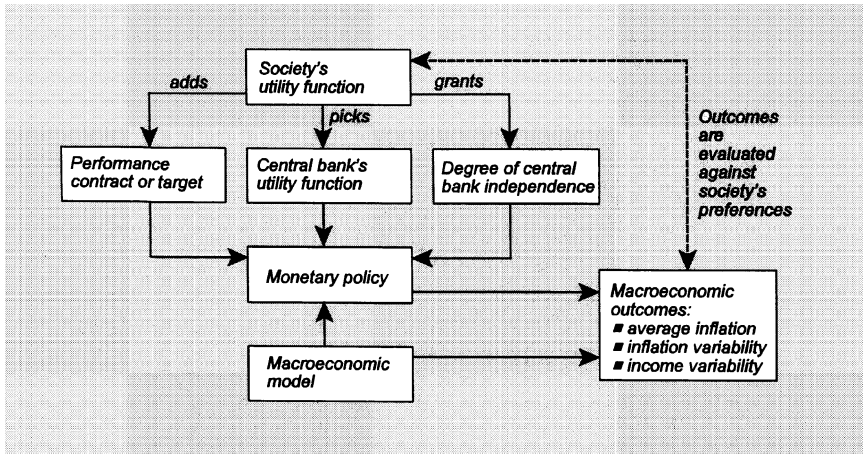
Fig. 24.3 Evaluating outcomes against society's preferences

One way to achieve price stability without distorting the stabilization of shocks is to set new incentives for the central bank by picking a progressive central bank that cares a lot about income, making it independent of the government, and adding an inflation target to its environment.

The current applied discussion of rules and targets questions all three building blocks that characterize the public choice approach:

1. Society has no more desire for income to exceed potential income. So, preferences are compatible with what can be achieved in the long run, both regarding price stability and the level of income. This eliminates the inflation bias, and, hence, the dilemma of a potential trade-off between an inflation and a stabilization bias. In fact, monetary policy governed by society's preferences generates an optimal long-run equilibrium and stabilization as desired.
2. The central bank has no preferences of its own. It can "either be assigned" a loss function (as for instance in the inflation targeting approach), or a reaction function.
3. Finally, and this is one of the strong points, current research about rules and targets accepts as a fact that economists do not agree on a correct macroeconomic model.

On the issue of whether there is a basis for time inconsistency and excessive inflation, it is hard to see why society should settle for potential income as its optimal choice. If potential income is indeed the result of a series of distortions, as is argued for most industrial countries, and comes along with such burdens as involuntary unemployment, shouldn't we want higher income? Do Europeans really not want their 10% a priori risk of being unemployed to fall? This is, in effect, what we are claiming when we argue that the desire for income not to exceed potential income is in our *preferences*. It is something entirely different if we decide that we do not want to draw one monetary policy to raise income. This would be the

result of a cost-benefit calculation on the basis of the macroeconomic options, from which we *might* conclude that a short-lived income hike was not worth the price of a lasting increase in inflation. Our preferences are an element in this calculation, but must not be confused with the calculation itself.[28]

Discarding the central bank's generic preferences and assuming it can simply be assigned any utility function or instrument rule is similarly worrisome. This might be a plausible approximation when fines for deviations from the assigned instrument or targeting rules are so large that personal preferences are dwarfed. But this does not really seem to and cannot really be the idea in a world of change in which rules are at best a frame of reference for policy decisions. None of these rules tells us how to adjust target levels in an evolving macroeconomic environment, how to implement a rule or switch from one to another, how to respond to financial bubbles, or other phenomena outside our standard models.

So, measured against what political macroeconomics achieved and contributed to monetary policy making and designed, current developments may be seen as a setback. Devising optimal rules and targets is certainly useful, but so are plans of how to eat right. The problem is that even its proponents see and sell monetary policy rules as a general framework with plenty of discretion. But then, what are optimality and robustness studies that are based on the strict application of a particular rule worth, if we do not know under what circumstances, how often, and in what direction central banks will deviate from or even change the rule? Such questions obviously cannot be addressed without returning the preferences of society and the central bank back into the equation and, hence, reactivating the public-choice element in monetary policy research.

References

Alesina, A. (1987). "Macroeconomic policy in a two-party system as a repeated game." *Quarterly Journal of Economics*, 102, 651–678.

Alesina, A. and Summers, L. (1993). "Central bank independence and macroeconomic performance." *Journal of Money, Credit, and Banking*, 25, 157–162.

Ball, L. (1999). "Efficient rules for monetary policy." *International Finance*, 2, 63–83.

Barro, R. and Gordon, D. (1983). "Rules, discretion, and reputation in a model of monetary policy." *Journal of Monetary Economics*, 12, 101–121.

Bernanke, B.S. and Mishkin, F.S. (1997). "Inflation targeting: a new framework for monetary policy." *Journal of Economic Perspectives*, 11, 97–116.

Blinder, A. (1995). "Central banking in theory and practice: lecture II: credibility, discretion, and independence." Marshall Lecture, University of Cambridge, May.

Calvo, G. (1983). "Staggered prices in a utility maximizing framework." *Journal of Monetary Economics*, 12, 383–398.

Clarida, R., Gali, J., and Gertler, M. (1998). "Monetary policy rules in practice: some international evidence" *European Economic Review*, 42, 1033–1067.

---

[28] For further arguments on the pros and cons of an income target in excess of potential income, see Walsh (1998), p. 370f.

Clarida, R., Gali, J., and Gertler, M. (1999). "The science of monetary policy: a new Keynesian perspective." *Journal of Economic Literature*, 37, 1661–1707.

Dittmer, R. and Gavin, W.T. (2000). "What do New-Keynesian Phillips curves imply for price-level targeting?" *Federal Reserve Bank of St. Louis Review,* March/April, 21–27.

Drazen, A. (2000b) "The political business cycle after 25 years." *NBER Macroeconomics Annual*, May.

Eijffinger, S. and Hoeberichts, M. (1998). "The trade off between central bank independence and conservativeness." *Oxford Economic Papers*, 50, 397–411.

Forder, J. (1998a). "The case for an independent European central bank: a reassessment of evidence and sources." *European Journal of Political Economy*, 14, 53–71.

Forder, J. (1998b). "Central bank independence—conceptual clarifications and interim assessment." *Oxford Economic Papers*, 50, 307–334.

Frey, B.S. and Ramser, H.-J. (1976). "The political business cycle: a comment." *Review of Economic Studies*, 43, 553–555.

Gärtner, M. (1994). "Democracy, elections and macroeconomic policy: two decades of progress." *European Journal of Political Economy*, 10, 85–109.

Gärtner, M. (2000). "Political macroeconomics: a survey of recent developments." *Journal of Economic Surveys*, 14, 527–561.

Hafer, R.W. and Wheelock, D.C. (2001). "The rise and fall of a policy rule: monetarism at the St. Louis Fed, 1968–1986." *Federal Reserve Bank of St. Louis Review*, January/February, 1–24.

Henderson, D.W. and Mc Kibbin, W.J. (1993). "An assessment of some basic monetary policy regimes pairs: analytical and simulation results from simple multi-region macreocnomic models," in R.C. Bryant, P. Hooper, and C.L. Mann (eds.) *Evaluating Policy Regimes: New Research in Empirical Macroeconomics*. Washington, DC: The Brookings Institution, 45–218.

Herrendorf, B. and Lockwood, B. (1997). "Rogoff's conservative central banker restored." *Journal of Money, Credit, and Banking*, 29, 476–495.

Hibbs, D.A. (1977). "Political parties and macroeconomic policy." *American Political Science Review*, 71, 1467–1487.

Kydland, F.E. and Prescott, E.J. (1977). "Rules rather than discretion: the inconsistency of optimal plans." *Journal of Political Economy*, 85, 473–491.

MacRae, D.C. (1977). "A political model of the business cycle." *Journal of Political Economy*, 85, 239–263.

McCallum, B.T. (1977). "The political business cycle: an empirical test." *Southern Economic Journal*, 43, 504–515.

McCallum, B.T. (1988). "Robustness properties of a rule for monetary policy." *Carnegie-Rochester Conference Series on Public Policy*, 29, 173–204.

McCallum, B.T. (1997). "Critical issues concerning central bank independence." *Journal of Monetary Economics*, 39, 99–112.

Muscatelli, A. (1995). "Delegation vs. optimal inflation contracts: do we really need conservative central bankers?" Discussion Paper No. 9511, University of Glasgow.

Nordhaus, W.D. (1975). "The political business cycle." *Review of Economic Studies*, 42, 169–190.

Peersman, G. and Smets, F. (1998). "Uncertainty and the Taylor rule in a simple model of the euro-area economy," mimeo.

Persson, T. and Tabellini, G. (1990). *Macroeconomic Policy, Credibility and Politics*. Chur: Harwood Academic.

Roberts, J.M. (1995). "New Keynesian economics and the Phillips curve." *Journal of Money, Credit, and Banking*, 27, 974–984.

Rogoff, K. (1985). "The optimal degree of commitment to an intermediate monetary target." *Quarterly Journal of Economics*, November, 1169–1189.

Svensson, L.E.O. (1997). "Optimal inflation targets, 'conservative' central banks, and linear inflation contracts." *American Economic Review*, 87, 98–114.

Svensson, L.E.O. (1999a). "Inflation targeting as a monetary policy rule." *Journal of Monetary Economics*, 43, 607–654.

Svensson, L.E.O. (1999b). "Price-level targeting vs. inflation targeting: a free lunch?" *Journal of Money, Credit, and Banking*, 31, 276–295.

Taylor, J.B. (1983). "Comments on rules, discretion and reputation in a model of monetary policy by R.J. Barro and D.B. Gordon." *Journal of Monetary Economics*, 12, 123–125.

Taylor, J.B. (1993). "Discretion vs. policy rules in practice." *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

Taylor, J.B. (ed.) (1999). *Monetary policy rules. A National Bureau of Economic Research Conference Report*. Chicago: University of Chicago Press.

Theil, H. (1961). *Economic Forecasts and Policy. Volume XV: Contributions to Economic Analysis*. 2nd edn. Amsterdam: North-Holland.

Tinbergen, J. (1952). *On the Theory of Economic Policy. Volume I: Contributions to Economic Analysis*. 2nd edn. Amsterdam: North-Holland.

Waller, Ch. J. (1995). "Performance contracts for central bankers." *Federal Reserve Bank of St. Louis Review*, September/October, 3–14.

Walsh, C.E. (1995). "Optimal contracts for central bankers." *American Economic Review*, 85, 150–167.

Walsh, C.E. (1998). *Monetary Theory and Policy*. Cambridge, MA: MIT Press.

Chapter 25
Regulation And Antitrust[1]

William F. Shughart II

> The state—the machinery and power of the state—is a potential resource or threat to every
> industry in the society. With its power to prohibit or compel, to take or give money, the state
> can and does selectively help or hurt a vast number of industries (Stigler, 1971, p. 3).

In theory, public policies toward business—the regulation of prices and conditions
of entry into specific industries, and the enforcement of antitrust laws that circum-
scribe the conduct of firms more broadly—serve as bulwarks of a freely functioning
market economy. Without such public-sector controls, profit-seeking firms, it is
commonly thought, inevitably would acquire market power and exploit it by
restricting output and raising price, benefiting themselves at consumers' expense.
Government agents must therefore vigilantly stand guard, intervening when neces-
sary to limit the potential abuses of monopoly. Such intervention supposedly is
guided by the goals of ensuring that prices are kept in line with costs, that scarce
productive resources remain fully employed, that technological progress is rapid,
and that economic growth is vigorous.

From this point of view, regulation and antitrust are thrust upon unwilling produc-
ers in order to channel and redirect their behavior away from privately rational, but
socially harmful ends. Business decisions motivated solely by the quest for profit are
displaced by those of public policymakers who pursue broader objectives. Assigning
greater weight to the interests of society as a whole, the antitrust and regulatory
authorities act quickly and appropriately to correct the failures that seem to flourish
in unfettered markets, thereby redistributing wealth back to consumers and enhancing
economic efficiency.

Public choice theory, by contrast, resists modeling public policymakers as disin-
terested, benevolent maximizers of society's welfare (Buchanan and Tullock, 1962).
Built on a foundation of methodological individualism, public choice closes the
behavioral system by assuming that all human actors, in or out of government, pursue
similar objectives (utility maximization) and employ the same rational-choice

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia
of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by
Kluwer Academic Publishers, Volume I, 263–283.

calculus to select the alternative that yields the greatest personal benefit net of cost (Buchanan, 1972). This assumption of universal self-interest, coupled with the logic of collective action (Olson, 1965), implies that the individuals responsible for formulating and executing public policies toward business will have powerful incentives, not selflessly to promote the public interest, but to enhance their own well-being by catering to the demands of politically well-organized special-interest groups.

Applied to public utilities, common carriers, and other "natural monopolies", the economic theory of regulation has revolutionized the study of public policies toward business. As a result of the empirical evidence accumulated over the past quarter century, lending broad support to the theory's implications, few economists now take seriously the naïve view "that regulation is a device for protecting the public against the adverse effects of monopoly" (Posner, 1971, p. 22). The public interest "theory" of regulation (Hotelling, 1938; Joskow and Noll, 1981), which is not in fact a theory in the accepted scientific sense (Posner, 1974; Aranson, 1990), has been displaced by models that bring the tools of microeconomics to bear in analyzing regulation as the product of the supply and demand for wealth transfers. Initially articulated as a theory of regulatory agency "capture" in which "as a rule, regulation is acquired by the industry and operated primarily for its benefit" (Stigler, 1971, p. 3) the theory has been generalized and extended to allow for more complex patterns of wealth transfers amongst the many and varied groups having stakes in regulatory outcomes (Peltzman, 1976; McCormick and Tollison, 1981; Becker, 1983). The constellation of forces at work has been shown to include the industry's customers (Posner, 1974), subsets of heterogeneous producers (Marvel, 1977; Maloney and McCormick, 1982; Anderson and Tollison, 1984; Anderson et al., 1989), and politicians themselves (Crain and McCormick, 1984; McChesney, 1987, 1991, 1997). While disputes continue about the efficiency of the regulatory process (Becker, 1985; Wittman, 1989, 1995; Lott, 1997; Rowley, 1997)—that is, whether competition in the market for wealth transfers is sufficiently robust so as to minimize regulation's deadweight social costs—the economic theory of regulation, which models regulation exclusively as a mechanism of wealth redistribution, is now the reigning paradigm of regulatory analysis.

Such a revolution has not yet materialized fully in the study of antitrust policy. Despite efforts to bring public choice principles to bear in explaining the origins (Baxter, 1980; DiLorenzo, 1985; Stigler, 1985; Libecap, 1992; Boudreaux et al., 1995; Ekelund et al., 1995; Troesken, 2000) and enforcement of the antitrust laws (Faith et al., 1982; Shughart and Tollison, 1985; Shughart, 1990; McChesney and Shughart, 1995), the conventional wisdom that antitrust serves the interests of that most unorganized of groups—consumers—still holds sway. Even the late George Stigler, who did much to undermine the idea "that regulation is instituted primarily for the benefit of the public at large" (Stigler, 1971, p. 3), once called antitrust a "public interest law in the same sense in which. . . private property, enforcement of contracts, and suppression of crime are public-interest phenomena" (Hazlett, 1984, p. 46).

The scholarly disconnect between antitrust and regulation rests partly on a failure to appreciate the regulatory character of many antitrust decrees

(Easterbrook, 1984). In addition, while regulatory policies normally are tailored narrowly to apply to specific firms and industries—and the interest groups having stakes in regulatory outcomes can therefore be identified easily—the antitrust laws supply a broad set of proscriptions on firm behavior that apply to the economy generally. Antitrust's wide reach complicates the identification of winners and losers. Because no one group consistently benefits from antitrust enforcement, the special-interest basis of antitrust policy is less apparent than is the case with other forms of regulation. Last, there is widespread resistance to the idea that the law enforcement agencies and judges who interpret and give effect to the vague language of the antitrust statutes are vulnerable to political influence.

As we shall see, however, antitrust is regulation and, hence, both can be analyzed with the same set of tools. Despite the tenacity of the public-interest view of competition policy (McCormick, 1984), the economic theory of regulation, embellished by public choice principles, helps to illuminate the causes and consequences of antitrust and to situate it within a general economic model of public policies toward business.

Policy Responses to "Market Failure"

Orthodox welfare economics (Pigou, 1932) justifies government intervention into the private economy on the basis of a perceived failure of market institutions always "to sustain 'desirable' activities or to estop 'undesirable' activities" (Bator, 1958, p. 351). Such situations arise when the benefits or costs of a decision or choice at the level of the individual diverge from the corresponding benefits or costs at the level of society, that is, when the parties interacting in a market cannot capture the full social benefits—or do not bear the full social costs—of their resource-allocation decisions. Potential gains from trade remain unexploited when private benefits and costs are not equal to social benefits and costs and, in principle, society's welfare can be improved by appropriate policy intervention.

Consider the case of environmental pollution. In the stylized Pigouvian world, manufacturers of goods that generate toxic wastes as by-products of the production process have little incentive to take account of the costs the pollutants impose on others. The marginal private cost of production, which consists only of the explicit and implicit costs borne by the firm's owners in bringing the product to market, is consequently less than the marginal social cost of production, which includes the additional healthcare expenses and other costs incurred by third parties exposed to the environmental contaminants. Because private costs are less than social costs, the firm produces a quantity of output that is greater than is optimal from society's point of view. Intervention in the form of an effluent fee equal to the difference between private and social costs is the prescribed policy response. Such a tax forces the firm to "internalize the externality", thereby reducing production to the socially optimal rate and supplying tax revenue that can in principle be used to compensate those who are injured by the residual pollutants.

Private markets may likewise fail to achieve ideal results when the social benefits of an activity exceed its private benefits. In deciding whether to be inoculated against a communicable disease, for instance, rational individuals understandably pay greater attention to the expected reduction in their own risks of infection than to the benefits conferred on others, whose risks are also lowered by virtue of immunity that is more widespread. Because inoculated individuals cannot appropriate personally the positive spillover effects of their choices, a public subsidy for vaccines helps align private benefits with social benefits, inducing more individuals to become inoculated than otherwise and thereby correcting the market's undersupply of immunizations. Market failures are also thought to be common when transaction-relevant information is distributed asymmetrically between buyers and sellers and the better-informed parties can exploit their superior knowledge strategically: purchasers of insurance may misrepresent their own risk characteristics in order to obtain coverage at actuarially favorable rates, for example. Similarly, sellers of "experience goods" may, because quality claims cannot be verified prior to purchase, misrepresent product attributes in order to increase their profits at buyers' expense (Nelson, 1970). Public intervention to ensure appropriate information provision is routinely called for in such circumstances. In the limit, private markets may fail completely—and production rights must therefore be assigned to the public sector if any output is to be supplied at all— in the case of pure "public goods" (national defense, for example) whose consumption is nonrivalrous and from whose benefits nonpayers cannot easily be excluded (Samuelson, 1954).

As the foregoing discussion suggests, conventional welfare economics assumes (often implicitly) that while markets are beset with imperfections, the public policy process is not so encumbered. The costs of transacting, including the costs of acquiring, collating, and utilizing information about resource values and of contracting for their exchange—costs that may prevent private economic actors from exploiting all available gains from trade—are ignored when corrective government action is prescribed. Social welfare is invariably enhanced when government intervenes because policymakers are presumed to be fully informed about the social costs and social benefits of resource allocation decisions not taken into account by private decision makers and, moreover, unselfishly to select the appropriate policy response.

This line of reasoning commits what Harold Demsetz (1969) calls the "nirvana fallacy". Market outcomes are generated in a setting in which information is costly to acquire, the future is uncertain, and choices consequently are "boundedly rational" (Simon, 1957). Nevertheless, market performance is usually evaluated, not by way of comparison with other, necessarily imperfect alternatives, but rather in light of the outcomes that would materialize in some idealized and unattainable world in which decision makers are fully informed and endowed with perfect foresight.

Modern approaches to the study of the imperfections associated with externalities, asymmetric information, and public goods raise doubts about their empirical importance (Demsetz, 1970; Coase, 1974; Cawley and Philipson, 1999). More fundamentally, the Coase Theorem (Coase, 1959, 1960) highlights the incentives

of private parties to take account of the external costs and benefits of their resource allocation decisions and to contract their own way around "market failure". Consider apple growing and beekeeping (Cheung, 1973). Apple growers benefit from the pollination services of bees and beekeepers benefit from a ready supply of apple blossom nectar. A complex set of bilateral contracts has evolved that compensates each party for their joint contributions to the apple and honey crop. Such contractual solutions to market failures require only that property rights be defined clearly and that transaction costs be less than expected gains. Indeed, the contours of the efficient solution (but not the distribution of income) are invariant to the initial property rights assignments. The Coase Theorem emphasizes that not every potential market failure demands a government response: private parties may fail to reach agreement, not only because the costs of doing so are high, but also because the anticipated benefits are low. The theorem emphasizes in addition that, even when social welfare potentially can be enhanced by government intervention, the knowledge limitations confronting policymakers and the costs of government intervention must be considered before corrective action is taken.

The Regulatory Nirvana

Nowhere is the nirvana fallacy committed more regularly than in the analysis of perceived market failures due to monopoly, to which regulatory and antitrust policies have arisen in response. The textbook model of "perfect competition" remains the standard by which the conduct of flesh-and-blood producers is evaluated by those who formulate and execute public policies toward business. In that model, rivalry between firms, by any commonsensical definition of the concept, is assumed away. Competition is "perfect" in the model of perfect competition because large numbers of firms offering identical products for sale interact with large numbers of consumers making offers to buy, there are no barriers to the entry of new firms into the market (and no barriers to the exit of old ones), and all transaction-relevant information, including information about the locations of sellers and the prices they charge, the quality attributes of their products, and the requirements and creditworthiness of buyers, is freely available to all.

Under such circumstances, long-run market equilibrium is characterized by *allocative efficiency* and by *productive efficiency*. Since the product offered for sale by any one firm is, by assumption, identical in all respects to the products offered by its "rivals", no seller can charge a price greater than marginal production cost. Because there is no product differentiation in the model of perfect competition, buyers select among sellers solely on the basis of price; they are otherwise indifferent as to the identity of the firm from which they make their purchases. The demand curve facing an individual seller consequently is perfectly elastic (horizontal) at the market-determined price (which is equal to marginal cost): any firm attempting to raise its price above marginal cost would immediately see its customers switching their purchases to rivals charging lower prices. No firm possesses

market power, defined as the ability to raise price without losing all of its sales; each is a *price taker*, whose only decision is how much output to produce at the going market price.

When the firm (and the industry) expands output to the point at which price is equal to marginal cost, the value consumers place on the last unit produced (the amount they are willing to pay for it) is just equal to the value (opportunity cost) of the resources consumed in producing that unit. From society's point of view, price equal to marginal cost yields an efficient allocation of the economy's scarce sources in the sense that producers' decisions about how much to produce dovetail with consumers' decisions about how much to buy. Neither too few nor too many resources are devoted to the production of the good in question. As Goldilocks might say, the quantity of resources consumed by the perfectly competitive industry is "just right".

In addition, since there are, again by assumption, no barriers to the entry of new firms into the industry, sellers cannot earn positive economic profits in the long run. The rate of return on invested capital in the perfectly competitive industry is driven to normal levels—equal to the rate of return on the next best alternative investment opportunity. With above-normal profits eliminated by new entry (and below-normal profits eliminated by exit), market price (average revenue) is not only equal to marginal cost, it is equal to average cost as well. Given that marginal cost is equal to average cost only at the latter's minimum point, which is in turn the only point consistent with long-run, zero-profit industry equilibrium. What is true for the industry must also be true for every firm in it. Price equal to marginal cost and zero profits implies that the horizontal demand schedule perceived by price-taking firms must be tangent to the minimum points on their respective average cost curves. This is the hallmark of *productive efficiency*. Given existing technologies and resource prices, the perfectly competitive industry produces its product at the lowest possible cost per unit. Things are once again "just right": the industry consists of the socially optimal number of firms each of which employs its production capacity at the efficient (cost-minimizing) rate.

The Welfare Costs of Monopoly

The belief that actual markets frequently fail to achieve ideal textbook results supplies the principal justification for antitrust and regulatory intervention into the private economy. But using the model of perfect competition in this way commits the nirvana fallacy. Real producers do not conduct business in a frictionless world of homogeneous products, zero transaction costs, and perfect knowledge. Owing to differences in quality, reputation, location, and so on, each seller's product or service has one or more unique attributes that distinguish it in the minds of consumers from the products or services sold by its rivals. The offerings of sellers in most markets are good, but not perfect substitutes for one another and buyers typically have prefer-ences for one particular brand (and are therefore willing to pay more for it). The demand schedule confronting each firm slopes downward under these very

common circumstances, and because of this neither allocative nor productive efficiency can possibly be attained.

In the presence of product differentiation (and the downward-sloping demand curves to which it gives rise), the assumptions of the model of perfect competition no longer apply. In order to attract customers away from the sellers of substitutable products and to increase its own sales, each firm must be prepared to engage in the commonplace types of rivalry assumed away for model-building purposes. They advertise and promote their products, engage in research and development, and offer pre- and post-sale services, warranties, convenient locations and hours of operation, to name a few of the many available methods of nonprice competition. In addition, of course, the quantity sold by a firm facing a downward-sloping demand curve can be increased only if its product's price is reduced. Downward-sloping demand also implies market power: the firm can raise its price without losing all of its sales.

The firm with market power does not take price as given, but instead searches for the price that maximizes its profits and, as every sophomore knows, that profit maximum occurs at an output rate that is lower (and a price that is consequently higher) than would be chosen by a perfectly competitive industry facing the same demand and cost conditions. In other words, the firm exploits its market power by restricting the number of units offered for sale below the competitive level. This output restriction reduces the welfare of consumers in two ways. First, because price exceeds average cost, at least in the short run (see below), income is redistributed from buyers to the seller in the form of pure economic profit. (This redistribution, by itself, usually is treated as a pure income transfer having no impact on the welfare of society as a whole: the seller's gains exactly offset consumers' losses.) Second, because price also exceeds marginal cost, additional surplus is transferred away from consumers which, not being captured by the seller, imposes a "deadweight" welfare loss on society (Harberger, 1954). When all units of output are sold at the same price (i.e., the seller does not engage in price discrimination), this deadweight social welfare loss materializes because by restricting production below the competitive level, the firm fails to supply units of output for which consumers are willing to pay more than it would cost to produce.

Markets populated by firms possessing market power thus fail to achieve desirable results in the sense that fewer units of output are produced (and fewer resources are therefore allocated to production) than is optimal when benchmarked against the textbook model of perfect competition. Price in excess of marginal cost impairs allocative efficiency. Moreover, although productive efficiency is achieved by firms with market power under constant-cost conditions, only by coincidence will such firms produce their outputs at rates corresponding to minimum average cost with more generic U-shaped cost curves.

The polar case of market power is monopoly, defined as a market served by a single firm producing a product having no close substitutes. Whether any firm possessing market power, including a monopolist, is able to earn above-normal profits in the long run depends critically on the conditions of entry facing newcomers to the industry. Consider a market that is perfectly "contestable", for example. Firms contemplating entry into such a market do not bear any costs not borne by the

established firm(s)—and firms exiting the industry can recoup their prior investments net of depreciation. Under these conditions, prices and profits must stay at competitive levels regardless of the number and size distribution of incumbent producers (Baumol et al., 1982).

Efficiency or Redistribution?

It is not the distribution of income between producers and consumers that is the stated concern of public policies toward business. Profits, after all, play an indispensable role in market economies, helping to guide alert entrepreneurs to redeploy scarce productive resources from less highly valued to more highly valued uses. Rather, it is the existence of allocative inefficiency ("deadweight" social welfare loss) that supplies a theoretical justification for government intervention into sectors of the economy ostensibly plagued by market power. Although the deadweight losses due to monopoly do not seem to loom large empirically (Harberger, 1954; Posner, 1975) and therefore are offset by even modest efficiency gains (Williamson, 1968a, 1977), the presumption is that appropriate public policies can and will be employed systematically to identify and correct these market failures, thereby restoring competitive results. In principle, the public's interest will be served—society will experience a net gain from such intervention—as long as the cost of implementing pro-competitive public policies is less than the associated improvement in market efficiency.

The proponents of an activist anti-monopoly policy have also pointed to the possibility of "X-inefficiency" (Leibenstein, 1966, 1978), the idea that the managers of firms insulated from competitive market forces have weak incentives to employ the resources at their command cost-effectively. Less competition leads to internal waste and therefore less efficiency. While Leibenstein did not identify the sources of such waste precisely (hence the "X" in "X-inefficiency"), he argued that such losses would far outweigh any cost savings (i.e., scale economies) otherwise associated with monopoly. Thus, in contrast to Williamson's trade-off model, it is productive efficiency, not allocative efficiency, that should loom large in justifying public policies toward business. If freedom from competition makes it possible for managers to be "X-inefficient", then policies aimed at increasing firms' exposure to competitive market forces will produce significant efficiency gains. Subsequent work has indicated, however, that this remains an open question (Stigler, 1976; Jensen and Meckling, 1976; De Alessi, 1983; Bertoletti and Poletti, 1997; Schmidt, 1997; for a general application of these ideas to regulation and antitrust, see, Rowley and Peacock, 1975).

On the other hand, the public choice model stresses that issues of income distribution will tend to carry greater weight in the public policy process than concerns of economic efficiency. Groups that stand to gain or lose wealth because of policies targeting perceived sources of market failure will coalesce around the policy process in order to advance or protect their own parochial interests. Politicians and policymakers will respond rationally to and balance these competing

demands, and in doing so tend to give preference to those constituencies best able to support them politically in exchange for favorable treatment. The economic theory of regulation (including antitrust regulation) is thus about the political pressures that impinge on the elected officials who enact the legal rules delineating regulation's scope, and on the agencies whose bureaucrats enforce those rules. Depending on the policy process in question, the beneficiaries of regulation may turn out to be almost any well-organized special-interest group. Owing to the fact that "the public" is numerous, geographically dispersed, and, in general, unorganized politically, its influence on the policy process is necessarily weak. Public regulation of private industry therefore will rarely, if ever, serve the public's interest.

Regulation

> The 'protection of the public' theory of regulation must say that the choice of [oil] import quotas is dictated by the concern of the federal government for an adequate supply of petroleum in the event of war—a remark calculated to elicit uproarious laughter at the Petroleum Club. (Stigler, 1971, p. 4)

The extent of public regulation of industry in the United States—and elsewhere—is both broad and deep. To name just a few, rules—and agencies to enforce them—have been established to require the disclosure of financial information to investors; to license physicians, hospitals, attorneys, accountants, stockbrokers, barbers, electricians, plumbers, morticians, and taxicab operators; to regulate advertising claims; to enforce environmental quality, workplace safety and product safety standards; and to promote equal opportunity in employee hiring and promotion decisions. While all such regulatory regimes are worthy of study from an interest-group perspective (Stigler, 1988), because of its historical significance in justifying regulatory controls on private industry, the case of immediate interest here is that of so-called natural monopoly.

Natural monopoly "does not refer to the actual number of sellers in a market but to the relationship between demand and the technology of supply" (Posner, 1969b, 1999, p. 1). In particular, a monopoly is said to be "natural" if, first, the production of the good or service in question is characterized by robust economies of scale, that is, long-run average costs fall sharply over the relevant range of output rates. Scale economies will loom large if production technologies are subject to increasing returns (proportional increases in input usage produce greater than proportional increases in output), if large capital investments must be made before production begins, but the cost of producing additional units or of serving additional customers is comparatively low from then on, or both. Second, monopoly is natural if, in the presence of significant economies of scale in production, the demand schedule intersects the long-run average cost curve at a point where the latter is still declining.

The existence of scale economies up to the level of market demand is sufficient for establishing natural monopoly when the firm produces a single product. In modern parlance, scale economies are a strong form of "cost sub-additivity",

meaning that there is no feasible way of subdividing the firm's quantity of output, Q, having each subpart produced by separate firms, without incurring higher total costs. In other words, "the cost of producing the whole is less than the sum of the costs of producing the parts" (Baumol et al., 1982, p. 17). Strict cost sub-additivity (and, hence, natural monopoly) in the multi-product case requires both economies of scale and economies of joint production, the latter representing situations in which the total cost of producing the individual products by separate firms is greater than the total cost of having all of them produced by the same firm (Tirole, 1988, pp 19–20; Sherman, 1989; Spulber, 1989).

The efficiency results of free and open competition may not be achievable under the conditions defining natural monopoly. Because of the peculiar relationship between demand and cost, one firm can supply the entire market more efficiently than two or more firms, each of which, owing to the strictures of cost sub-additivity, would necessarily incur higher unit costs. Indeed, an alternative definition of natural monopoly is an industry in which one firm is viable (i.e., earns positive economic profits), but two or more firms are not (Tirole, 1988, p. 20). Two unhappy outcomes are then possible: "either the firms will quickly shake down to one through mergers or failures, or production will continue to consume more resources than necessary" (Posner, 1969b, 1999, p. 1). In the former case, left unrestrained, the sole survivor rationally will restrict the number of units it produces below the competitive level and raise its price to the monopoly profit-maximum. In the latter case, capital investments will be wastefully duplicated from society's point of view in the sense that production on a larger scale by a single firm would yield substantial improvements in economic efficiency.

Regulation of natural monopoly is thus justified normatively on the grounds that, while society would benefit from the production efficiencies achieved by having the market served by a single firm, allocative efficiency will be impaired if the monopolist remains free to exercise his market power. Society can in principle have it both ways if government intervenes by, on the one hand, assigning exclusive rights to produce the good or service in question to one firm and, on the other hand, imposing regulatory controls that require the franchisee to expand production and lower price, thereby approximating competitive market outcomes.

Construed narrowly, the natural monopoly justification for regulation rests on the fulfillment of extreme assumptions and, as such, applies only to a limited set of "public utilities". Water and sewer systems, electric power grids, and telecommunication networks, long regulated by local, state, and national governments in the United States and elsewhere, are prime examples. Even in these textbook cases, however, the theoretical rationale for regulation is weak. Competition for the field can substitute for competition within it (Demsetz, 1968; Williamson, 1976). Additionally, "access pricing", whereby rival suppliers pay for the right to utilize the large-scale infrastructure necessary to serve public utility customers, avoids duplicative investments in production capacity and promotes efficient utilization of that capacity (Shy, 2001, p. 8). It also turns out that if an unregulated natural monopolist operates under conditions of contestability, it will charge Ramsey-optimal prices in all markets

(i.e., prices that are inversely proportional to the elasticity of demand in each market; see Ramsey, 1927), subject to a minimum profit constraint that ensures viability (Baumol et al., 1977; Baumol et al., 1982; Tirole, 1988, pp 308–309). Hence, while the particular configurations of cost and demand defining natural monopoly prevent the attainment of first-best outcomes, they may not preclude second-best optima even in the absence of regulation.

In practice, even otherwise staunch supporters of active government involvement in the economy admit that arguments based on "trumped-up claims of monopoly" (Scherer, 1980, p. 482) frequently have been appealed to in order to widen regulation's scope far beyond the limited set of industries for which public-sector controls might theoretically be defensible. Indeed, public regulation of industry in the United States began at the federal level in 1887 with passage of the Act to Regulate Commerce, which established the Interstate Commerce Commission (ICC) and delegated to that agency authority to ensure that railway rates were "just and reasonable". The ICC's regulatory powers ultimately were expanded to include control over most surface (and some subsurface) interstate transportation modes, including inland water carriers, trucks, busses, and crude oil pipelines—industries to which the natural monopoly label does not obviously apply.

Public regulation of industry in the United States in fact predates the Act to Regulate Commerce by a decade. In 1877, the U.S. Supreme Court ruled that when private property is "affected with a public interest", regulation is constitutionally permissible despite the Fourteenth Amendment's guarantee that "no State shall. . . deprive any person of life, liberty, or property without due process of law". That ruling was handed down in *Munn* v. *Illinois*, 94 U.S. 113, a case challenging a provision of the Constitution of the State of Illinois designating privately owned grain elevators as public warehouses as well as a law passed by the Illinois legislature in 1871 prescribing maximum rates for grain storage. Munn and Scott, two grain elevator operators, had been convicted of charging higher rates than the law allowed; the Supreme Court upheld their conviction. In the following years, the Court construed the "public interest" standard strictly, approving state regulation of only a select group of industries, including in addition to grain storage, banks, fire insurance companies, and insurance agents. Limited as the early extensions of regulation may have been, none of the newly regulated industries plausibly were monopolies, natural or otherwise.

But the Court rejected even the narrow construction of *Munn* in 1934, declaring that "the phrase 'affected with a public interest' can, in the nature of things, mean no more than that an industry, for adequate reason, is subject to control for the public good' and that 'there can be no doubt that on proper occasion and by appropriate measures the state may regulate a business in any of its aspects, including the prices to be charged for its products or the commodities it sells". In that 1934 case, *Nebbia* v. *New York*, 291 U.S. 502, a case in which the Court upheld the right of New York's Milk Control Board to regulate milk prices in the state, the Court in effect ruled that there is no constitutional distinction between public utilities and other industries. The states were thereafter free to regulate any business operating within their jurisdictions for any reason public officials could rationalize

as promoting the public interest, so long as the regulation was "neither arbitrary nor discriminatory".

Thus were the regulatory floodgates opened. Any pretense that natural monopoly conditions explained the onset of economic regulation was gone. (The evolution of legal doctrines underpinning economic regulation in the United States is discussed more fully in Shughart, 2004.)

With their stronger and more overt socialist heritages, many European governments adopted a different method for dealing with alleged natural monopolies. At least until recently, public ownership rather than regulation of privately owned firms has been the norm there. (Though much rarer, such a policy approach is not unknown in the United States: the electric power industry, e.g., contains a mix of investor-owned, publicly owned, and customer-owned companies; virtually all of America's local public transit systems, many of its municipal refuse collection services and, most infamously, its postal delivery service, are also operated as government enterprises.) As with public regulation in the United States, public ownership in Europe expanded far beyond the bounds set by the peculiar configurations of cost and demand defining natural monopoly. In addition to the traditional public utilities, commercial airlines, railroads, banks, television and radio networks, and telephone systems are (or have been) nationalized. So have the manufacturing of steel, automobiles and aircraft, and the extraction and processing of oil, coal, and other natural resources. While an analysis of state-owned enterprises is beyond the scope of the present discussion, it suffices to say that, because of weaker incentives for using resources efficiently (Alchian, 1965; De Alessi, 1982, 2001), publicly owned firms are predicted to perform poorly by market standards. That prediction is borne out by an extensive empirical literature (Shughart, 1997, pp 295–301). Public enterprise even seems to be inferior to (less technically efficient than) regulation (Rowley and Yarrow, 1981).

The American and European paths continue to diverge even now. Beginning with the domestic commercial airlines, a wave of deregulation has been underway in the United States since the late 1970s. The United Kingdom embarked on a program of privatization the following decade, and the number of state-owned enterprises on Europe's endangered species list increased dramatically with the collapse of the Soviet Union. But public ownership of industry in Europe has been replaced, not by a hands-off approach to the private sector, but by a new emphasis on regulation. A comparison of traditional and interest-group theories of regulation helps to shed light on these developments.

The Standard Theory

Given that, absent contestable market conditions, an unregulated natural monopolist rationally would restrict output, raise price, and thereby earn above-normal profits, the case for government intervention rests on the theory that an industry-specific regulatory agency can and will impose controls that allow the substantial economies of single-firm production to be achieved, while at the same time forcing

prices and profits to competitive levels. But a serious problem arises at the outset: because marginal cost lies below average cost when the latter is falling, mandating a price equal to marginal cost would cause the regulated firm to incur losses and a subsidy would therefore be necessary for it to remain viable in the long run. Market demand would be satisfied and the regulated firm would be constrained to earning a normal profit if it were allowed to charge a price equal to average cost, but that price would necessarily exceed marginal cost. Thus, all orthodox theories of regulation are inevitably concerned with trade-offs between productive efficiency, allocative efficiency, and sustainability.

There is a second, perhaps more serious, problem confronting the regulators. The ostensible goal of regulation is to induce the regulated firm to produce and price "optimally" (Train, 1991). If that goal is to be achieved, the regulatory agency must be fully informed about the cost and demand conditions facing the firms it is responsible for regulating. It is reasonable to assume, though, that regulated firms will have both more and better information about their own costs and the values their own customers place on the goods or services they produce than will the regulatory agency's staff, no matter how expert they may be. The very real possibility exists that regulated firms will exploit their superior knowledge to persuade regulators to approve rate requests that depart from optimality. Mechanisms must therefore be designed that provide incentives for regulated firms truthfully to reveal the specialized information in their possession. The complexity of the regulatory process is increased—and strategic misrepresentation of relevant information becomes more likely—owing to the fact that considerations of allocative efficiency and sustainability typically force regulatory agencies to contrive a schedule of allowable rates rather than permitting them to deal with the much simpler problem of approving a single price that all customers will pay.

The necessity of creating a schedule of allowable rates arises whenever customers differ in their marginal valuations of the good or service supplied by the regulated firm. Under such quite common circumstances, requiring the regulated firm to charge the same price to all buyers generates allocative inefficiency even if that price is set at the proper break-even level that allows the firm to earn a normal profit. This allocative inefficiency results from the fact that some customers—those who would choose to purchase the good or service if it were priced at marginal cost—will not be willing to buy at the higher average-cost price. In addition, as noted above, because marginal cost lies continuously below average price under natural monopoly conditions, no single market-clearing price equal to marginal cost exists at which the regulated firm can avoid economic losses.

Price discrimination is the standard solution to this problem. Following this approach, customers are segregated into different classes based on their elasticities of demand for the regulated firm's product and a separate rate is set for each class that is inversely proportional to its demand elasticity (Ramsey, 1927). The result of price discrimination is higher rates for those customers having less elastic demands and lower rates for those customers having more elastic demands. It is not unusual, for example, for industrial customers to be required to pay more for electric power than commercial (small business) customers, who are in turn required to pay more

than residential customers. Such discriminatory rate structures help achieve regulation's twofold objective, at least in principle. First, charging different prices to different classes of customers increases the regulated firm's revenues over and above those that would be earned under a single-price policy. Price discrimination thus makes it more likely that the firm will break even. Second, tailoring prices more closely to customers' marginal valuations works to mitigate allocative inefficiency.

Other regulatory pricing schemes for helping resolve the twin problems of efficiency and sustainability include "peak-load pricing", which involves charging higher prices to all customers when the demand for the regulated firm's good or service rises systematically relative to normal demand. Differentially higher electricity rates during the summer months and differentially higher public transit fares during "rush-hour" are relevant examples. "Multi-part pricing", in which customers pay a fixed service connection charge upfront plus a price per unit of service consumed that approximates the marginal cost of supplying them, is another alternative, as is a rate schedule that declines in stepwise fashion as additional "blocks" of service are consumed.

It should be obvious, however, that, in the presence of imperfect (and perhaps strategically false) information, diversity in customers' demands and differences in the costs of serving them, "optimal" regulation will be elusive (Coase, 1946). The orthodox case for regulating natural monopoly is undermined further by considering some rational behavioral responses to it. Traditional public utility regulation requires the regulatory agency to establish schedules of allowable rates consistent not only with the goal of improving allocative efficiency, but also with an eye toward preventing the regulated firm's revenues from breeching an overall profit constraint, thereby ensuring that the firm's owners earn only a normal or "fair" return on their investments. Under such a regulatory regime, the regulatory agency is obligated to pass through to customers the cost of any physical capital it permits the firm to add to its installed "rate base" (the value of the stock of capital on which the regulated rate of return is computed) plus an allowance for normal profit. Because prices must be increased by more than the cost of additions to the rate base in order to ensure that the regulated firm continues to earn a "fair" return, the cost of capital is effectively lowered. As a result, the regulated firm has an incentive to invest in more capital than it would in the absence of regulation (Averch and Johnson, 1962; Baumol and Klevorick, 1970). Rate-of-return regulation may therefore compromise the regulated firm's productive efficiency by inducing it to select an input combination that is too capital-intensive compared with the combination that is optimal from society's point of view. Firms subject to rate-of-return regulation also have incentives opportunistically to evade the regulatory profit constraint by diversifying into unregulated lines of business and then adopting internal-to-the-firm transfer pricing policies that reallocate recorded profits away from core activities subject to regulatory control.

So-called price-cap regulation supplies a somewhat different set of incentives (Acton and Vogelsang, 1989; Train, 1991: 317–319). Adopted by the U.S. Federal Communications Commission in mid-1990 for regulating long-distance telephone rates and by regulatory authorities in the United Kingdom for regulating natural

gas, electric power, and water utilities, the regulated firm is permitted to earn a rate of return that exceeds the ceiling that would otherwise be imposed in exchange for agreeing not to raise its prices by more than allowed under a predetermined formula. That formula is of the form *CPI—X*, where *CPI* is the annual rate of increase in an index of retail prices and *X* is some specified percentage less than the measured economy-wide inflation rate. In other words, the public utility or common carrier is authorized to raise its prices only if the rate of inflation is greater than *X*, and then only to the extent that the *CPI* exceeds that threshold.

Price-cap regulation has two advantages over traditional rate-of-return regulation. First, because input prices are not distorted and the regulated firm can keep any and all of the profits it earns under the price cap, it will choose efficient methods of production. In addition, the firm has an incentive to implement any cost-reducing innovations it discovers, again because it can keep all of the realized profits. (It does not have an incentive to pass those cost savings on to consumers in the form of lower prices, however.) Second, regulatory rate hearings are greatly simplified: requests for price increases are approved automatically, subject only to the regulatory agency's determination that the proposed increase satisfies the agreed-to pricing formula.

It is nevertheless true that, as with all other forms of regulation, the information required to implement price caps largely must be obtained from the regulated firm itself. Depending on how methodically regulatory rate hearings are conducted and how aggressively the regulatory agency adjusts price caps over time to take account of changing conditions of cost and demand, the utility may be able to exploit its superior knowledge to benefit its owners and managers at consumers' expense.

In the end, and in spite of a large and elegant scholarly literature prescribing mechanisms for dealing with the complexities of regulation,

> its contribution to social and economic welfare is very possibly negative. The benefits of regulation are dubious, not only because the evils of natural monopoly are exaggerated but also because the effectiveness of regulation in controlling them is highly questionable. (Posner [1969b] 1999, p. 106)

The conclusion that regulation often fails to achieve its stated goals garnered empirical support in an initial series of studies examining its actual effects. A study of electric utility regulation, for example, found that it had little or no impact on the level of prices or on the rates of return to investments in that industry (Stigler and Friedland, 1962). In another study, investors were found to have obtained few benefits from the regulatory oversight of new stock issues by the U.S. Securities and Exchange Commission (Stigler, 1964). Indeed, a survey of the early empirical literature supporting the "capture" theory of regulation suggested that, while regulatory intervention was not always as ineffective as Stigler and Friedland had found—as a matter of fact, they were later shown to have been wrong (Peltzman, 1993)—in those industries where regulation did affect prices and profits, the effects were perverse (Jordan, 1972). In particular, when applied to naturally competitive industries, such as air and surface transportation, regulation uniformly was found to have reduced the number of competitors and to have raised prices. On the other

hand, when applied to industries more plausibly characterized by natural monopoly conditions, regulation had no effect on prices.

These empirical findings raised two important questions. If consumers' interests are not materially advanced by regulation, why is regulation adopted in the first place? Given the very real costs of regulation—the costs to taxpayers of defraying the expenses of the regulatory agencies and the costs to society in the form of the resources misallocated by poorly crafted regulatory constraints and the resources consumed in attempts to influence the regulatory process—why does regulation persist?

The initial answer to both of these questions was that, despite their well-intentioned purposes, regulatory agencies are vulnerable to "capture" by the very firms they were created to oversee. Public institutions ostensibly designed to protect consumers from the abuses of monopoly in practice catered chiefly to the interests of producers. This "capture" theory of regulation has subsequently been formalized and extended into what is currently known as the economic theory of regulation, to which the discussion now turns.

The Economic Theory of Regulation

Regulation creates rents for the regulated. By virtue of the exclusive franchises they have been granted, the owners of regulated firms are in position to earn profits in excess of normal levels. Regulatory agencies are of course charged with the responsibility of ensuring that public utilities and other natural monopolies do not exercise their market power, imposing controls on price that allow the owners to earn only "fair" rates of return on their investments. But if the profits of regulated natural monopolies were in fact typically equal to the rate of return available in the next best alternative investment opportunity, there would be no need for rules governing the conditions of entry into the regulated industry. The regulation of entry into a market with natural monopoly characteristics can be justified on the basis of a social-welfare standard only if prices would be too high after entry, not if they would be too low. If prospective entrants anticipate that entry will cause prices to fall below average cost so that all firms, including the incumbent, stand to incur economic losses, *then they will not enter*. Legal barriers to entry, usually imposed in the form of requirements that newcomers to the market obtain a "certificate of convenience and necessity" from the regulatory agency prior to entering, consequently supply prima facie evidence that the profits of regulated firms are often above normal levels.

Regulated firms may earn supranormal profits in only some of their markets. Regulatory mandates requiring public utilities to serve all of the customers in their territories regardless of cost (so-called "universal service" requirements) and rules designating shippers as "common carriers", may force regulated firms to serve markets they would not serve in the absence of regulation. In such cases, and as an alternative to explicit subsidy, the regulatory agency must permit the regulated firm to make up losses in markets where revenues are less than costs by charging higher

prices (and earning profits exceeding normal levels) in more remunerative markets. A case can be made for erecting legal barriers to entry into the firm's paying markets on the grounds that, without the supranormal returns obtainable there, owners would not earn a "fair" return overall.

Regulatory cross-subsidies of these kinds are quite common, so common in fact that Richard Posner (1971) calls such pricing schemes "taxation by regulation": some of the excess returns associated with exclusive natural monopoly franchises are taxed away by regulators in the form of requirements to serve customers that would not be served otherwise. The implication is that regulators allow regulated firms to charge some of their customers prices that exceed the costs of serving them so that other customers can be served at prices that are less actual costs. One allocative inefficiency is introduced to sponsor another.

The history of deregulation (about which more later) suggests, however, that prices exceed costs in most, if not all, regulated markets. The rates paid by long-distance telephone customers, for example, supposedly kept high by regulators in order to subsidize local telephone customers, have fallen dramatically since MCI and other competitors began entering the industry in 1982. (Deregulation was initially opposed by the incumbent regulated monopolist AT&T on the basis that, if MCI was allowed to "skim the cream" from its most profitable market, AT&T's ability to fulfill its universal local telephone service obligations would be severely compromised.) But local telephone rates are also falling as competition emerges in those markets. To be sure, lower local and long-distance telephone rates are explained in part by the rapid pace of technological change in the telecommunications industry since the early 1980s—events which themselves owe much to the competitive market forces unleashed by deregulation. It is nevertheless reasonable to conclude from this and similar experiences in the commercial airline industry, the trucking industry, and the natural gas transmission industry, among others, that prices (and profits) under regulation tend to exceed normal levels.

Whether rents exist in all or only some of a regulated firm's markets, however, their existence begets rent-seeking (Tullock, 1967; Krueger, 1974). Individuals and groups rationally strive to put themselves in position to earn above-normal returns and, moreover, are willing to invest resources for the purpose of capturing them equal to the expected value of their anticipated gains (Posner, 1975; Tullock, 1980). In the case at hand, rent-seeking materializes in the form of lobbying activities calculated to influence the regulatory process. George Stigler (1971) modeled regulation largely as a struggle between producers and consumers for access to the rents associated with conditions of natural monopoly. Given that the members of the regulated industry normally would be better informed about the regulatory process, have greater financial stakes in regulatory outcomes, and, owing to their smaller numbers and more cohesive objectives, be better organized and, hence, more effective in bringing influence to bear on the regulatory agency, regulators would tend to favor their interests over those of consumers. The essence of the "capture" theory of regulation is that "consumers are the least organized and therefore typically the least effective interest group. The long-run consumer interest in particular has no lobby" (Posner [1969b] 1999, p. 67).

In Stigler's formulation of the problem, producers dominate the regulatory process and no one should therefore be surprised that, from society's perspective, public regulation of industry is ineffective or perverse: "Consumers never asked for an Interstate Commerce Commission to prevent new truckers from entering the business. Nor had consumers been heard from when the federal government set up milk marketing boards to restrict the supply of milk and drive up the price. The main players were truckers and milk producers, who wanted to limit competition" (Henderson, 1995, p. 62).

A subsequent formalization of the economic theory of regulation (Peltzman, 1976) supplies a more general framework for thinking about the problem. In that more general theory, the regulators themselves are portrayed as rational, self-interested actors whose objective is to maximize their own political support. Where they hold elective office, "political support" can be defined in terms of votes, campaign contributions, or both, in which case regulators are assumed to be motivated by the goal of maximizing their probability of reelection. Where they hold appointive office, regulators strive to maximize their probability of reappointment or some other index of job security. An even more universal behavioral assumption is utility (wealth) maximization, a maximand which includes the regulator's salary and perquisites of public office as well as income received from post-government employment, which, because of the specialized knowledge gained in participating in the regulatory process, not infrequently will be a job in the regulated industry itself. In any case, the interest-group theory of regulation rejects the analytical inconsistencies of the "public-interest" theory, which places regulators outside the model and does not therefore inquire into their motives. Everyone involved in the regulatory process, including the regulators themselves, is thereby brought within the ambit of positive economic analysis.

As in all economic models of human behavior, the regulator's pursuit of self-interest is not unconstrained. In Peltzman's framework, the regulator selects the price the regulated firm is permitted to charge. This price can be set at the competitive level, in which case the regulated firm earns a normal profit and consumers enjoy all of the gains associated with regulation. The price can also be set at the monopoly profit-maximizing level, in which case producers are regulation's sole beneficiaries. In general, however, the politically self-interested regulator must balance the demands of both groups. While an increase in price (and profit) elicits greater political support from the regulated firm(s), it also invites greater opposition from consumers. Lower prices invoke the opposite reactions. If the political returns to higher profit or lower price are diminishing at the margin, neither group will get all that it wants from regulation: from the regulator's point of view, the optimal price will lie somewhere between the extremes of competition and monopoly. Where the balance is struck in any particular case depends on the configurations of the costs and benefits of bringing political influence to bear on the regulatory process facing the groups having stakes in the outcome.

Like the public-interest theory, the Stigler–Peltzman model predicts that regulation will target natural monopolies and that, to the extent to which losses in political support from the regulated firm are offset by increases in support from consumers,

regulators will require the regulated firm to charge a lower price than it would otherwise. But unlike the public-interest theory, the Stigler–Peltzman model helps to explain why regulatory controls have in practice been applied to industries that would otherwise be competitive. If regulators can increase their political support by mandating that prices be raised above competitive levels, they will rationally do so up to the point where the additional support provided by producers equates at the margin to the loss in support from consumers.

Although the discussion thus far places the Stigler–Peltzman model of regulation in its original, highly stylized producer–consumer context, the economic theory of regulation is in fact much more general. Because "the political process automatically admits powerful outsiders to the industry's councils" (Stigler, 1971, p. 7), regulatory outcomes will assimilate the interests of any individual or group that can bring effective influence to bear on the regulators. For example, "it is well known that the allocation of television channels among communities does not maximize industry revenue but reflects pressures to serve many smaller communities" (ibid.). The regulatory subsidies granted to rural electric power and telephone customers are further examples of this point. The economic theory of regulation accommodates such diversity. It places regulation in political context and argues that the observed level and pattern of regulatory intervention into the private economy is the logical outcome of a process that tends to favor groups having comparative advantages in exploiting regulatory institutions and processes to their own self-serving ends.

One of the most fruitful applications of the interest-group model recognizes not only that a constellation of interests frequently impinges on the regulatory process, but also that the groups seeking influence are themselves not monolithic. The producers in any industry, for example, differ as to size, cost-efficiency, geographic location, and so on. Heterogeneity on these and other competitive margins gives rise to the possibility that a subset of firms within an industry will be able to utilize regulatory processes to benefit themselves at the expense of their rivals.

To illustrate, consider an industry whose members employ two distinct production technologies. Assume that one technology is relatively labor intensive and that the other is relatively capital intensive. The firms using capital-intensive production methods negotiate a contract with labor union representatives that raises wage rates *industry-wide*. All firms face higher costs as a result, but the costs of the labor-intensive firms rise proportionately more than those of the capital-intensive firms. Marginal producers employing labor-intensive production methods are forced to exit the industry, and if the ensuing increase in market price outweighs the increase in costs for the surviving low-cost producers, their inframarginal rents increase. Moreover, these rents are protected by the fact that the now higher industry wage rates erect a barrier to entry by labor-intensive firms (Williamson, 1968b).

Regulatory processes can be exploited to produce similar intra-industry redistributions of wealth. Consider the Factory Acts adopted by the British government during the 1830s. These laws, which limited the hours women and children could legally work, are widely seen as public-spirited measures designed to end the cruel

exploitation of vulnerable members of the labor force. Howard Marvel (1977), however, argues that a key impetus for passage of the Factory Acts was that they benefited the owners of steam-powered textile mills at the expense of the owners of water-powered mills. The latter could operate only when water flows were adequate to power the textile machinery; production had to be curtailed during times of drought. By preventing the water-powered mills from working overtime when streams were in spate, the Factory Acts conferred a considerable competitive advantage on the owners of the steam-powered mills who were not constrained by river conditions and could therefore operate on a regular basis year round. In addition, Anderson and Tollison (1984) suggest that the interests of senior (male) factory operatives also played a role in the adoption of the Factory Acts insofar as the working-hour restrictions limited the extent to which women and children could compete for their jobs.

The heterogeneous firm approach has likewise been shown to be helpful in explaining the adoption of workplace safety rules (Maloney and McCormick, 1982) and environmental quality regulations (Pashigian, 1985; Bartel and Thomas, 1987). Requiring all producers to employ the same technologies for reducing the risk of on-the-job injuries or for controlling the emission of pollutants can benefit some firms at the expense of others. The actual (as opposed to the stated) purposes of regulation are frequently cloaked in high-minded ideals.

Deregulation

In seeking to explain why a regulatory policy has been adopted, the economic theory of regulation "tells us to look, as precisely and carefully as we can, at who gains and who loses, and how much..." (Stigler, 1975, p. 140). In seeking to explain why a regulatory policy persists, especially in the face of evidence that its actual effects are "unrelated or perversely related" to its announced goals, the interest-group theory tells us that "the *truly intended effects should be deduced from the actual effects*" (ibid.; emphasis in original). Errors are of course possible, but in the Stigler–Peltzman framework,

> errors are not what men live by or on. If an economic policy has been adopted by many communities, or if it is persistently pursued by a society over a long span of time, it is fruitful to assume that the real effects were known *and desired*. Indeed, an explanation of a policy in terms of error or confusion is no explanation at all—anything and everything is compatible with that 'explanation.' (ibid.; emphasis added)

Hence, if it is found, for example, that the regulatory policies of the U.S. Civil Aeronautics Board (CAB) placed the interests of the commercial airlines over those of the flying public, or that the regulatory policies of the ICC placed the interests of the railroads and motor carriers over those of their customers, then the interest-group theory teaches that it is reasonable to conclude that regulation was intended to have precisely those effects.

But what of deregulation? While it is relatively easy, after the fact, to identify the winners and losers from regulation's adoption, how is it possible to explain policies freeing an industry from regulatory control, thereby presumably confiscating its hard-won rents? One theory appeals to regulation's "unintended consequences": designed to ensure that regulated firms would earn profits in excess of normal levels, some regulatory regimes were in practice unable to deliver on that promise. Airline rate regulation under the auspices of the U.S. CAB, for instance, has been described as "sporadic, casual, and uninformed" (Wilcox, 1966, p. 424). The agency's failure in this regard has been attributed to a number of factors, not the least important of which was the problem of determining "the" cost of a seat on a particular flight, given the industry's complex mix of routes, traffic schedules, and capital equipment (Douglas and Miller, 1976). Lacking sufficient flexibility in the fares they were permitted to charge under regulation, the airlines rationally competed for passengers by a variety of nonprice means, including safety records, quality of in-flight meals, comfort of aircraft cabin interiors, and attractiveness of cabin attendants. The scheduling of frequent flights on major routes, offering passengers convenient departure and arrival times, was one of the more important margins of competition. The result was chronic overcapacity: "for all flights by all major airlines in 1977, the composite load factor stood at only 55.5 percent, which meant that on average each plane was flying a little more than half full" (McGraw, 1984, p. 261). Thus, despite regulatory rate fixing, which generated markups ranging from 20% to 95% over the fares charged on unregulated intrastate flights of equal distance (Keeler, 1972), the airlines' profits were eroded by the costs of inefficient scheduling and other forms of nonprice competition. Indeed, the airlines hardly ever earned what the CAB considered to be a "fair" rate of return (Moore, 1986: Douglas and Miller, 1974).

An important barrier to regulatory reform is what Gordon Tullock (1975) calls "the transitional gains trap": the promise of above-normal returns motivates resource owners to seek regulatory privileges from the state. But these gains are only transitory. First, the present value of the available rents is in some cases dissipated upfront in the form of expenditures incurred in the pursuit of monopoly rights. To the extent that these rent-seeking investments are "sunk", deregulation will not necessarily increase society's welfare (McCormick et al., 1984; Shughart, 1999). Second, regulatory rent streams may be eroded *ex post* by nonprice competition among the privileged franchisees. Last, if the monopoly franchise is subsequently sold, the rents will be capitalized in the purchase price of the monopolist's assets. In all of these cases, the rate of return on investments in the regulated industry is driven to normal levels. As a result, there seems to be no politically acceptable way of abolishing a regulatory program that is inefficient both from the standpoint of consumers, who pay artificially high prices, and from the standpoint of producers, who no longer make exceptional profits: "those persons and groups who have established what they consider to be entitlements in the positive gains that have been artificially created will not agree to change, and those persons and groups who suffer losses will not willingly pay off what they consider to be immoral gainers"

(Buchanan, 1980, p. 365). The controversy over compensation for the "stranded costs" of regulated electric utilities—investments made under regulation that are not viable in a competitive market environment—is illustrative (McChesney, 1999).

From a theoretical perspective, the uncomfortable fact of the matter is that the grip of the "dead hand" of monopoly (Buchanan and Tullock, 1968) has been loosened in the airline industry, the trucking industry, and elsewhere. While economists have not yet fully fleshed out a general theory of institutional change, at least some of the episodes in what has thus far been a highly selective deregulation movement seem amenable to explanation by the economic theory of regulation (Keeler, 1984; Peltzman [1989] 1998).

The railroad industry exemplifies a case in which "support for... regulation eroded along with the rent" (Peltzman [1989] 1998:, p. 307). Even though federal regulators kept prices artificially high, a secular decline in demand for rail transport and a regulatory rate structure that accommodated the interests of motor carriers, the ICC's other major constituency, ultimately squeezed railroad profit margins. A spate of bankruptcies in the early 1970s produced a situation in which the only viable political options were nationalization or deregulation. The commercial airline industry, where, as we have seen, profits were dissipated over time by nonprice competition between the major carriers and inefficient capacity utilization, is another instance in which the demand for deregulation seems to have originated from the regulated firms themselves. A demand-side theory of deregulation based on the interests of producers, who expected costs to fall faster than prices, is broadly consistent with the Stigler–Peltzman model. Other examples of deregulation (e.g., stock brokerage, bank deposits, oil) also seem to fit that model, while some (e.g., telecommunications, trucking) do not (Peltzman [1989] 1998).

Supply-side forces may also be at work. Politicians serve as brokers of wealth transfers in a public choice interpretation of the economic theory of regulation (McCormick and Tollison, 1981). If wealth transfers, not social welfare, are all the brokers care about—that is, they are "factionalist reformers" rather than "utilitarian reformers" (Tollison and Wagner, 1991)—then in the face of changes in underlying economic conditions or coalitional strength (producing corresponding changes in relative political prices), they may take advantage of opportunities to advance their own interests by deregulating selected industries, thereby redistributing wealth to newly important constituencies, even though the costs to society of doing so exceed the benefits.

In any case, privatization and deregulation pose major challenges to models in which the privileged holders of monopoly franchises and the other beneficiaries of regulation seem well positioned to resist reform when it is not in their interest. Analytical responses to these challenges merit high priority on the research agendas of political economists. This is especially so given that, paradoxically, the selective retreat of traditional economic regulation of price and entry has been accompanied by spirited growth of regulation in the areas of social and environmental policy (McGraw, 1984, p. 304).

Antitrust

> There is a specter that haunts our antitrust institutions. Its threat is that, far from serving as the bulwark of competition, these institutions will become the most powerful instrument in the hands of those who wish to subvert it. More than that, it threatens to draw great quantities of resources into the struggle to prevent effective competition, thereby more than offsetting the contributions to economic efficiency promised by antitrust activities. This is a specter that may well dwarf any other concern about the antitrust processes. We ignore it at our peril and would do well to take steps to exorcise it. (Baumol and Ordover, 1985, p. 247)

The stated goals of antitrust policy are much the same as those of regulatory policy. It too attempts to influence the pricing and output decisions of private business firms. But enforcement of the antitrust laws proceeds by indirect means rather than by way of the hands-on price and entry controls normally associated with public regulation. Stripped to their essentials, the antitrust laws declare private monopolies to be illegal. Law enforcement is then carried out on a number of fronts, including preventing monopolies from being created in the first place through the merger of former competitors or the orchestration of collusive agreements among them, requiring the dissolution of large firms that have attained monopoly positions in the past, and limiting the use of certain business practices thought to facilitate the acquisition or exercise of market power.

The Legal Framework

American common law in the late nineteenth century "still contained provisions that had been struck from the English common law by statutes", including proscriptions on forestalling and engrossing as well as prohibitions on combinations of workers in restraint of trade (Letwin, 1965: 52). U.S. antitrust policy's legislative history dates to state statutes, many of which were enacted in the 1880s (Libecap, 1992; Boudreaux et al., 1995). It began at the federal level with passage of the Sherman Act (1890), section 1 of which states that "every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, is declared to be illegal". The law's only other substantive section (section 2) declares that "every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony".

Some commentators have argued that the Sherman Act merely codified the common law treatment of restraints of trade (Demsetz, 1992; Kleit, 1993), layering on an apparatus of public enforcement (by the U.S. Department of Justice) and allowing certain mergers to be deemed unlawful, neither of which innovations produced significant changes in American competition policy. It is clear, however, that at least some freely entered into private contracts were newly brought within

statutory reach. Before the Sherman Act, price-fixing agreements were not presumptively illegal. Indeed, "the common law was inclined to uphold contracts in restraint of trade for the same reasons that moved it to sustain any good contract" (Letwin, 1965, p. 42). As was the case with the futures contracts banned in a Dutch edict of 1610, which proscribed "windhandel" or trading in shares not currently in the seller's possession, the courts did not impose sanctions on firms for participating in collusive agreements; "they simply refused legal enforcement of such contracts" (Garber, 2000, p. 36). In particular, "the modern common law on combinations in restraint of trade was established by the *Mogul Steamship* case [2 Chitty 407 (1815)], which laid down the principle that although a trade combination might be destroyed by attack from within, it could not be successfully attacked by an outsider" (Letwin, 1965, p. 49). Thus, "the Sherman Act went far beyond the common law when it authorized the Attorney General to indict violators of the Act, and gave injured persons the power to sue them" (ibid., p. 52).

The Sherman Act was innovative for a second reason. "Unlike statute law, common law allows people to contract around it" (De Alessi, 2001, p. 39). In other words, "all those parties who do not wish to be bound by a particular [common law] rule, . . . generally have the opportunity to adopt any other rule that is mutually satisfactory" (De Alessi and Staaf, 1991, p. 112). This was no longer possible after 1890. Controlled by the statute's language, individuals and firms were no longer free to enter into contracts that would restrain trade, even if such contracts made them jointly better off.

Be that as it may, desultory enforcement of the Sherman Act early on, combined with negative reactions from antitrust's partisans to early interpretations of it, produced a demand for new legislation that would define more sharply the boundaries of U.S. antitrust policy. Supporters of a vigorous antitrust policy were especially critical of the 1911 landmark decision ordering the dissolution of the Standard Oil trust, in which the Court announced a "rule of reason", declaring its unwillingness to condemn all restraints of trade, but only those determined to be "unreasonable". As a compromise between those pressing for a law that would incorporate a list of specific business practices to be declared unlawful (and made criminal offenses) and those pressing for a law that would provide broad, but unspecified enforcement powers, two additional antitrust statutes were enacted in 1914. One of these was the Clayton Act, which identified and declared illegal four specific business practices—price discrimination (section 2), exclusive dealing and tying contracts (section 3), mergers (section 7) and interlocking corporate directorates (section 8)—where their effect "may be to substantially lessen competition or tend to create a monopoly". The other was the Federal Trade Commission (FTC) Act, which created a five-member law enforcement body and delegated to it the responsibility for prosecuting unspecified "unfair methods of competition" (FTC Act §5).

Subsequent amendments to these two statutes strengthened and broadened the scope of the powers granted to the federal antitrust authorities. The most important of these were the Robinson–Patman Act (1936), which made it more difficult to mount defenses against charges of unlawful price discrimination; the Wheeler–Lea

Act (1938), which added the phrase "unfair or deceptive acts or practices in or affecting commerce" to section 5 of the FTC Act, thereby granting the commission authority to regulate advertising and other business activities, such as product warranties and credit terms, falling under the rubric of "consumer protection"; the Celler–Kefauver Act (1950), which closed a 'loophole' in section 7 of the Clayton Act allowing mergers consummated through the acquisition of stock to escape condemnation (but see Ekelund et al., 1995); and the Hart–Scott–Rodino Antitrust Improvement Act (1976), which established a formal premerger notification and review process.

Statutory antitrust policy is of much more recent vintage in Europe. Six pieces of legislation delineate its contours in the United Kingdom: the Monopolies and Restrictive Trade Practices Act (1948), the Fair Trading Act (1973), the Restrictive Trade Practices Act (1976), the Resale Prices Act (1976), and the Competition Acts of 1980 and 1998. The first of these laws established the Monopolies and Restrictive Practices Commission, a tribunal having the authority to investigate cases referred to it by the Board of Trade (see Rowley, 1966, for an analysis of the repercussions of the law). The second established the Office of Fair Trading, delegating to it responsibility for monitoring competition and granting it authority to refer to a lay body, the Monopolies and Mergers Commission (MMC), the power to investigate suspected "monopoly situations" (defined as a single firm or group of firms accounting for 25% of sales or purchases in the relevant market). Public utilities and "anti-competitive practices" were added to the MMC's charge by the Competition Act of 1980. The two 1976 statutes deal with price-fixing agreements and with vertical price restraints (e.g., resale price maintenance), respectively (Hay and Morris, 1991, pp 612–614); the Competition Act of 1998 aligns British law more closely with its European counterpart (Utton, 2000).

Competition policy in the European Union emanates from the Treaty of Rome (1957). The first of the Treaty's two substantive antitrust provisions prohibits agreements and other concerted actions, be they along horizontal or vertical lines or involve price or nonprice terms, which restrict competition within or among the member states. The second provision condemns abuses of dominant market positions, including "imposing unfair purchasing or selling prices or other unfair trading conditions", "limiting production, markets or technical development to the prejudice of consumers", "applying dissimilar conditions to equivalent transactions with other trading parties", and "making the conclusion of contracts subject to acceptance by the other parties of conditions which... have no connection with the subject of such contracts" (ibid., p. 617). Although not based on any clearly articulated theory of anticompetitive behavior, possible "abuses of dominant market positions" have been the chief concern of the EU's law enforcers in recent years. Reflecting the emerging globalization of antitrust, worries of incipient market dominance have provoked decisive European opposition to a number of high-profile mergers between major U.S. companies to which U.S. authorities had previously granted clearance.

The Process of Antitrust Law Enforcement

Despite differences in details, the stated purposes of competition policy in the United States and elsewhere in the western industrialized world rest squarely on the market-failure tradition. Antitrust's staunchest advocates see the laws as embodying values consistent with economic efficiency goals, ensuring that markets remain vigorously competitive and that consumers are thereby protected against the abuses of market power: "a much more widespread pattern of growth by merger, an efflorescence of collusive agreements of all sorts, and the use of various exclusionary and otherwise anticompetitive practices now forbidden would all follow on the abandonment of a procompetitive public policy" (Kaysen and Turner, 1959, p. 5). Indeed, at least one respected student of the legislative history of the Sherman Act has argued forcefully that antitrust's origins were explicitly based on a consumer-welfare standard (Bork, 1966, 1978).

The professed efficiency basis of competition policy has not gone unchallenged (Lande, 1982; DiLorenzo and High, 1988). What is more important, faith in the efficacy of the antitrust laws to deliver net social gains ignores the political pressures that impinge on the agencies created to enforce them, pressures marshaled by groups perceiving opportunities to exploit antitrust processes strategically, not to promote competition, but to subvert it (Baumol and Ordover, 1985). A law that declares mergers to be illegal where their effect "may be to substantially lessen competition or tend to create a monopoly" is also a law that affords the merger partners' rivals the opportunity to block a transaction that promises to create a larger, more efficient competitor. A law that makes it illegal for a firm to charge different prices to different customers not justified by differences in the cost of serving them is also a law that affords rivals the opportunity to seek relief from prices that are "predatorily" low. The Robinson–Patman Act was in fact drafted and passed in response to the political influence mobilized by independent grocers, druggists and other small retailers, who complained loudly that, under the Clayton Act's original language, the FTC was either unable or unwilling to prevent the emerging national chain stores from using their mass buying power to sell goods to consumers at prices below those charged by the independents (Ross, 1984).

Observers of the antitrust enforcement process have long been critical of individual applications of it (for recent surveys of the case-study literature, see Armentano, 1990; Rubin, 1995). A typical antitrust case study finds that the evidence presented in behalf of the plaintiff was "weak and at times bordered on fiction" and that "neither the government nor the Courts seemed able to distinguish between competition and monopolizing" (Peterman, 1975, p. 143). Even when the law conceivably has struck at acts and practices that resulted in injury to consumers, the effectiveness of the penalties imposed on guilty defendants has been called into question (Elzinga, 1969; Rogowsky, 1986, 1987). Systematic empirical studies of the antitrust case-selection process have produced *no* support for the hypothesis that the process is guided by social-welfare criteria (Long et al., 1973; Asch, 1975; Siegfried, 1975) or that antitrust law enforcement has had measurable pro-competitive effects on the

behavior of firms (Stigler, 1966; Asch and Seneca, 1976; Shaw and Simpson, 1986; Sproul, 1993).

Viewed through the lens of public choice, these apparent empirical anomalies are easily explained: social-welfare criteria carry little or no weight in the objective functions of the politicians and policymakers charged with drafting and enforcing the antitrust laws. Writing in 1969, Richard Posner charged that the FTC's stated mission of promoting competitive markets had been significantly impaired by reason of its dependence on Congress, which must approve budget requests and confirm presidential appointments to senior policymaking positions. He emphasized the obvious point that in a geographically based representative democracy, each member of the legislature is obligated to protect and further the provincial interests of those who have elected him to office. More specifically, "the welfare of his constituents may depend disproportionately on a few key industries. The promotion of the industries becomes one of his most important duties as a representative of the district" (Posner, 1969a, p. 83). The ability to do so would accrue disproportionately to the members of the committees and subcommittees of Congress vested with oversight responsibilities with respect to antitrust law enforcement generally and the FTC in particular: a legislator holding such a position will have "a great deal of power to advance the interests of businesses located in his district however unimportant the interests may be from a national standpoint" (ibid.). A subsequent test of this antitrust "pork barrel" hypothesis found that cases instituted against firms headquartered in the jurisdictions of key committee members were more likely to be dismissed than cases instituted against firms not so represented (Faith et al., 1982).

Merger law enforcement seems to be particularly vulnerable to political influence. Two studies have found that, holding its staff's evaluation of the merits of a proposed merger constant, the commission is more likely to vote to oppose a transaction the more pressure is brought to bear on it in the form of news coverage and summonses to appear before congressional committees (Coate et al., 1990; Coate and McChesney, 1992). Similarly, the only two factors found to increase the probability of a merger challenge by the United Kingdom's Monopolies and Mergers Commission were whether the proposed merger would affect the balance of payments adversely and whether the firm targeted for takeover contested the bid (Weir, 1992). The available evidence from capital market event studies suggests that the mergers challenged by the U.S. antitrust authorities tend to be efficiency enhancing on balance, and that the merger partners' rivals therefore appear to be the chief beneficiaries of merger law enforcement (e.g., Eckbo and Wier, 1985). This evidence offers further support for the contention that, because many investigations of alleged violations of the law are initiated "at the behest of corporations, trade associations, and trade unions whose motivation is at best to shift the costs of private litigation to the taxpayer and at worst to harass competitors", antitrust seldom serves the public's interest (Posner, 1969a, p. 87).

In sum, the empirical case for characterizing antitrust processes as a mechanism of wealth redistribution is strong. From the perspective of public choice, antitrust is simply another form of regulation, having the same causes and consequences.

Although this conclusion has not yet gained wide acceptance, the mounting evidence of the politicization of antitrust law enforcement produced by recent high-profile cases brought against some of the world's most successful business enterprises—cases instigated not in response to complaints by consumers but at the prompting of competitors and other special pleaders—promises eventually to bring antitrust within the ambit of the economic theory of regulation.

Summary

The economic theory of regulation generally, and antitrust in particular, looks behind the stated intentions of the proponents of government intervention into the private economy to uncover hidden agendas of wealth redistribution. The theory's main thrust is that the formulation and enforcement of public policies toward business has, in fact, tended to protect politically powerful constituencies at the expense of competition and economic efficiency. That is, the theory explains many (if not all) policy decisions as rational political responses to the demands of well-organized pressure groups. These demanders of protectionism offer political support (votes, campaign contributions, and the like) in return for favored treatment. These favors include the right to charge prices in excess of costs, the erection of barriers to the entry of new rivals, and the proscription of business practices and contractual agreements that would enhance overall economic efficiency, but harm them personally. Importantly, the strategic exploitation of regulation and antitrust by well-organized groups does not represent "abuse" of the policy process in any meaningful sense. The demand for protectionism—and the political response to it—is simply rational behavior under a particular set of institutional constraints.

Competing with this general public-choice description of the purposes and effects of government policies toward business is the public-interest "theory", which contends that regulatory and antitrust policies are the product of well-intentioned, but fallible, public servants. Whether justified or not in the economic theories or situational facts they rely on in any particular case, the function of the relevant laws, regulations, and enforcement agencies is to serve what are believed to be the best interests of society as a whole. While mistakes are certainly possible, public policies toward business, or so it is thought, are designed and generally work to improve the allocation of scarce productive resources.

Strongly held a priori beliefs in the efficacy of governmental processes are the principal sources of support for the public-interest theory. The empirical evidence is almost universally consistent with the predictions of the interest-group, public-choice theory. As a result, the benefits of regulation are now seen to accrue chiefly, not to the public at large, but to politically well-organized pressure groups. While antitrust policy has only recently been exposed to the analytical power of the public-choice model, the idea that it, uniquely among public policies toward business, is immune to political influence, is now in significant retreat.

References

Acton, J.P. and Vogelsang, I. (1989). "Price-cap regulation: introduction." *RAND Journal of Economics*, 20, 369–372.

Alchian, A.A. (1965). "Some economics of property rights." *Il Politico*, 30(4), 816–829.

Anderson, G.M., Ekelund, R.B. Jr, and Tollison, R.D. (1989). "Nassau senior as economic consultant: the factory acts reconsidered." *Economica*, 56, 71–92.

Anderson, G.M. and Tollison, R.D. (1984). "A rent-seeking explanation of the British factory acts," in D.C. Collander (ed.) *Neoclassical Political Economy: The Analysis of Rent- Seeking and DUP Activities*. Cambridge, MA: Ballinger, 187–201.

Aranson, P.H. (1990). "Theories of economic regulation: from clarity to confusion." *Journal of Law and Politics*, 6, 247–286.

Armentano, D.T. (1990). *Antitrust and Monopoly: Anatomy of a Policy Failure*, Second Edition. Oakland, CA: Independent Institute.

Asch, P. (1975). "The determinants and effects of antitrust policy." *Journal of Law and Economics*, 18, 575–581.

Asch, P. and Seneca, J.J. (1976). "Is collusion profitable?" *Review of Economics and Statistics*, 58, 1–12.

Averch, H. and Johnson, L.L. (1962). "Behavior of the firm under regulatory constraint." *American Economic Review*, 52, 1052–1069.

Bartel, A.P. and Thomas, L.G. (1987). "Predation through regulation: the wage and profit effects of the occupational safety and health administration and the environmental protection agency." *Journal of Law and Economics*, 20, 239–264.

Bator, F.M. (1958). "The anatomy of market failure." *Quarterly Journal of Economics*, 72, 351–379.

Baumol, W.J., Bailey, E., and Willig, R. (1977). "Weak invisible hand theorems on the sustainability of prices in a multiproduct monopoly." *American Economic Review*, 67, 350–365.

Baumol, W.J. and Klevorick, A.K. (1970). "Input choices and rate-of-return regulation: an overview of the discussion." *Bell Journal of Economics and Management Science*, 1, 162–190.

Baumol, W.J., Panzar, J.C., and Willig, R.D. (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich.

Baumol, W.J. and Ordover, J.A. (1985). "Use of antitrust to subvert competition." *Journal of Law and Economics*, 28, 247–265.

Baxter, W.F. (1980). "The political economy of antitrust," in R.D. Tollison (ed.) *The Political Economy of Antitrust: Principal Paper by William Baxter*. Lexington, MA: Lexington Books, 3–49.

Becker, G.S. (1983). "A theory of competition among pressure groups for political influence." *Quarterly Journal of Economics*, 98, 371–400.

Becker, G.S. (1985). "Public policies, pressure groups, and dead weight costs." *Journal of Public Economics*, 28, 329–347.

Bertoletti, P. and Poletti, C. (1997). "X-inefficiency, competition and market information." *Journal of Industrial Economics*, 45, 359–375.

Bork, R.H. (1966). "Legislative intent and the policy of the Sherman act." *Journal of Law and Economics*, 9, 7–48.

Bork, R.H. (1978). The Antitrust Paradox: A Policy at War with Itself. New York: Basic Books.

Boudreaux, D.J., DiLorenzo, T.J., and Parker, S. (1995). "Antitrust before the Sherman act," in F.S. McChesney and W.F. Shughart II (eds.) *The Causes and Consequences of Antitrust: The Public-Choice Perspective*. Chicago: University of Chicago Press, 255–270.

Buchanan, J.M. (1972). "Toward analysis of closed behavioral systems," in J.M. Buchanan and R.D. Tollison (eds.) *Theory of Public Choice*. Ann Arbor: University of Michigan Press, 11–23.

Buchanan, J.M. (1980). "Reform in the rent-seeking society," in J.M. Buchanan, R.D. Tollison, and G. Tullock (eds.) *Toward a Theory of The Rent-Seeking Society*. College Station: Texas A&M University Press, 359–367.

Buchanan, J.M. and Tullock, G. (1962). *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Buchanan, J.M. and Tullock, G. (1968). "The 'dead hand' of monopoly." *Antitrust Law and Economic Review*, 85–96.

Cawley, J. and Philipson, T. (1999). "An empirical examination of barriers to trade in insurance." *American Economic Review*, 89, 827–846.

Cheung, S.N.S. (1973). "The fable of the bees: an economic interpretation." *Journal of Law and Economics*, 16, 11–33.

Coase, R.H. (1946). "The marginal cost controversy." *Economica*, 13, 169–182.

Coase, R.H. (1959). "The federal communications commission." *Journal of Law and Economics*, 2, 1–40.

Coase, R.H. (1960). "The problem of social cost." *Journal of Law and Economics*, 3, 1–44.

Coase, R.H. (1974). "The lighthouse in economics." *Journal of Law and Economics*, 17, 357–376.

Coate, M.B., Higgins, R.S., and McChesney, F.S. (1990). "Bureaucracy and politics in FTC merger challenges." *Journal of Law and Economics*, 33, 463–482.

Coate, M.B. and McChesney, F.S. (1992). "Empirical evidence on FTC enforcement of the merger guidelines." *Economic Inquiry*, 30, 277–293.

Crain, W.M. and McCormick, R.E. (1984). "Regulators as an interest group," in J.M. Buchanan and R.D. Tollison (eds.) *The Theory of Public Choice-II*. Ann Arbor: University of Michigan Press, 287–304.

De Alessi, L. (1982). "On the nature and consequences of private and public enterprise." *Minnesota Law Review*, 67, 191–209.

De Alessi, L. (1983). "Property rights, transaction costs, and X-efficiency: an essay in economic theory." *American Economic Review*, 73, 64–81.

De Alessi, L. (2001). "Property rights: private and political institutions," in W.F. Shughart II and L. Razzolini (eds.) *The Elgar Companion to Public Choice*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar, 33–58.

De Alessi, L. and Staaf, R.J. (1991). "The common law process: efficiency or order?" *Constitutional Political Economy*, 2, 107–126.

Demsetz, H. (1968). "Why regulate utilities?" *Journal of Law and Economics*, 11, 55–65.

Demsetz, H. (1969). "Information and efficiency: another viewpoint." *Journal of Law and Economics*, 12, 1–22.

Demsetz, H. (1970). "The private production of public goods." *Journal of Law and Economics*, 13, 293–306.

Demsetz, H. (1992). "How many cheers for antitrust's 100 years?" *Economic Inquiry*, 30, 207–217.

DiLorenzo, T.J. (1985). "The origins of antitrust: an interest-group perspective." *International Review of Law and Economics*, 5, 73–90.

DiLorenzo, T.J. and High, J.C. (1988). "Antitrust and competition, historically considered." *Economic Inquiry*, 26, 423–435.

Douglas, G.W. and Miller, J.C. III (1974). "Quality competition, industry equilibrium, and efficiency in the price-constrained airline market." *American Economic Review*, 64, 657–669.

Douglas, G.W. and Miller, J.C. III (1976). *Economic Regulation of Domestic Air Transport: Theory and Policy*. Washington, DC: Brookings Institution.

Easterbrook, F.H. (1984). "The limits of antitrust." *University of Texas Law Review*, 63, 1–40.

Eckbo, B.E. and Wier, P. (1985). "Antimerger policy under the Hart-Scott-Rodino act: a reexamination of the market-power hypothesis." *Journal of Law and Economics*, 28, 119–149.

Ekelund Jr, R.B., McDonald, M.J., and Tollison, R.D. (1995). "Business restraints and the Clayton Act of 1914: public- or private-interest legislation?" in F.S. McChesney and W.F. Shughart II (eds.) *The Causes and Consequences of Antitrust: The Public-Choice Perspective*. Chicago: University of Chicago Press, 271–286.

Elzinga, K.G. (1969). "The antimerger law: Pyrrhic victories?" *Journal of Law and Economics*, 12, 43–78.

Faith, R.L., Leavens, D.R. and Tollison, R.D. (1982). "Antitrust pork barrel." *Journal of Law and Economics*, 25, 329–342.

Garber, P.M. (2000). *Famous First Bubbles: The Fundamentals of Early Manias*. Cambridge and London: MIT Press.

Harberger, A.C. (1954). "Monopoly and resource allocation." *American Economic Review Papers and Proceedings*, 44, 77–87.

Hay, D.A. and Morris, D.J. (1991). *Industrial Economics and Organization: Theory and Evidence*, Second Edition. Oxford: Oxford University Press.

Hazlett, T.W. (1984). "Interview with George Stigler." *Reason*, January: 44–48.

Henderson, D.R. (1995). "Antitrust busters." *Reason*, August/September: 62–64.

Hotelling, H. (1938). "The general welfare in relation to problems of taxation and railway and utility rates." *Econometrica*, 6, 242–269.

Jensen, M.C. and Meckling, W.H. (1976). "Theory of the firm: managerial behavior, agency costs and ownership structure." *Journal of Financial Economics*, 3, 305–360.

Jordan, W.A. (1972). "Producer protection, prior market structure and the effects of government regulation." *Journal of Law and Economics*, 15, 151–176.

Joskow, P. and Noll, R. (1981). "Regulation in theory and practice: an overview," in G. Fromm (ed.) *Studies in Public Regulation*. Cambridge: MIT Press, 1–65.

Kaysen, C. and Turner, D.F. (1959). *Antitrust Policy*. Cambridge: Harvard University Press.

Keeler, T.E. (1972). "Airline regulation and market performance." *Bell Journal of Economics and Management Science*, 3, 399–424.

Keeler, T.E. (1984). "Theories of regulation and the deregulation movement." *Public Choice*, 44(1), 103–145.

Kleit, A.N. (1993). "Common law, statute law, and the theory of legislative choice: an inquiry into the goal of the Sherman Act." *Economic Inquiry*, 31, 647–662.

Krueger, A.O. (1974). "The political economy of the rent-seeking society." *American Economic Review*, 64, 291–303.

Lande, R.H. (1982). "Wealth transfers as the original and primary concern of antitrust: the efficiency interpretation challenged." *Hastings Law Journal*, 34, 65–151.

Letwin, W. (1965). Law and Economic Policy in America: The Evolution of the Sherman Antitrust Act. Chicago and London: University of Chicago Press.

Libecap, G.D. (1992). "The rise of the Chicago packers and the origins of meat inspection and antitrust." *Economic Inquiry*, 30, 242–262.

Leibenstein, H. (1966). "Allocative efficiency vs. 'x-efficiency'." *American Economic Review*, 56, 392–415.

Leibenstein, H. (1978). "X-inefficiency xists—reply to an xorcist." *American Economic Review*, 68, 203–211.

Long, W.F., Schramm, R., and Tollison, R.D. (1973). "The determinants of antitrust activity." *Journal of Law and Economics*, 16, 351–364.

Lott, J.R. Jr (1997). "Donald Wittman's *The myth of democratic failure*: review article." *Public Choice*, 92, 1–13.

Maloney, M.T. and McCormick, R.E. (1982). "A positive theory of environmental quality regulation." *Journal of Law and Economics*, 25, 99–124.

Marvel, H.P. (1977). "Factory regulation: a reinterpretation of early English experience." *Journal of Law and Economics*, 20, 379–402.

McChesney, F.S. (1987). "Rent extraction and rent creation in the economic theory of regulation." *Journal of Legal Studies*, 16, 101–118.

McChesney, F.S. (1991). "Rent extraction and interest-group organization in a Coasean model of regulation." *Journal of Legal Studies*, 20, 73–90.

McChesney, F.S. (1997). Money for Nothing: Politicians, Rent Extraction, and Political Extortion. Cambridge: Harvard University Press.

McChesney, F.S. (1999). "Of stranded costs and stranded hopes: the difficulties of deregulation." *Independent Review*, 3, 485–509.

McChesney, F.S. and Shughart, W.F. II (eds.) (1995). *The Causes and Consequences of Antitrust: The Public-Choice Perspective*. Chicago: University of Chicago Press.

McCormick, R.E. (1984). "The strategic use of regulation: a review of the literature," in R.A. Rogowsky and B. Yandle (eds.) *The Political Economy of Regulation: Private Interests in the Regulatory Process*. Washington, DC: Federal Trade Commission, 13–32.

McCormick, R.E., Shughart, W.F. II, and Tollison, R.D. (1984). "The disinterest in deregulation." *American Economic Review*, 74, 1075–1079.

McCormick, R.E. and Tollison, R.D. (1981). Politicians, Legislation, and the Economy: An Inquiry into the Interest-Group Theory of Government. Boston: Martinus Nijhoff.

McGraw, T.K. (1984). Prophets of regulation: Charles Francis Adams, Louis D. Brandeis, James M. Landis, Alfred E. Kahn. Cambridge and London: Belknap Press of Harvard University Press.

Moore, T.G. (1986). "U.S. airline deregulation: its effects on passengers, capital, and labor." *Journal of Law and Economics*, 29, 1–28.

Nelson, P. (1970). "Information and consumer behavior." *Journal of Political Economy*, 78, 311–329.

Olson, M. (1965). The Logic of Collective Action: Public Goods and the Theory of Groups. Cambridge: Harvard University Press.

Pashigian, B.P. (1985). "Environmental quality regulation: whose self-interests are being protected?" *Economic Inquiry*, 23, 551–584.

Peltzman, S. (1976). "Toward a more general theory of regulation." *Journal of Law and Economics*, 19, 211–240.

Peltzman, S. (1993). "George Stigler's contribution to the economic analysis of regulation." *Journal of Political Economy*, 101, 818–832.

Peltzman, S. (1989, 1998). "The economic theory of regulation after a decade of deregulation," in S. Peltzman (eds.) *Political Participation and Government Regulation*. Chicago: University of Chicago Press, 286–323.

Peterman, J.A. (1975). "The Brown Shoe case." *Journal of Law and Economics*, 18, 81–146.

Pigou, A.C. (1932). *The Economics of Welfare*, Fourth Edition. London: Macmillan.

Posner, R.A. (1969a). "The federal trade commission." *University of Chicago Law Review*, 37, 47–89.

Posner, R.A. (1969b, 1999). *Natural Monopoly and its Regulation*. Washington, DC: Cato Institute.

Posner, R.A. (1971). "Taxation by regulation." *Bell Journal of Economics and Management Science*, 2, 22–50.

Posner, R.A. (1974). "Theories of economic regulation." *Bell Journal of Economics and Management Science*, 5, 335–358.

Posner, R.A. (1975). "The social costs of monopoly and regulation." *Journal of Political Economy*, 83, 807–827.

Ramsey, F.P. (1927). "A contribution to the theory of taxation." *Economic Journal*, 37, 47–61.

Rogowsky, R.A. (1986). "The economic effectiveness of section 7 relief." *Antitrust Bulletin*, 31, 187–233.

Rogowsky, R.A. (1987). "The Pyrrhic victories of section 7: a political economy approach," in R.J. Mackay, J.C. Miller III, and B. Yandle (eds.) *Public Choice and Regulation: A View From Inside the Federal Trade Commission*. Stanford, CA: Hoover Institution Press, 220–239.

Ross, T.W. (1984). "Winners and losers under the Robinson-Patman Act." *Journal of Law and Economics*, 27, 243–271.

Rowley, C.K. (1966). *The British Monopolies Commission*. London: George Allen and Unwin.

Rowley, C.K. (1997). "Donald Wittman's *The myth of democratic failure*: review article." *Public Choice*, 92, 15–26.

Rowley, C.K. and Peacock, A.T. (1975). *Welfare Economics: A Liberal Restatement*. Oxford: Martin Robertson.

Rowley, C.K. and Yarrow, G.K. (1981). "Property rights, regulation and public enterprise: the case of the British steel industry 1957–1975." *International Review of Law and Economics*, 1, 63–96.

Rubin, P.H. (1995). "What do economists think about antitrust? A random walk down Pennsylvania avenue," in F.S. McChesney and W.F. Shughart II (eds.) *The Causes and Consequences of Antitrust: The Public-Choice Perspective*. Chicago: University of Chicago Press, 33–61.

Samuelson, P.A. (1954). "The pure theory of public expenditure." *Review of Economics and Statistics*, 36, 387–389.

Scherer, F.M. (1980). *Industrial market structure and economic performance*, Second Edition. Boston: Houghton Mifflin.

Schmidt, K.M. (1997). "Managerial incentives and product market competition." *Review of Economic Studies*, 64, 191–213.

Shaw, R.W. and Simpson, P. (1986). "The persistence of monopoly: an investigation of the effectiveness of the United Kingdom monopolies commission." *Journal of Industrial Economics*, 34, 355–372.

Sherman, R. (1989). *The Regulation of Monopoly*. Cambridge and New York: Cambridge University Press.

Shughart, W.F. II (1990). *Antitrust Policy and Interest-Group Politics*. New York: Quorum Books.

Shughart, W.F. II (1997). *The Organization of Industry*, Second Edition. Houston, TXDame Publications.

Shughart, W.F. II (1999). "The reformer's dilemma." *Public Finance Review*, 27, 561–565.

Shughart, W.F. II (2004). "Bending before the storm: the US Supreme Court in economic crisis, 1934–1937." *Independent Review*, 9, 55–83.

Shughart, W.F. II and Tollison, R.D. (1985). "The positive economics of antitrust policy: a survey article." *International Review of Law and Economics*, 5, 39–57.

Shy, O. (2001). *The Economics of Network Industries*. Cambridge and New York: Cambridge University Press.

Siegfried, J.J. (1975). "The determinants of antitrust activity." *Journal of Law and Economics*, 17, 559–574.

Simon, H.A. (1957). *Models of Man*. New York: Wiley.

Sproul, M.T. (1993). "Antitrust and prices." *Journal of Political Economy*, 101, 741–754.

Spulber, D.F. (1989). *Regulation and Markets*. Cambridge and London: MIT Press.

Stigler, G.J. (1964). "Public regulation of the securities market." *Journal of Business*, 2, 117–142.

Stigler, G.J. (1966). "The economic effects of the antitrust laws." *Journal of Law and Economics*, 9, 225–258.

Stigler, G.J. (1971). "The economic theory of regulation." *Bell Journal of Economics and Management Science*, 2, 3–21.

Stigler, G.J. (1975). "Supplementary note on economic theories of regulation (1975)," in G.J. Stigler (ed.) *The Citizen and The State: Essays on Regulation*. Chicago: University of Chicago Press, 137–141.

Stigler, G.J. (1976). "The xistence of x-efficiency." *American Economic Review*, 86, 213–216.

Stigler, G.J. (1985). "The origin of the Sherman Act." *Journal of Legal Studies*, 14, 1–12.

Stigler, G.J. (ed.) (1988). *Chicago Studies in Political Economy*. Chicago: University of Chicago Press.

Stigler, G.J. and Friedland, C. (1962). "What can regulators regulate? The case of electricity." *Journal of Law and Economics*, 5, 1–16.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge and London: MIT Press.

Tollison, R.D. and Wagner, R.E. (1991). "Romance, realism, and economic reform." *Kyklos*, 44 (1), 57–70.

Train, K.E. (1991). Optimal Regulation: The Economic Theory of Natural Monopoly. Cambridge: MIT Press.

Troesken, W. (2000). "Did the trusts want a federal antitrust law? An event study of state antitrust enforcement and passage of the Sherman Act," in J.C. Heckelman, J.C. Moorhouse, and R.M. Whaples (eds.) *Public Choice Interpretations of American Economic History*. Boston: Kluwer Academic Publishers, 77–104.

Tullock, G. (1967). "The welfare costs of tariffs, monopolies and theft." *Western Economic Journal*, 5, 224–232.

Tullock, G. (1975). "The transitional gains trap." *Bell Journal of Economics and Management Science*, 6, 671–678.

Tullock, G. (1980). "Efficient rent seeking," in J.M. Buchanan, R.D. Tollison, and G. Tullock (eds.) *Toward a Theory of The Rent-Seeking Society*. College Station: Texas A&M University Press, 97–112.

Utton, M. (2000). "Going European: Britain's new competition law." *Antitrust Bulletin*, 45, 531–551.

Weir, C. (1992). "Monopolies and mergers commission, merger reports and the public interest: a probit analysis." *Applied Economics*, 24, 27–34.

Wilcox, C. (1966). *Public Policies Toward Business*, Third Edition. Homewood, IL: Irwin.

Williamson, O.E. (1968a). "Economies as an antitrust defense: the welfare tradeoffs." *American Economic Review*, 58, 18–36.

Williamson, O.E. (1968b). "Wage rates as a barrier to entry: the Pennington case in perspective." *Quarterly Journal of Economics*, 82, 85–116.

Williamson, O.E. (1976). "Franchise bidding for natural monopolies—in general and with respect to CATV." *Bell Journal of Economics*, 7, 73–104.

Williamson, O.E. (1977). "Economies as an antitrust defense revisited." *University of Pennsylvania Law Review*, 125, 699–736.

Wittman, D. (1989). "Why democracies produce efficient results." *Journal of Political Economy*, 97, 1395–1424.

Wittman, D. (1995). The Myth of Democratic Failure: Why Political Institutions are Efficient. Chicago: University of Chicago Press.

Chapter 26
International Trade Policy: Departure From Free Trade[1]

Arye Hillman

In a world with no international boundaries and no sovereign governments, all trade would be domestic and there could be no international trade policy. Governments and national sovereignty introduce *international* trade, but the gains from free trade (Kemp 1962; Samuelson 1962) remain unaffected. Yet governments have often chosen to depart from free trade. Economic research has taken two approaches to the departures from free trade. A conventional view in the literature of international economics has been normative in developing a research agenda that shows how departure from free trade can enhance efficiency and maximize social welfare. A political-economy view synonymous with public (or rational) choice has approached departure from free trade from a positive perspective (explaining and predicting rather than recommending), and has shown why trade policy might *compromise* the efficiency of free trade for political and income-distribution reasons. The conventional normative views have origins in classical nineteenth century justifications put forward as exceptions to the case for the efficiency of free trade.

The Terms of Trade

A classical nineteenth century argument recognized that departure from free trade may increase the welfare of a population by improving the terms of trade. Gain through the terms of trade requires a population to collectively have monopsony power in the world market for imported goods. The usual outcome of a tariff (income effects can result in unusual outcomes) is an increased domestic (relative) price of imports and reduced domestic demand, and the terms of trade improve if the reduced domestic demand decreases world demand so that the relative price of

---

[1] This chapter is a reprint of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 129–138.

imported goods falls in world markets. Cheaper imports are a source of social benefit. There are accompanying losses because of declines in the amount of trade and domestic inefficiency because of the tariff. An *optimum tariff* balances these losses against the gains from improvement in the terms of trade.

Since the gain to a population through an optimum tariff is at the expense of the people in other countries whose terms of trade have deteriorated, the optimum tariff is known as a beggar-thy-neighbor policy. World efficiency is also compromised for the benefit of a local population. Populations in countries that do not seek gain at the expense of others will not wish to have their governments impose optimum tariffs. There may in any event be no prospect of gain through an optimum tariff, since there may be no goods for which a country's population has a sufficiently large share of world consumption for a collective monopsony power to be present. Whenever populations face given market-determined world prices, there is no collective monopsony power and the optimum tariff is zero.

There are problems other than willingness to take advantage and feasibility in seeking gains through optimum tariffs. Where feasible, optimum tariffs may result in foreign retaliatory tariffs that reverse the beneficial terms of trade changes while further reducing the volume of trade (although terms of trade gains from an optimum tariff may be sustainable despite retaliation (Johnson 1953–54)). Benefits to a population also require that revenue from the tariff be used to finance increased public spending or to reduce other taxes. For example, the government of a country whose population has a collective monopsony power might decide to use a tariff on coffee to reduce the world price of coffee and so improve the country's terms of trade as an importer of coffee. Domestic consumers of coffee lose when the domestic price of coffee increases. The offsetting gain to consumers is through the tariff revenue that the government has collected. There is however considerable evidence of wasteful government spending (Tanzi and Schuknecht 2000). A government that does not spend the revenue in a socially beneficial way, or does not reduce other taxes, fails to deliver the offsetting gain. The legacy of the tariff for domestic consumers is the higher domestic price of coffee. The country's tariff will have provided benefits to coffee consumers in other countries through the reduced world price of coffee.

Market power in world markets has generally been exercised through monopoly (e.g., the OPEC oil cartel) rather than through monopsony. Documented cases of optimum tariffs improving the terms of trade are uncommon in empirical literature. Also uncommon is the documentation of governments declaring that the purpose of a tariff is to mobilize monopsony power of the domestic population to improve the terms of trade.

Infant Industries

The optimum tariff is one of two classical cases for departure from free trade. The second classical argument justifies temporary protection to allow a new or infant domestic industry to establish itself (Kemp 1960). The theme of the infant industry

argument has also reappeared in a literature that rediscovered learning externalities to explain why diminishing returns do not constrain growth. More direct domestic policies can correct for the market imperfections that underlie the infant-industry argument (Baldwin 1969). The infant industry argument is therefore a '*second-best*' case for public policy when the 'first-best' corrective policies are unavailable. Uncompensated private learning externalities, which are often proposed as underlying a case for infant industry protection, call for compensating subsidies as the 'first-best' response. Unless the infant industry is a monopoly, protection does not compensate a domestic producer for beneficial externalities provided to other domestic competitors. There are, in addition, moral hazard problems associated with the protection of infant industries. Since the reward for doing well is the end of protection, it may be preferable for a producer with infant status never to perform too well, and so to remain a protected infant. Since there are many potential infant industries, a government has to also decide which industry to protect, and has to avoid political favors.

Distortions or the Theory of the Second Best

In the second-best situation there are uncorrectable domestic market inefficiencies (or "distortions"). The unresolved domestic inefficiencies can be due to externalities as in the case of the infant industry, or can be due to domestic monopoly, public goods, or restrictions on prices in markets such as minimum wages. The theory of the second best proposes that, if *all* domestic market imperfections cannot be corrected, departures from free trade may be efficient and increase social welfare. Minimum wages provide one example. In the minimum-wage case, a country's international comparative advantage is in labor-intensive production. The direction of international trade has however been distorted by the minimum wage, which has artificially increased the domestic cost of labor. The 'first-best' policy is to eliminate the minimum wage. However, with the minimum wage present, the realized direction of international trade may be contrary to true comparative advantage, since domestic labor looks scarce or expensive because of the minimum wage but is actually relatively abundant and cheap. The 'second-best' theory in that case proposes elimination of the "incorrect" international trade.

Another example of a second-best case for departure from free trade is based on the presence of environmental externalities. Computation of the true cost of production of a good when environmental costs are included can switch a country's comparative advantage. The first-best response is to correct the environmental externality domestically at its source. If, however, correction of the externality cannot take place at the domestic source, the "second-best" trade policy may no longer be free trade. The efficient second-best policy depends on whether the domestic industry that is the source of environmental damage is an exporter or confronts import competition. If the industry exports its output, an export tax decreases domestic production and thereby reduces domestic environmental

damage. If a polluting industry confronts import competition, government subsidy to imports is the appropriate second-best policy, since, by making competing imports cheaper, the government reduces domestic output of the local industry. There is a compendium of cases where the theory of the second best shows how efficiency gains can be achieved through departures from free trade (Bhagwati 1971).

Strategic Trade Policy

Strategic trade policy is a second-best proposal for government intervention where the second-best enters because of imperfect competition in international markets. When international markets are not competitive, rents (or excess profits) may be present. Strategic trade policy devises means of capturing the rents for a country's own nationals rather than leaving the rents with foreigners. Strategic trade policy arose as an adjunct to a body of literature that called itself the "new" international trade theory. The new theory differed from the old in recognizing that international markets might not be competitive and in emphasizing the potential importance of economies of scale. Many variants of strategic trade policy have been proposed (Brander 1995). In the basic Cournot duopoly model, for example, a domestic firm was described as confronting a foreign firm in a third market. A subsidy by the government to its domestic firm allowed the firm to credibly expand output beyond the Cournot equilibrium output, and profits or rents of the domestic firm then increased at the expense of the foreign firm. The same type of rent transfer to a domestic firm could take place through an import duty if a foreign firm were selling in the home market.

Proposals for strategic trade policy are related to the two classical cases for departure from free trade. Like the optimum tariff, strategic trade policy is based on gains in non-competitive markets at the expense of foreigners, while, in third markets, problems of retaliation arise, since a foreign government can neutralize gains from a strategic trade policy by subsidizing its own national firm. Since resources and personnel attracted to an industry favored by strategic trade policy are unavailable for other industries, policies that favor one domestic firm or industry are at the expense of other domestic firms or industries (Dixit and Grossman 1986). As with the infant-industry case, a belief in the effectiveness of strategic trade policy requires an accompanying belief that political decision makers can maximize social welfare by knowing "how to pick winners and losers" from among the domestic firms that are eligible under the theory for government assistance. All domestic firms facing foreign competition in imperfect markets are in principle eligible for assistance through strategic trade policy.

Strategic trade policy envisages policies as chosen to maximize social welfare (defined as profits of the domestic firm plus welfare of domestic consumers when intervention is in domestic and not in third markets). Nonetheless strategic trade policy benefits the firms whose profits increase (unless all profit increases can be

discriminately taxed). Given the broad scope of eligibility, beneficiaries of strategic trade policy can be selected to reward political support, such as campaign contributions, which is a different problem from that of a government having inadequate information to pick winners and losers.

Global capital markets allow individual shareholders to diversify risk by owning stock in both "domestic" and "foreign" firms. A diversified shareholder has no need for strategic trade policy. Indeed, calls for strategic trade policy introduce extraneous uncertainty into asset diversification decisions, since, in deciding on an asset portfolio, investors need to guess whether the government will heed a proposal of intervention on behalf of a firm (Feeney and Hillman 2001).

Characteristics of strategic trade policy are present in government policies toward agriculture (Bagwell and Staiger 2001). Studies have also pointed to the world duopoly of aircraft frames and have considered possibilities in semiconductors and automobiles. With the exception perhaps of agricultural subsidies, cases of policy makers following recommendations of strategic trade policy are uncommon.

Revenue Motives

Government revenue can be a motive for taxes on international trade. Taxation of internationally traded goods has the administrative advantage of goods passing through a limited number of geographic locations where revenue can be collected. Because of the ease of collection, taxes on international trade (or taxes for right of passage) were often the first taxes historically levied. Taxes on international trade have remained significant government revenue sources where domestic taxes cannot be levied because of ineffective tax administration. Where possible, domestic taxes, however, provide broader tax bases than taxes on international trade. A domestic sales tax has in particular a broader base for taxation than an import tariff, which only taxes imported goods.

Taxes on imports are often too high to maximize revenue: with sufficiently high import duties, there is of course no tax revenue at all, since there are no imports. If a country's population has the collective monopsony power necessary for an optimum tariff, the revenue-maximizing tariff exceeds the optimum tariff. By maximizing revenue from the tariff, a government would be shifting the real income abroad.

More significantly, a revenue motive is at odds with restrictions on international trade through import quotas that are freely assigned to private importers. Governments seeking revenue would auction the quotas, but auctions have been rare. In another type of import quota known as a voluntary export restraint, governments have forgone revenue by assigning rights to foreign firms to sell in domestic markets. Historical cases and contemporary instances where poorer countries lack effective tax administrations aside, revenue needs do not explain departure from free trade.

Protection and Political Economy

The normative descriptions of the beneficial consequences of departure from free trade have in common the point of departure that markets have failed to provide efficient outcomes. The 'second-best' policies specify uncorrectable market inefficiencies. Strategic trade policy is based on rents in inefficient non-competitive markets. In the classical precursors, the optimum tariff argument required non-competitive markets that allowed realization of monopsony power; and the infant-industry argument was based on markets that were inefficient because of non-internalized beneficial externalities. A political economy or public choice view in contrast accepts that markets and therefore free trade policies are proximately efficient, and looks for political incentives for policy makers to choose departures from free trade. Economic theory shows how some groups benefit from the inefficiency of a departure from free trade. Protectionist policies can benefit broad factor classes. More particularly, beneficiaries tend to be identified with incomes from non-diversified industry-specific sources (Jones, 1971). Rather than second-best *corrections* for inefficiency, the public choice or political economy view has approached departure from free trade as *creating* inefficiency, for political gain related to incomes in import-competing industries.

Feasible policies depend on (that is, are endogenous to) institutions and laws. Policy outcomes also depend on abilities of interest groups to organize and mobilize resources for collective political action (Olson 1965). Organized interest groups are generally better able to influence policy decisions than the broad population. The per capita stakes of special interests are also higher: special interests are seeking to increase their incomes, while the losses of consumers from protection of any one industry are small, because spending on the products of the industry in general comprises only a small part of an individual's or household's total spending. A public choice view predicts that, under these conditions, political-economy considerations can result in socially undesirable protectionist policies. Incumbent governments or politicians may seek maximal political support by trading off the political benefits from providing increased income to organized industry interests against the political cost of dissatisfaction of disorganized voters with departures from free trade (Hillman 1982). The incumbent policy maker may confront many organized interest groups and may be able to design a combination of policies to maximize the payments received from selling protection to the different organized interests (Grossman and Helpman 1994). Rather than decided by incumbents, policies may be determined through proposals made by candidates competing for political office (Hillman and Ursprung 1988; Magee et al. 1989). Political-support considerations have also been linked to the sudden collapse of domestic industries that have lost comparative advantage (Cassing and Hillman 1986), and to the choice of the means of protection (Cassing and Hillman 1985; Hillman and Ursprung 1988). Empirical studies have confirmed that departures from free trade are in general not the consequence of second-best intent to improve efficiency or maximize social welfare, but reflect protection related to political support and domestic income distribution (Baldwin 1984; Hillman 1989, Chap. 11; Rodrik 1995).

Contingent Protection

Contingent protection differs from protection in place. A level of protection defines protection in place. Contingent protection is defined through legal rules that specify conditions under which protection can be provided. Anti-dumping duties are a form of contingent protection. Producers can successfully undertake legal proceedings to request anti-dumping duties, if foreign firms can be shown to be causing injury through unfair competitive practices. Evidence of unfair practices (or unfair trade) may be domestic sales by foreign producers at less than cost, or sales in the domestic market at prices less than in the foreign producers' home markets. A claim of dumping is similar to a claim of predatory pricing (where firms are claimed to be selling at below cost with the intent of eliminating rivals from a market). Anti-dumping and predatory-pricing laws are complex, and are open to ambiguities in interpretation, since costs may be difficult to define and competitors reduce prices in the normal course of competition. While proven cases of predatory pricing are uncommon, claims of injury through the trade-related counterpart of dumping tend to be more often accepted by courts.

A second form of contingent protection consists of import duties that neutralize (or countervail) subsidies that foreign producers are shown to be receiving from their governments. Or the subsidies may be implicit within ownership of foreign competitors by foreign governments.

Contingent protection can also be provided without the requirement of demonstrating unfair foreign competition through escape clauses or safeguard provisions. The escape is from prior trade liberalization commitments, to safeguard an industry that is being injured by import competition. The relief from import competition is intended to be temporary (as in the infant industry case), to give a domestic industry time to adjust to competition from imports.

Contingent protection is encoded in the rules of the World Trade Organization (WTO) and the pre-1995 predecessor General Agreement on Tariffs and Trade (GATT) (Jackson 1997). Trade liberalization agreements are negotiated under conditions of uncertainty about comparative advantages in the future. Contingent protection facilitates ex-ante trade liberalization agreements under conditions of uncertainty, since governments know that liberalization can be reversed in cases where ex-post contingencies call for protection (Ethier 2002).

The legalistic language of contingent protection differs from concepts of economic theory. The unfair competition and injury defined in laws on contingent protection contradict the perspective of economic theory that competition is socially beneficial. The harm or injury defined in contingent-protection laws is incurred by producers, who benefit from less rather than form more competition. The benefit from competition in economic theory is to consumers or society at large. Contingent-protection laws therefore reflect political sensitivity to producer interests and unemployment in import-competing industries.

It is irrational for foreign producers to pay anti-dumping duties if the duties can be avoided by charging higher prices. The initiation (or threat thereof) of

anti-dumping procedures is therefore often sufficient to lead foreign producers to increase prices (Prusa 1992; Schuknecht 1992). Anti-dumping laws can thereby sustain non-competitive pricing in domestic markets by disciplining foreign producers to cooperate in accepting the price leadership role of domestic producers in domestic markets (Hillman 1990).

Escape clauses or safeguard provisions introduce moral hazards into incentives of producers to claim injury. There can be asymmetric information: producers may know, but the government and the courts may not know, whether producers' injury is due to imports, or is due to reasons such as a decline in domestic demand or inept management of domestic firms. The asymmetric information allows spurious claims of injury to be made in order to obtain the benefits of protection (Leidy and Hoekman 1991).

A mechanism of contingent protection that is not part of formal national trade law or GATT/WTO procedures takes the form of voluntary restraints on exports negotiated between governments of importing and exporting countries. The restraints set limits on total allowable foreign sales in the domestic market. To ensure adherence to the limit on imports, foreign exporters are assigned domestic market quotas. As with anti-dumping duties, pre-conditions are established for non-competitive practices. Domestic producers can set domestic prices or quantities to be sold with foreknowledge of supply by the foreign cartel that has been created by the inter-governmental agreement (Hillman 1990). The price to domestic consumers increases and domestic and foreign producers earn higher profits in the domestic market. The higher profits of domestic firms reflect the successful protectionist objective. The higher profits of foreign firms are compensation for the protection that has been provided to domestic producers (Hillman and Ursprung 1988; Hillman 1990; Ethier 1991, 2002). There are similarities and also links (Rosendorff 1996; Ethier 2002) between voluntary export restraints and anti-dumping duties. In both cases, trade policies allow non-competitive behavior that increases domestic and foreign producer profits.

Protectionism as Insurance

Contingent protection suggests insurance. Through the rules of contingent protection, import-competing producers are provided with insurance against cheaper imports. Since contingent protection is usually discriminatory, it also provides insurance to third-countries whose exports are not constrained (Ethier 1991, 2002). Protectionism has been interpreted as insurance against trade-related income losses provided by government maximizing social welfare (Eaton and Grossman 1985). Protectionism as social insurance (insurance provided by government) is another normative second-best case for departure from free trade. Social insurance is a second-best policy, because private insurance markets do not provide the income protection that people seek. Protection as social insurance has also been proposed as a positive theory to explain observed conservative income-maintaining policies in industries confronting import competition (Corden 1974).

There is a problem with a second-best normative interpretation of protection as social insurance. Asymmetric information that prevents private insurance markets from efficiently providing insurance also prevents government from replicating missing insurance markets (Dixit 1992).

A public choice perspective also notes that political motives for providing protection can look like replication of missing or incomplete insurance markets. In an expanding industry there are ambiguities in distinguishing politically provided benefits from incomes earned through personal merit and effort. The same ambiguities about sources of benefit are not present when protection increases incomes in an industry in decline because of lost comparative advantage. It is therefore politically advantageous to assist declining industries, because the benefits from political favors are clear to the beneficiaries (Hillman 1982). Characteristics of insurance are present when protection provides benefits to industries in decline. If protection is insurance, the insurance coverage is however selective and incomplete. Only import-competing industries are eligible, and import-competing industries do not benefit equally from the insurance provided by government. Industry collapse can take place in the face of cheaper imports (Cassing and Hillman 1986). The selective insurance reflects different political benefits from ex-post protection. In cases of contingent protection in particular, where an insurance motive is explicitly indicated, ambiguities about the existence and source of injury have allowed decisions about whether to provide protection to become politicized (Finger et al. 1982; Schuknecht 1992).

Domestic Political Objectives and the Terms of Trade

Domestic political objectives have been linked to effects through the terms of trade (Bagwell and Staiger 1999). The domestic efficiency costs of protectionist policies are reduced or are not incurred at all, if the efficiency costs can be shifted to people abroad through improvements in the terms of the trade. In contrast to the optimum tariff argument, the objective of government in this scenario is not necessarily gain to society through improved terms of trade, but to provide protection. Whether governments can provide politically motivated protection while felicitously increasing social welfare is an empirical question. Feasibility depends on terms of trade gains to offset domestic efficiency losses.

Whether efficiency costs of protection can be moved to foreigners at all is also an empirical question. The answer depends on the ability to influence the terms of trade, and, if the terms of trade can be influenced, on the absence of retaliation and the realization of social benefits through government revenue. If social welfare increases because of the changes in the terms of trade even though there is a political interest in providing protection, there is a normative case for departure from free trade. In this case, pursuing a political objective of protection can be socially beneficial.

Rent Seeking

A rent is income that can be increased or taken away without changing behavior. Rents are therefore earned by industry-specific factors of production that have no substitution possibilities in production, and the activity of seeking policies that increase incomes or prevent income declines in import-competing industries is therefore a form of rent seeking (Tullock 1967). Protectionist policies also provide rents for importers who obtain quota rights (Krueger 1974). Anti-dumping laws and voluntary export restraints negotiated between governments provide rents for both domestic producers and for foreign sellers. Rents and rent seeking are therefore parts of a political-economy view of international trade policy. The social losses due to trade-related rent seeking depend on how resources used in rent seeking influence political decisions, and on whether the rents that are sought are income transfers from others through protection or are in place through import (or export) quotas (Hillman and Riley 1989). The efficiency losses from willingness to depart from free trade consist of the resources attracted to rent seeking, and are an addition to the losses from protection due to substitution effects in production and consumption. Although not incorporated in the conventional normative analyses, incentives for rent seeking are also part of strategic trade policy. A government considering to follow the recommendations of strategic trade policy would face rent-seeking activity from the diverse potential beneficiaries of government assistance.

Voting

When trade policy is decided by majority voting as an election issue, there is no assurance that free trade will be chosen. A self-interested median voter will want free trade only if his or her personal assets and income sources correspond to the average asset composition and income sources for the economy at large (Mayer 1984). Trade policy can be the dominant issue in an election (Irwin 1994). In general, however, unless voters happen to live in Switzerland (Weck-Hannemann 1990), thry do not have opportunities to vote on trade policy directly. Political representatives are then in a position to decide on trade policy.

Why is Trade Policy Used to Redistribute Income?

There remains the question why political decision makers should wish to use protectionist trade policy to redistribute income. A country whose population has collective monopsony power in world markets has reason to use a tariff to achieve a domestic income distribution objective because of the benefits from the terms of trade improvements that offset, *in whole or in part*, the domestic inefficiencies of tariffs. Part of the cost of protection can thereby be transferred to foreigners. Yet, if there are gains from an optimum tariff and a government has no qualms about

imposing costs on foreigners for the benefit of the local population, we might expect the government to seek to impose the optimum tariff in any event without regard for the domestic income distribution objective. Also, optimum tariffs do not seem all that relevant for many goods and many governments.

If world prices are more or less independent of domestic demand, protectionist policies create domestic inefficiencies without offsetting the changes in the terms of trade. The domestic inefficiencies could be avoided if non-distorting lump-sum taxes and subsidies were available to redistribute income. Since non-distorting means of redistributing income are in practice not feasible, policy makers have no choice but to use some form of inefficiency-creating mechanism to redistribute income. Still, this does not answer the question why trade policy should be used to redistribute income, since there are in general income transfer mechanisms that incur smaller efficiency losses (Mayer and Riezman 1990). Governments should be expected to use these more efficient means of income transfer, since, by consensus, everybody in the population would wish the inefficiency associated with redistribution to be minimized.

The consensus in favor of efficiency has been the basis for a prediction that *in practice* political redistribution *is always undertaken* in the most efficient way (Wittman 1995). If that were so, departures from free trade should be observed as a means of income distribution only when more efficient means of redistribution are unavailable. All observed trade restrictions could then be interpreted ex-post as having been the most efficient ex-ante means of achieving the policy makers' objectives of income redistribution.

Choice of the efficient means of income redistribution is however compromised by political benefits from information asymmetries. Information about government policy has political consequences. Political decision makers gain by not publicizing to voters at large policies that benefit special interests. Surreptitious or hidden income transfers are politically more advantageous. Departures from free trade are an obtuse means of transferring income. Voters may not be aware that a tariff that taxes foreign goods is at the same time a subsidy to domestic import-competing producers. The rhetoric of unfair foreign competition or protecting domestic jobs against foreign competition may be used. Voluntary export restraints are a particularly obtuse means of income redistribution through trade restrictions. The government sets limits on permissible quantities of imports and directs foreign exporters to set market shares. Foreign exporters thereby establish a cartel for supply to the domestic market. The restricted domestic supply increases the domestic price, which provides the protectionist income transfer to import-competing domestic producers. The benefits to domestic producers from protection have been achieved through voluntary compliance with foreign competitors.

If the information is not personally useful, voters have reason to be "rationally ignorant" of trade policy issues. Voters are however not equally ignorant of all income transfer mechanisms. A direct income transfer that is "hidden" in a one line item of a government's budget can be found if someone is looking for it. The transfer of income via a tariff from consumers to protected producers is indirect and less obvious. Tariffs have the politically expedient characteristic that domestic

buyers directly make income transfers to domestic producers through the increased domestic price facilitated by the tariff. The income transfer achieved through protectionism does not require intermediation of government through taxation and budgetary allocations. The indirect nature of redistribution by trade policy therefore explains why international trade restrictions are used as means of income redistribution when more efficient but more obvious means of income transfer are available (Magee et al. 1989). Protection then makes clear to the beneficiaries that the government has provided them with benefits, when voters at large have reason to be "rationally ignorant" of trade policy issues.

Agriculture

Agriculture has been a special case for government intervention. Rarely have governments left agriculture to the intervention-free determination of markets (Anderson and Josling 1993). Agriculture is often taxed in poorer countries, where agriculture is a large part of national income and agricultural goods are exported. An export tax is sometimes directly levied or government enforces a position for itself as monopsonistic domestic buyer and pays farmers a low price and sells in the world market at a higher price. To obtain revenue from an export tax or domestic monopsony, the government needs to be effective in preventing smuggling, which creates a need for resources for policing of borders. The benefits from goods escaping export tax or the monopsony price and reaching the market outside the country introduce gains from corruption through the participation of border officials in smuggling activities. If the corruption reaches the government, smuggling can be extensive and little official government revenue may be provided.

The taxes on agriculture in poorer countries reflect the search for extensive and available tax bases, and also, since agricultural sectors are large in poorer countries, the taxes on agriculture also reflect the principle that larger groups face higher costs of collective action. Since many of these countries are dictatorships or quasi-dictatorships, the taxes also reflect the fact that those with power exploit the powerless.

The principle of organizational advantage applied to the effectiveness of organization of small groups underlies government assistance to agriculture in richer countries, where agriculture has been extensively subsidized or protected. The policies that support agriculture in richer countries are also sometimes explained as justified by an objective of sustaining traditional rural life and avoiding depopulation of the countryside. The beneficiaries of agricultural subsidies are however often large firms rather than smaller family farms.

Trade conflicts involving agriculture have often been framed in terms of motives other than protection. For example, European restrictions on imports of U.S. beef have been framed in terms of the purported health hazard from hormones given to U.S. cattle. Protectionism has reflected former colonial ties in discrimination by the European Union in favor of imports of bananas from former European colonies, to the disadvantage of bananas grown (often on U.S.-owned plantations) in Central America.

National Security

Protection of agriculture is often justified on grounds of national security. Consequences of vulnerability to foreign suppliers were demonstrated when the international oil cartel OPEC imposed export embargos. There have also been cases where countries under threat from foreign aggressors found that defense equipment, which had been ordered and paid for, were withheld by foreign suppliers. In other cases, when foreign-purchased defense equipment has been required for self-defense, foreign governments have withheld spare parts. Trade embargos provide a normative case for self-reliance because of national security concerns (Mayer 1977; Arad and Hillman 1979). Countries also impose restrictions on exports because of national security concerns.

Views of Government

With national security and some other limited cases as exceptions (for example, trade in heroin), there is a compelling case for free trade independent of international boundaries. Departures from free trade have however often taken place. The political economy premises of the public choice approach point to political motives and income distribution as underlying the departures from free trade and to inefficiencies incurred, including through rent seeking. Theories set out in the conventional normative view have, in contrast, described how governments can act in the public interest by correcting inefficiencies when departing from free trade.

Since the political economy premises of public choice offer positive conclusions and the conventional theories offer normative recommendations, the two approaches have been complementary. Open lines of communication between the approaches, however, require that a clear distinction between normative and positive analysis be maintained. The distinction is lost and lines of communication are not present when a normative belief that government *should* act in the public interest becomes a prediction that government *will* always act in the public interest, because government *should* be benevolent. The censorship that is then implicitly imposed limits politically correct economic analysis to normative theory where government can do no wrong (Hillman 1998). Since non-virtuous government is by hypothesis excluded from economic analysis, the consequent theories *can only* be normative. Addressing *why* governments have chosen to depart from free trade may require introducing non-virtuous government into economic analysis. A public choice perspective would advise caution in pursuing a research agenda that provides a repertoire of normative arguments consistent with departure from free trade by a virtuous government. When policy makers are politically motivated, the normative proposals can be misused to justify politically expedient policy decisions.

In the mid-1990s, the political economy premises of public choice began to be widely adopted in descriptions of departure from free trade (Grossman and Helpman 2002). With the exception of agriculture and national security, and limited incidents

of contingent protection, governments were at the same time, after extensive liberalization, no longer significantly departing from free trade (*see trade liberalization and globalization*).

References

Anderson, K. and Josling, T. (1993). "The challenge to economists of multilateral trade negotiations on agricultural protection." *Food Research Institute Studies*, 24, 275–304; reprinted (1995) in G.H. Peters (ed.) *Agricultural Economics. International Library of Critical Writings in Economics*, ch. 21. Cheltenham, U.K.: Edward Elgar Publishing.

Arad, R.W. and Hillman, A.L. (1979). "Embargo threat, learning and departure from comparative advantage." *Journal of International Economics*, 9, 265–275.

Bagwell, K. and Staiger, R.W. (1999). "An economic theory of GATT." *American Economic Review*, 89, 215–248.

Bagwell, K. and Staiger, R.W. (2001). "Strategic trade, competitive industries, and agricultural trade disputes." *Economics and Politics*, 13, 113–128.

Baldwin, R.E. (1969). "The case against infant-industry protection." *Journal of Political Economy*, 77, 295–305.

Baldwin, R.E. (1984). "Trade policies in developed countries," in R. Jones and P. Kenen (eds.) *Handbook of International Economics*. Amsterdam: North-Holland, pp. 571–619.

Bhagwati, J.N. (1971). "The generalized theory of distortions and welfare," in J.N. Bhagwati, R.W. Jones, R.A. Mundell, and J. Vanek (eds.) *Trade, Balance of Payments, and Growth: Essays in Honor of Charles P. Kindleberger*. Amsterdam: North-Holland, pp. 69–90.

Brander, J.A. (1995). "Strategic trade policy," in G.M. Grossman and K. Rogoff (eds.) *Handbook of International Economics*. Amsterdam: North-Holland, pp. 1395–1455.

Cassing, J.H. and Hillman, A.L. (1985). "Political influence motives and the choice between tariffs and quotas." *Journal of International Economics*, 19, 279–290.

Cassing, J.H. and Hillman, A.L. (1986). "Shifting comparative advantage and senescent industry collapse." *American Economic Review*, 76, 516–523; reprinted (2004) in J.N. Bhagwati and B.P. Rosendorff (eds.) *Readings in the Political Economy of Trade Policy*. Cambridge, MA: MIT Press.

Corden, W.C. (1974). *Trade Policy and Economic Welfare*, (2nd edn, 1997), Oxford: Oxford University Press.

Dixit, A. and Grossman, G.M. (1986). "Targeted export promotion with several oligopolistic industries." *Journal of International Economics*, 21, 233–249.

Dixit, A. (1992). "Trade policy with imperfect information, in R.W. Jones and A.O. Krueger (eds.). The Political Economy of International Trade: Essays in Honor of Robert E. Baldwin. Oxford: Blackwell, pp. 9–24.

Eaton, J. and Grossman, G.M. (1985). "Optimal commercial policy when markets are incomplete." *Canadian Journal of Economics*, 18, 258–272.

Ethier, W.J. (1991). "The economics and political economy of managed trade," in A.L. Hillman (ed.) *Markets and Politicians:* Boston: Kluwer, pp. 283–306.

Ethier, W.J. (2002). "Unilateralism in a multilateral world." *Economic Journal*, 112, 266–292.

Feeney, J. and Hillman, A.L. (2001). "Privatization and the political economy of strategic trade policy." *International Economic Review*, 42, 535–556.

Finger, J.M., Hall, H.K., and Nelson, D.R. (1982). "The political economy of administered protection." *American Economic Review*, 72, 452–466.

Grossman, G.M. and Helpman, E. (1994). "Protection for sale." *American Economic Review*, 84, 833–850; reprinted (2004) in J.N. Bhagwati and B.P. Rosendorff (eds.) *Readings in the Political Economy of Trade Policy*. Cambridge, MA: MIT Press.

Grossman, G.M. and Helpman, E. (2002). *Interest Groups and Trade Policy*. Princeton, NJ: Princeton University Press.

Hillman, A.L. (1982). "Declining industries and political-support protectionist motives." *American Economic Review*, 72, 1180–1187; reprinted (2004) in J.N. Bhagwati and B.P. Rosendorff (eds.) *Readings in the Political Economy of Trade Policy*. Cambridge, MA: MIT Press.

Hillman, A.L. (1989). *The Political Economy of Protection*. New York: Harwood.

Hillman, A.L. (1990). "Protectionist policies as the regulation of international industry." *Public Choice*, 67, 101–110.

Hillman, A.L. (1998). "Political economy and political correctness." *Public Choice*, 96, 219–239 (Presidential Address, European Public Choice Society, Prague, April 1997).

Hillman, A.L. and Riley, J. (1989). "Politically contestable rents and transfers." *Economics and Politics*, 1, 17–39; reprinted (2004) in J.N. Bhagwati and B.P. Rosendorff (eds.) *Readings in the Political Economy of Trade Policy*. Cambridge MA: MIT Press.

Hillman, A.L. and Ursprung, H.W. (1988). "Domestic politics, foreign interests and international trade policy." *American Economic Review*, 78, 729–745; reprinted (2004) in J.N. Bhagwati and B.P. Rosendorff (eds.) *Readings in the Political Economy of Trade Policy*. Cambridge MA: MIT Press.

Irwin, D. (1994). "The political economy of free trade: Voting in the British general election of 1906." *Journal of Law and Economics*, 37, 75–108.

Jackson, J.H. (1997). The World Trading System: Law and Policy of International Relations, 2nd edn. Cambridge, MA: MIT Press.

Johnson, H.G. (1953–54). "Optimum tariffs and retaliation." *Review of Economic Studies*, 21, 142–153.

Jones, R.W. (1971). "A three-factor model in theory, trade, and history," in J.N. Bhagwati, R.W. Jones, R.A. Mundell, and J. Vanek (eds.) Trade, the Balance of Payments, and Growth: Essays in Honor of Charles B. Kindleberger. Amsterdam: North-Holland, pp. 3–21.

Kemp, M.C. (1960). "The mill-bastable infant-industry dogma." *Journal of Political Economy*, 68, 65–67.

Kemp, M.C. (1962). "The gains from international trade." *Economic Journal*, 72, 803–819.

Krueger, A.O. (1974). "The political economy of the rent seeking society." *American Economic Review*, 64, 291–303; reprinted (1980) in J.M. Buchanan, R.D. Tollison, and G. Tullock (eds.) *Toward a Theory of the Rent-Seeking Society*. Texas: Texas A&M Press, pp. 51–70.

Leidy, M.P. and Hoekman, B.M. (1991). "Spurious injury as indirect rent seeking: free trade under the prospect of protection." *Economics and Politics*, 3, 111–137.

Magee, S., Brock, W.A., and Young, L. (1989). *Black Hole Tariffs and Endogenous Policy Theory*. Cambridge, MA: Cambridge University Press.

Mayer, W. (1977). "The national defense tariff argument reconsidered." *Journal of International Economics*, 7, 363–377.

Mayer, W. (1984). "Endogenous tariff formation." *American Economic Reviewv*, 74, 970–985; reprinted (2004) in J.N. Bhagwati and B.P. Rosendorff (eds.) *Readings in the Political Economy of Trade Policy*. Cambridge, MA: MIT Press.

Mayer, W. and Riezman, R. (1990). "Voter preferences for trade policy instruments." *Economics and Politics*, 2, 259–273.

Olson, M. (1965). The Logic of Collective Action: Public Goods and the Theory of Groups. Cambridge, MA: Harvard University Press.

Prusa, T.J. (1992). "Why are so many anti-dumping petitions withdrawn?" *Journal of International Economics*, 33, 1–20.

Rodrik, D. (1995). "Political economy of trade policy," in G.M. Grossman and K. Rogoff (eds.) *Handbook of International Economics*. Amsterdam: North-Holland, pp. 1457–1495.

Rosendorff, B.P. (1996). "Voluntary export restraints, antidumping procedure, and domestic politics." *American Economic Review*, 86, 544–561; reprinted (2004) in J.N. Bhagwati and B.P. Rosendorff (eds.) *Readings in the Political Economy of Trade Policy*. Cambridge MA: MIT Press.

Samuelson, P.A. (1962). "The gains from trade once again." *Economic Journal*, 72, 820–829.

Schuknecht, L. (1992). *Trade Protection in the European Community*. New York: Harwood.

Tanzi, V. and Schuknecht, L. (2000). *Public Spending in the 20th Century*. New York: Cambridge University Press.

Tullock, G. (1967). "The welfare costs of tariffs, monopoly, and theft." *Western Economic Journal*, 5, 224–232; reprinted (1980) in J.M. Buchanan, R.D. Tollison, and G. Tullock (eds.) *Toward a Theory of the Rent-Seeking Society*. Texas: Texas A&M Press, 39–50.

Weck-Hannemann, H. (1990). "Protection under direct democracy." *Journal of Theoretical and Institutional Economics*, 146, 389–418.

Wittman, D. (1995). *The Myth of Democratic Failure: Why Democratic Institutions are Efficient*. Chicago: The University of Chicago Press.

Chapter 27
Trade Liberalization and Globalization[1]

Arye Hillman

Trade liberalization is the reverse process of protectionism. After previous protectionist decisions, trade liberalization occurs when governments decide to move back toward free trade. Trade liberalization may take place unilaterally. Extensive trade liberalization that occurred among the richer countries in the second half of the twentieth century was however reciprocal and multilateral. Many governments reciprocated each other's liberalization decisions, and the liberalization was nondiscriminatory in applying to all liberalizing trading partners. The trade liberalization was accompanied by liberalization of international capital markets and by substantial international migration, both legal and illegal. International agreements and conventions also unified rules of conduct concerning protection of property rights, including intellectual property rights related to proprietary knowledge. The outcome of these liberalizing and integrating processes is known as globalization. The issues are why or how did globalization occur, and whether the outcome has been good for humanity.

Reciprocal Liberalization

A government can transfer income to an import-competing industry through unilateral protection. Incomes in export industries can be unilaterally increased through export subsidies. The articles of the pre-1995 General Agreement on Tariffs and Trade (GATT) and the rules of the successor World Trade Organization (WTO) do not view export subsidies favorably, and allow duties to be levied to counter export subsidies. If there has been past protection, income can instead be transferred to an export sector by unilateral liberalization, provided that exporters have access to foreign markets. Foreign protectionist barriers may limit access to foreign markets to exporters. When foreign import quotas limit exports, exporters need to be included in the quota to have market access and cannot sell more than the allowed quantities. Foreign tariffs impose additional costs of selling in the foreign markets. Exporters

---

[1] This chapter first appeared as an essay in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 312–320.

will regard the elimination or reduction of foreign import quotas and tariffs as providing them with market access, or market access under more favorable conditions.

Policy makers dealing with trade liberalization are, in practice, vitally concerned with such notions of market access. By exchanging market access for each other's exporters, governments hope to be seen as reciprocally providing benefits for each other's export sectors with an effectiveness, or a visibility, that may not be possible through unilateral liberalization. Each government benefits politically from the market access provided for its export industries by the other reciprocating government (Hillman et al. 1995; Hillman and Moser 1996).

This practical concern with the exchange of market access reflects a mercantilist view of international trade by policy makers. The mercantilist view is that granting of access of foreign goods to the domestic market is not a socially beneficial policy for the liberalizing economy. Rather, allowing foreigners to sell in the home market under more favorable conditions of market access is a "concession" or "favor" that compromises the rights of domestic producers to their own market and requires a reciprocal favor of foreign market access in return.

How much protection each government retains after providing and receiving market access for exporters depends on the relative political influence, or political importance, of export and import-competing industries. Politically optimal exchange of market access need not therefore result in an agreement to eliminate all trade barriers to implement free trade. Reciprocal liberalization will in particular be gradual, if governments are constrained in the magnitude of the income losses that can be imposed at any point in time on the import-competing industries that lose from the reciprocal liberalization (Ethier 2001, 2002).

Consider a country that is small relative to the world market. The world market consists of the combined national markets of all potential trading partners. With open access to the world market under conditions of free trade, exporters in such a small country have no problems of market access. They can sell all they wish at given world prices. If however the world market consists of segmented national markets with import quotas or tariffs, and if the small country is large in any of these markets, its exporters will benefit from improved conditions of foreign market access. This potentially provides governments of even small countries with incentives to participate in negotiations for reciprocal exchange of market access.

Any particular exporting firm in any country is likely to be grateful for perceived improvements in conditions of foreign market access, either through elimination or relaxation of quota restrictions or through reduced foreign tariffs. The firm may be able to receive a higher price for its exports, which from the country's perspective is an improvement in the terms of trade. Improved foreign market access is beneficial for the firm even if new export sales take place without price increases. The benefit is evident if the firm is not capacity constrained. If the firm has no immediate excess capacity, new investment can increase capacity. With the firm's selling price providing a mark-up of costs, providing long-run average costs does not significantly increase when production capacity is expanded; profits increase through increased export sales. If the expansion of firms is at the expense of the import-competing

sector, through diversion of investment and labor hired away from import-competing firms, incomes in the import-competing sector decline. A government that reciprocally liberalizes imposes losses on the domestic import-competing sector.

Since new markets that allow increased sales without increases in price are valuable to exporting firms, the exchange of market access that reciprocally benefits export firms in different countries does not require improvements in the terms of trade. Indeed, with only two countries, it is impossible for liberalization through reciprocal exchange of market access to improve the terms of trade of both countries. Each country benefits from more trade at unchanged terms of trade because protection causes each to value imports higher than what it must pay to buy them from its partner.

Unilateral liberalization would however deteriorate the terms of trade of the liberalizing country (if the country's population can collectively influence the terms of trade). Reciprocal liberalization through exchange of market access avoids significant losses in the terms of trade for any country. Reciprocal tariff liberalization could, for example, leave the terms of trade unchanged. Export sectors in both countries will have however benefited from an increase in the relative domestic price of their products that is not neutralized, even in part, by a deterioration in the terms of trade.

Through reciprocal liberalization, two liberalizing governments forgo tax revenue to transfer income to their own exporters, while reciprocal liberalization prevents part of that revenue from being captured by foreign exporters. The reciprocal trade liberalization that occurs is also beneficial for society at large in each country, by moving countries closer to free trade. The social benefits of more liberal trade policies have come about, however, because of the political interest in opening foreign markets to export industries (Hillman et al. 1995; Hillman and Moser 1996).

Terms of Trade Changes

Trade liberalization as exchange of market access is consistent with the political-economy premises of a public choice view of policy determination, because of the focus on income distribution and political motives for policy decisions. An alternative view emphasizes the effects of terms of trade changes on social welfare. After governments have reciprocally imposed tariffs with the intent of improving the terms of trade, there are in general reciprocal incentives to liberalize trade. Countries can possibly be better off in the Nash equilibrium than in free trade (see *international trade policy: departure from free trade*), but losses incurred because of the tariffs through domestic inefficiencies and reduced volumes of trade more generally provide mutual gains from trade liberalization. Although mutual tariffs are the Nash equilibrium outcome in a single-move prisoners' dilemma game where governments impose tariffs with the intent of improving the terms of trade, applications of the theory of repeated games, and in particular the folk theorem of repeated games, show how mutually beneficial self-enforcing

contracts to move toward free trade can be an equilibrium outcome. It is typically assumed that such contracts are supported by the threat that any deviation by a government would result in reversion to the static Nash equilibrium tariffs. However, in actual practice, deviations normally trigger negotiations (which in the WTO are governed by an elaborate dispute settlement procedure) to determine what retaliatory tariffs should be imposed. This raises the possibility that renegotiation could undermine the threats on which the contracts are based, and when tariff contracts are restricted to being "renegotiation proof", free trade cannot be achieved (Ludema, 2001).

When negotiations however take place to liberalize trade, the issues actually on the agenda involve willingness to reduce protectionist barriers in order to exchange market access, rather than how trade liberalization will affect the terms of trade. The contingent protection (see *international trade policy: departure from free trade*) that accompanies trade liberalization agreements confirms the political sensitivity to income distribution consequences of trade liberalization.

Still, if terms of trade changes are significant, governments would have incentives to take terms of trade changes into account when formulating policies to achieve domestic income distribution objectives (Bagwell and Staiger, 1999, 2001). Because of changes in the terms of trade, a government may not be prepared to liberalize unilaterally to assist export sectors. If the terms of trade are not determined in a broader world market, prices received by home exporters fall when the terms of trade decline because of unilateral liberalization. There are therefore two influences on prices of the export sector's output when trade liberalization is unilateral. The export sector benefits with unchanged terms of trade. If however the terms of trade deteriorate, the sensitive problem for a government may not be the social loss, but the compromise of the objective of liberalizing to benefit the export sector. Reciprocal liberalization avoids or moderates the deterioration in the terms of trade that would disadvantage exporters.

While terms of trade effects influence incentives to negotiate reciprocal liberalization when terms of trade effects are significant, the incentives for reciprocal trade liberalization are present without terms of trade changes, through the mutual political benefits from increasing exporters' incomes through reciprocal liberalization to exchange market access (Hillman et al. 1995; Hillman and Moser 1996; Ethier 2001, 2002).

Multilateral Liberalization

Trade liberalization has not been bilateral, but negotiated in a sequence of multilateral agreements. The multilateral agreements were based on a most-favored nation clause (MFN), which required any market access for exporters from one country to be provided to exporters from all countries (Horn and Mavriodis 2001). MFN is a means of confronting opportunism in the exchange of market access in a multi-country world (Ethier 2001). Market access can be simultaneously exchanged with many countries, in principle at different terms. For example, two countries might

negotiate trade liberalization and agree to levy tariffs of 20% on each other's trade, when trade with other countries is subject to higher tariffs. Afterwards, one of the two countries can proceed to negotiate further liberalization with a third country with (for example) reciprocal 10% import duties. The market access provided to the first trading partner is thereby devalued, since the first trading country still confronts the 20% tariff while the new trading partner is advantaged by the superior market access of a 10% tariff. The way to avoid such devaluation of benefit of negotiated market access is to insist that whatever "favors" are granted to one country are also granted to all other countries with whom trade agreements are in place. The outcome is multilateralism in trade liberalization.

Non-discrimination in trade liberalization through the MFN clause was therefore not due to principles of "fairness" and enlightenment in seeking equal treatment to establish a liberal international trading environment. The universal openness of multilateral liberalization was rather the equilibrium outcome of the non-sustainability of bilateralism in the face of potential opportunism in exchange of market access. Multilateralism protected the prior benefits of producers that had been negotiated by exchange of market access (Ethier 2001, 2002).

Liberalization and Contingent Protection

Contingent protection (see *international trade policy: departure from free trade*) allows governments to protect selected industries even though prior commitments were made that trade was to be liberal. Contingent protection encourages ex-ante agreement on trade liberalization, since liberalization agreements do not have to be complete in covering all possible contingencies regarding future changes in international comparative advantage.

Contingent protection is therefore a form of political insurance when trade liberalization is negotiated. Future outcomes that are politically non-tenable can be addressed if the need arises through the provisions of contingent protection. Multilateralism adds a further dimension to the insurance role of contingent protection (Ethier 2002). Countries negotiate trade liberalization realizing that, ex post, no sovereign government will do anything that is against its own interest. Negotiated liberalization therefore cannot exceed the liberalization preferred, ex post, by the country most reluctant to liberalize. When negotiations take place, it is not known how future changes in comparative advantage will affect exporters of different countries. Future outcomes can result in leaders (whose exporters are the most efficient), followers (whose exporters compete abroad with those of the leaders), and laggards (whose import-competing firms compete with exporters from both the leaders and the followers). The leaders will capture export markets from the laggards, and also from followers who export to the same markets. The followers, whose goods are less attractive to buyers for price or quality reasons, will then lose sales in export markets. Discriminatory protection by laggards against leaders through anti-dumping duties or voluntary export restraints protects the laggard's own home market and also protects followers from the advantages of the leaders.

This form of protection however, provides the leaders with compensatory rent transfers. Adequate compensation provides leaders with incentives to accept the protection of the laggards without retaliation. Governments know at the stage of negotiation of liberalization agreements that, should their exporters in the future be followers rather than leaders, protection by the laggards will discriminate against the leaders. Under conditions of uncertainty about whose exporters will be leaders, followers, or laggards, contingent protection therefore makes liberalization more attractive than if the discriminatory contingent protection were not part of future policy possibilities. Thus, ex ante, more liberalization can be negotiated. In a multilateral context, the insurance against adverse comparative advantage outcomes provided through contingent protection is, therefore, also multilateral (Ethier 2002).

Preferential Trading and Regionalism

Preferential trading arrangements (Pomfret 1988) depart from MFN treatment in providing selective discriminatory exchange of market access to participating countries. While discriminatory preferences contradict the MFN treatment, the arrangements are permissible in the GATT—WTO framework, provided the participating countries substantially reduce internal trade barriers among themselves and trade barriers against other countries do not rise on an average.

A preferential trading agreement can take the form of a customs union such as in the case of the European Union, or a free-trade area as in the case of NAFTA (North American Free Trade Agreement). The members of a customs union have a common foreign trade policy. A free-trade agreement permits each member country to maintain an independent trade policy with other non-member countries.

A free-trade agreement therefore requires internal border policing to certify the origin of goods and to prevent goods entering the free-trade area through the least protectionist country and then moving to more protectionist countries. Preferential trading arrangements are usually regional, and are part of a liberalizing process that has been called regionalism. A free-trade agreement avoids opportunist behavior in exchange for market access. By completely liberalizing all the way to free trade, the countries in the regional agreements are left with no scope for depreciating the value of the agreements by offering better terms of market access to others.

There is no assurance a customs union is on balance beneficial for a domestic population. A customs union provides benefits through liberal internal trade among member countries according to comparative advantage, and also provides benefits through a greater variety of products available to consumers in the expanded market (Levy 1997). These gains are balanced against losses from trade diversion.

Trade diversion occurs when, because of the preference in market access to member countries, the customs union makes more expensive goods seem cheaper to a member country's consumers. For example, a good from the cheapest foreign source may cost $00, but the country is outside the customs union and an import

duty of 50% is levied on the good. Consumers would therefore pay $150 if they purchased imports from the cheapest foreign source. Producers from a country within the customs union can supply the same good for $140, and there is no duty because of the free market access of the member country. Consumers therefore buy from the cheaper member country's producers. There is therefore a $40 loss on every unit of the good imported. If domestic consumers were to buy from the cheapest foreign source, the cost would be $100 paid to the foreign producer, and an import duty of $50 paid to the home government. A benevolent government would however return the $50 to consumers through public spending or reduced taxation, leaving a net cost to consumers of $100. In these circumstances, a country's consumers confront a multi-person prisoners' dilemma (Hillman 1989). The country's consumers would all be better off if they collectively ignored the privately cheaper duty-free good, and bought the privately more expensive good that is less expensive at world prices.

Trade diversion can occur when a customs union complies with the GATT/WTO restriction that trade barriers against non-member countries do not increase. Under a first-best policy of non-discriminatory free trade, losses from trade diversion could not occur. A customs union is a case of the second-best, because free trade is not with everybody (Viner 1950). Because of trade diversion, the question whether joining a customs union is beneficial for a country's population requires an empirical answer.

A free trade agreement allows each government to set import duties to avoid losses from trade diversion. If external duties in a customs union are low, prospects for trade diversion are also low: free trade within the European Union emerged in the latter parts of the twentieth century in the aftermath of extensive multilateral trade liberalization that reduced the scope of anticipated losses from trade diversion. Changes in political will to protect after entry into a customs union can also diminish trade diversion (Richardson 1993).

Although the formation of a customs union can result in net losses because of trade diversion, in principle every move towards world free trade through membership of countries in customs unions can be manipulated to be beneficial for members of the customs union without loss to other countries (Kemp and Wan 1976). Whenever a group of countries forms a customs union, it is possible to find a common external tariff for the customs union and compensating lump-sum payments between members of the union, such that no person, whether in the customs union or not, is worse off than before the formation of the union. The lump-sum transfers here are among the governments of the countries, in the first instance. Further lump-sum transfers would distribute the compensating payments to individuals. With the provision that lump-sum transfers among individuals are feasible, the path to world free trade through the formation of groups of countries into customs unions can be made Pareto-improving (that is, some people could always be made better off in each step without anyone being made worse off). A problem is that members of a customs union need have no particular incentive to constrain themselves to adopt a common external tariff that does not harm outsiders.

Non-Economic Motives

The motives for preferential trading agreements are often non-economic. Regional governments do not usually commission measurement of gains from trade creation and losses from trade diversion. The motives for formation of the European Union were not principally economic. The intent of the founders was to end the European conflicts that had been due to past animosities among nation states, and to provide a counter to the United States. There was also a prescience that a united Germany would fit better within a united Europe.

Foreign Investment and Migration

Regional preferential trading arrangements are a means of poorer countries competing for foreign direct investment from richer countries (Ethier 1998a, 1998b). The attraction for foreign investment is that a free-trade agreement allows duty-free import of inputs and duty-free export of goods produced in poorer countries to markets of richer countries. For example, the exchange of market access through NAFTA was one sided, with low-income Mexico eliminating more protectionist barriers than the United States and Canada. In mercantilist terms, Mexico made more concessions or gave up more than it received, since it already had quite free access to the U.S. and Canadian markets. NAFTA however provided incentives for foreign investment to go into Mexico, rather than into other countries like Mexico, because of the marginal increase in market access to its northern neighbors that NAFTA gave to Mexico. From the vantage of the United States, foreign investment in Mexico could reduce incentives for illegal immigration by increasing the demand for local labor.

Unilateral Trade Liberalization

Governments have sometimes liberalized trade policy unilaterally, without the strategic considerations of exchange of market access. By liberalizing unilaterally, a country's government places itself in a situation where there are no "concessions" left to exchange with foreign governments that have not likewise completely liberalized. Unilateral liberalization took place as a part of a process of development assistance to poorer countries, although evidence indicates that political influence affected the market-access concessions that were granted (Ray, 1987). Unilateral liberalization in poorer countries has also sometimes occurred as part of policy conditionality for World Bank assistance.

The most prominently researched case of unilateral trade liberalization is the repeal of the Corn Laws in nineteenth century England (Irwin 1989; Schonhardt-Bailey 1996, 1997). The Napoleonic Wars had provided natural protection for English agriculture. Afterwards, the natural protection was replaced by protection through government policy. One hypothesis is that repeal of the Corn Laws, as the

protectionist polices were known, reflected enlightened liberal trade policy responding to the case for free trade that had been made by David Ricardo through his theory of comparative advantage. A political economy view looks in the direction of political popularity and private self-interest to explain the unilateral trade liberalization. England imported food, and the real wage was determined in terms of food. Those who benefited from the protection provided by the Corn Laws had however successfully resisted trade liberalization for some decades to maintain their agricultural interests. The trade liberalization took place when the previous stringent opposition of agricultural interests had subsided. An investigation of the asset composition in the estates left by landowners reveals diversification of asset ownership out of land and into transport and industry (Schonhardt-Bailey, 1991). No one in a country would have an interest in protectionist policies and there would be national consensus for free trade, if domestic asset markets permit the population of a country to perfectly diversify asset holdings to reflect the composition of national productive assets or resources. Free trade, which maximizes national income, would then also maximize each individual's personal income. It is *asymmetric* domestic asset ownership that defines special interests, and which underlies the political-economy relation between income distribution and endogenous protectionist policies. The asset diversification of prior landed interests is consistent with a change from previous support for protection to support for trade liberalization.

More generally, amenability to trade liberalization can be linked to development of domestic and global asset markets that have allowed diversification of personal income sources (Feeney and Hillman 2001). The asset diversification moderates or eliminates the association between individuals' incomes and special-interest industry identities. Industry-specific factors of production continue to exist, but the income from the industry-specific factors is spread by opportunities for diversification in asset ownership. As asset markets became more developed in the second half of the twentieth century, the asset diversification reduced industry-specific associations and changed personal interests to be more supportive of trade liberalization. When governments negotiated trade liberalization, asset markets moderated prior domestic opposition to liberal trade policy, and there were more beneficiaries of the expanded national income from exchange of market access.

Markets allowing people directly to diversify their human capital holdings do not exist, and there is a question whether opportunities for diversification of physical capital allow a fully diversified portfolio to be achieved by balancing non-diversifiable human capital against diversifiable physical capital. There will also be principal-agent problems when individuals are completely diversified. Risk-averse investors invest in mutual funds to diversify their asset portfolios, and, if fully diversified, have an interest in trade liberalization. At the same time, stock option schemes, which are intended to give managers an interest in the profitability of individual firms, also give managers an interest in lobbying for industry protection that increases the value of the stock options. Solving the shareholder—manager principal—agent problem through stock options therefore gives rise to another principal—agent problem where managers resist the liberal trade policies that benefit diversified shareholders (Cassing, 1996).

Globalization

Liberalization of international trade, and also of international capital transactions, resulted toward the end of the twentieth century in a phenomenon known as globalization. The term globalization refers to the integration of national markets into global markets. Globalization occurred as multilateral trade liberalization and the regional agreements reduced trade barriers among the richer countries, and as restrictions on international capital market transactions were lifted. Globalization was also facilitated by the new ease of international communications. Trade patterns, which had previously been based on the richer countries trading among themselves and importing raw materials and low-valued goods from poorer countries, changed. In the United States, the proportion of imports from the poorer countries increased from 14% in 1970 to 35% in 1990. In Western Europe the increase was from 5% to 12% in the same period (inclusive of intra-European trade). The types of goods exported by the poorer countries also changed. By 1992, 58% of the exports from the developing countries to the developed world were light manufactured goods, compared with 5% in 1955 when many of the poorer countries were still colonies of European countries. The change in the volume and composition of the international trade of poorer countries was accompanied by increased international mobility of capital. Political risk was reduced in those poorer countries where the rule of law prevailed and private property rights were protected. Adherence to the rule of law provides assurance for foreign investors that their ownership rights were protected. At the same time, domestic changes in the poorer countries led to improvements in education and health of the local populations, which permitted domestic labor to be mobilized for organized market activity.

The changes affected domestic income distribution. Economic theory predicts that, without international investment and migration, free trade in goods in the long run equalizes real wages internationally or at least reduces wage differences. When liberalization of capital movements and foreign investment equalizes risk-adjusted returns to capital across countries, real wages tend to be equalized in the short run. As trade liberalization proceeds, domestic changes are in particular predicted to take place in income distribution to the detriment of the unskilled low-human capital workers in the richer human-capital abundant countries of the world. The equalizing tendencies in real wages become more pronounced when substantial migration from poorer to richer countries takes place, legally and illegally. When income distribution becomes globalized, personal incomes tend to depend more on individuals' personal capabilities and education rather than on where they live. Thus, in the latter part of the twentieth century, real incomes of unskilled workers in richer countries declined absolutely, and also relative to skilled workers' incomes. For example, the U.S. male college-education high-school premium was 40% in 1979 and 74% in 1996. During the same period, the male college premium for completion of high school increased from 73% to 157%. For women, the college-high school graduation premium increased from 50% in 1979 to 72% in 1989, and then remained more or less constant throughout the remainder of the century.

Inequality also increased in the relative incomes of younger and older workers: the mean annual income premium for male workers aged 45–54 relative to those aged 25–34 rose from a ratio of 1.15 in 1979 to 1.27 in 1989, and then to 1.35 by 1995 (Brauer 1998). Such changes in income distribution in the United States have been described as an "economic disaster (that) has befallen low-skilled Americans, especially young men" (Freeman 1995). In the "social markets" of Europe, the adverse effects for people with low skills were felt more in terms of unemployment levels rather than reduced market incomes.

Trade liberalization is not the sole reason for decline in incomes and employment of unskilled workers in the richer countries. Technological change also reduced demand for unskilled workers. The liberalization of international trade occurred simultaneously with a technological revolution in information technologies that required complementary skills and education (Krueger 1993; Burtless 1995). There were also new standards of employee responsibility, since capricious or inept employee behavior became extremely costly for employers. The changes in income distribution are a consequence of influences of both trade liberalization and technology. An approach to identifying the contribution of trade liberalization is to look for possible relative price changes that would have given rise to the income changes. Yet reciprocal and multilateral liberalization through exchange of market access also neutralizes or dampens any terms of trade changes that would be associated with unilateral liberalization.

Unskilled workers were disadvantaged by incentives for the introduction of the new technologies associated with trade liberalization (Wood 1994). Through the new technologies, producers in high labor-cost countries sought ways to compete with low labor-cost foreign goods by substituting domestic unskilled labor with domestic skilled labor. Because domestic unskilled labor could not compete with the cheaper foreign unskilled labor used to produce lower-quality imports, a change to higher quality production also took place.

Demand for low-skill labor also declined in richer countries as producers responded to trade liberalization through outsourcing. Imports from foreign cheap-labor sources thereby replaced domestic production of intermediate goods that used low skilled labor, contracting employment opportunities and wages of local unskilled workers then declined. Evidence suggests that outsourcing explains some 20% of the substitution toward skilled non-production workers in the United States in the 1980s (Feenstra and Hansen 1996). Outsourcing takes place when trade liberalization allows foreign intermediate goods to be used in place of domestic production without a penalty for use of foreign goods.

Opposition to Globalization

Low-skilled persons in richer countries have self-interested reasons to oppose globalization. Although low skilled, these people have had expectations of a standard of living that is due to them because they live in a country that is on an

aggregate, wealthy. After globalization, the principal difference between being low skilled in rich and poor countries is in the role of the state in richer countries as a provider of last resort (Rodrik 1998).

Trade liberalization and outsourcing increase incomes in the poorer countries of the world. Opponents of globalization have however pointed to adverse effects in the poorer countries through labor standards and the environment (Bhagwati and Hudec 1996; Anderson 1997). Where child labor is a social norm, trade liberalization increases the demand for child labor because of improved foreign market access for goods produced by children. The opposition to child labor as a matter of social conscience then becomes opposition to globalization (or trade liberalization).

Domestic industries in richer countries may confront competition from foreign goods produced by children or foreign investment may take place to take advantage of the low costs of using child labor. Higher labor standards in poor countries benefit domestic producers and workers in rich countries by increasing production costs in poor countries. Again economic and humanitarian objectives become intertwined (Hefeker and Wunner 2002). Protectionist policies in richer countries would also protect producers and workers in richer countries and protect foreign workers from low foreign labor standards by denying market access for the foreign produced goods. If foreign labor standards cannot be changed, foreign labor standards become the basis for a case for protection to keep the foreign goods produced with the foreign labor standards out of the markets of richer countries (Agel and Lundborg 1995). Coalitions composed of producers and groups seeking social policies can form to oppose globalization (Hillman and Ursprung 1992). Producers and workers seeking protection from import competition can be bed-fellows (wanted or unwanted) with groups seeking protection of the environment and seeking to improve working conditions and end child labor in poor countries. Opposition to globalization has been particularly fierce when multinational firms, whose owners are principally in richer countries, use child labor or employ foreign labor at prevailing market wages and conditions of the poorer countries.

Local employers in poorer countries have not been subject to the same scrutiny as multinational firms. Nor have local restrictions in poorer countries on the freedom and rights of women, nor practices such as genital mutilation of pubescent female children, been reasons for outrage with the opponents of globalization. Opponents of globalization have not protested the corrupt behavior of rulers in poorer countries and the policies that keep the poor in poor countries in sustained poverty. Nor is the observation heralded that the highest inequality is in the poorer countries of the world where the political elites rule the poor. The opponents of globalization have an agenda that blames poverty in poor countries on open world markets, rather than calling for change in the behavior of political elites in poorer countries who sustain poverty by failing to use aid resources to improve living conditions of the general population (Rowley 2000; Easterly 1991; Hillman 2002).

The environment is also an issue for opponents of globalization. Clearing of rain forests and destruction of animals and their habitats is linked to demand in international markets. The opponents of globalization blame open markets, rather than confronting the foreign political elites who sell rights to deforestation.

Preferences regarding labor standards, sending children to school and calling for environmental quality can reflect income differences rather than cultural attributes. As incomes increase in poorer countries, preferences can consequently be expected to become more uniform over time. Globalization furthers this objective by increasing incomes in poorer countries through market integration with richer countries, and is therefore pro-environment (Grossman and Kreuger 1993). Political rulers may however not be responsive to the preferences and wishes of their citizens who seek to end child labor and end clearing of rain forests and destruction of animal habitats. The source of the problems that concern social activists is then again not globalization, but governments in poor countries that do not provide adequate resources for schools and continue environmental degradation for personal profit. It is interesting that the opponents of globalization have preferred to blame markets rather than blaming governments and political elites in poorer countries for their exploitative labor standards and damaging environmental policies.

References

Agel, J. and Lundborg, P. (1995). "Fair wages in an open economy." *Economica*, 62, 335–351.

Anderson, K. (1997). "Social policy dimensions of economic integration: environmental and labor standards," in I. Takatoshi and A.O. Krueger (eds.) *Regionalism versus Multilateral Trade Arrangements*. Chicago: The University of Chicago Press.

Bagwell, K. and Staiger, R.W. (1999). "An economic theory of GATT." *American Economic Review*, 89, 215–248.

Bagwell, K. and Staiger, R.W. (2001). "Reciprocity, non-discrimination and preferential agreements in the multilateral trading system." *European Journal of Political Economy*, 17, 281–325.

Bhagwati, J.N. and Hudec, R.E. (eds.) (1996). *Fair Trade and Harmonization: Prerequisites for Free Trade*. Cambridge, MA: MIT Press.

Brauer, D. (1998). "The changing US income distribution: facts, explanations, and unresolved issues." Working Paper No. 9811, Federal Reserve Bank of New York.

Burtless, G. (1995). "International trade and the rise in earnings inequality." *Journal of Economic Literature*, 33, 800–816.

Cassing, J.H. (1996). "Protectionist mutual funds." *European Journal of Political Economy*, 12, 1–18.

Easterly, W. (1991). *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. Cambridge, MA: MIT Press.

Ethier, W.J. (1998a). "Regionalism in a multilateral world." *Journal of Political Economy*, 106, 1214–1245.

Ethier, W.J. (1998b). "The new regionalism." *Economic Journal*, 108, 1149–1161.

Ethier, W.J. (2001). "Theoretical problems in negotiating trade liberalization." *European Journal of Political Economy*, 17, 209–232.

Ethier, W.J. (2002). "Unilateralism in a multilateral world." *Economic Journal*, 112, 266–292.

Feeney, J. and Hillman, A.L. (2001). "Privatization and the political economy of strategic trade policy." *International Economic Review*, 42, 535–556.

Feenstra, R. and Hansen, G. (1996). "Globalization, outsourcing, and wage inequality." *American Economic Review* (Papers and Proceedings), 86, 240–245.

Freeman, R.B. (1995). "Are your wages being set in Beijing?" *Journal of Economic Perspectives*, 9, 15–32.

Grossman, G.M. and Krueger, A.B. (1993). "Environmental impacts of a North-American free trade agreement," in P. Garber (ed.) *The Mexican—U.S. Free Trade Agreement*. Cambridge, MA: MIT Press.

Hefeker, C. and Wunner, N. (2002). "The producer interest in foreign labor standards." *European Journal of Political Economy*, 18, 429–447.

Hillman, A.L. (1989). "Resolving the puzzle of welfare reducing trade diversion: a prisoners' dilemma interpretation." *Oxford Economic Papers*, 41, 452–455.

Hillman, A.L. (2002). "The World Bank and the persistence of poverty in poor countries." *European Journal of Political Economy*, 18, 783–795.

Hillman, A.L., Long, N.V., and Moser, P. (1995). "Modeling reciprocal trade liberalization: the political-economy and national-welfare perspectives." *Swiss Journal of Economics and Statistics*, 131, 503–515.

Hillman, A.L. and Moser, P. (1996). "Trade liberalization as politically optimal exchange of market access," in M. Canzoneri, W.J. Ethier, and V. Grilli (eds.) *The New Transatlantic Economy*. New York: Cambridge University Press, pp. 295–312; reprinted in K. Anderson and B. Hoekman (eds.) *The Global Trading System*, Volume 2, *Core Rules and Procedures* Tauris, London, (2002).

Hillman, A.L. and Ursprung, H.W. (1992). "The influence of environmental concerns on the political determination of international trade policy," in R. Blackhurst and K. Anderson (eds.) *The Greening of World Trade Issues*. Brookfield, Wheatsheaf, pp. 195–220.

Horn, H. and Mavriodis, P. (2001). "Economic and legal aspects of the most-favored-nation clause." *European Journal of Political Economy*, 17, 233–279.

Irwin, D. (1989). "Political economy and Peel's repeal of the Corn laws." *Economics and Politics*, 1, 41–59.

Kemp, M.C. and Wan, H. (1976). "An elementary proposition concerning the formation of customs unions." *Journal of International Economics*, 6, 95–97; reprinted in M.C. Kemp (ed.) *The Gains From Trade and the Gains From Trade: Essays in International Trade Theory*. London: Routledge, 1995, pp. 37–40.

Krueger, A.B. (1993). "How computers have changed the wage structure: evidence from microdata 1984–1989." *Quarterly Journal of Economics*, 107, 33–60.

Levy, P.I. (1997). "The political economy of free-trade agreements." *American Economic Review*, 87, 506–519.

Ludema, R.D. (2001). "Optimal international trade agreements and dispute settlement procedures." *European Journal of Political Economy*, 17, 355–376.

Pomfret, R. (1988). Unequal Trade: The Economics of Discriminatory Trade Agreements. Oxford: Basil.

Ray, E. (1987). "The impact of special interests on preferential trade concessions by the U.S." *Review of Economics and Statistics*, 69, 187–195.

Richardson, M. (1993). "Endogenous protection and trade diversion." *Journal of International Economics*, 34, 309–324.

Rodrik, D. (1998). "Why do more open economies have bigger governments?" *Journal of Political Economy*, 106, 997–1032.

Rowley, C. (2000). "Political culture and economic performance in sub-Saharan Africa." *European Journal of Political Economy*, 16, 133–158.

Schonhardt-Bailey, C. (1991). "Specific factors, capital markets, portfolio diversification, and free trade: domestic determinants of repeal of the Corn laws." *World Politics*, 43, 545–569.

Schonhardt-Bailey, C. (1996). *Free Trade: The Repeal of the Corn Laws*. Bristol: Thoemmes Press.

Schonhardt-Bailey, C. (1997). *The Rise of Free Trade*, Volumes 1–4. London: Routledge.

Viner, J. (1950). *The Customs Union Issue*. New York: Carnegie Endowment for International Peace.

Wood, A. (1994). *North-South Trade, Employment, and Inequality: Changing Fortunes in a Skill-Driven World*. Oxford: Clarendon Press.

Chapter 28
Shadow Economy[1]

Friedrich Schneider

Introduction

Crime and shadow economic activities are a fact of life around the world, and almost all societies engage in trying to control these activities through education, punishment, or prosecution. Gathering statistics about who is active in the shadow economy activities, the frequency with which underground activities occur and the magnitude of these activities is crucial for making effective and efficient decisions regarding allocating resources in this area. Obviously it is difficult to get accurate information about underground or shadow economy activities because individuals engaged in these activities wish to remain unidentified. Hence, estimation of shadow economy activities can be considered a scientific passion for knowing the unknowable.

The attempts at measurement are obviously problematic,[2] since shadow economy activities are performed in such a way as to avoid any official detection. Moreover, if you ask an academic, a public sector specialist, a policy or economic analyst, or a politician, what is going on in the shadow economy, and even just how big it is, you will get a wide range of answers.[3] In spite of this, there is growing concern over the phenomenon of shadow economy, and there are several important reasons why politicians and public sector workers should be especially worried about its rise and growth.

Among the most important of these are:

- If an increase of the shadow economy is caused mainly by a rise in the overall tax and social security burden, this may lead to an erosion of the tax and social security bases and finally to a decrease in tax receipts, and thus to a further increase in the

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 286–296.

[2] Compare e.g. feature "Controversy: on the hidden economy," in the Economic Journal, Vol. 109, No. 456, June 1999, and also Schneider and Enste (2000) and Schneider (2005, 2007).

[3] Compare the different opinions of Tanzi (1999), Thomas (1999), Giles (1999a, b), and Pedersen (2003).

budget deficit or to a further increase of tax rates with the consequence of an additional increase in the shadow economy, and so on. Therefore, a growing shadow economy can be seen as a reaction by individuals who feel overburdened by state activities.

- With a growing shadow economy, (economic) policy is based on erroneous "official" indicators (like unemployment, official labor force, income, consumption), or at least indicators that are inaccurate in magnitude. In such a situation, a prospering shadow economy may cause politicians severe difficulties because it provides unreliable official indicators, and the direction of intended policy measures may therefore be questionable.

- On one hand, a growing shadow economy may provide strong incentives to attract (domestic and foreign) workers away from the official economy. On the other hand, at least two-thirds of the income earned in shadow economy is immediately spent in the official economy[4] resulting in a considerable (positive) stimulating effect on the official economy.

What is Shadow Economy?

Studies trying to measure shadow economy first face the difficulty of defining it. For instance, one commonly used definition is that shadow economy includes all current economic activities which contribute to the officially calculated (or observed) Gross National Product.[5] However, Smith (1994, p. 18) defines it as "market-based production of goods and services, whether legal or illegal, that escapes detection in the official estimates of GDP." As these definitions leave open a lot of questions, Table 28.1 may be helpful for developing a better feeling for what could be a reasonable consensus definition of the legal and illegal underground or shadow economy.

From Table 28.1 it becomes clear that the shadow economy includes unreported income from the production of legal goods and services, either from monetary or barter transactions, hence, all economic activities which would generally be taxable were they reported to the tax authorities. In general, a precise definition seems quite difficult, if not impossible, as " Shadow economy develops all the time according to the 'principle of running water': it adjusts to changes in taxes, to sanctions from the tax authorities and to general moral attitudes, etc." (Mogensen et al., 1995, p. 5).[6] My survey does not focus on tax evasion or tax compliance. It rather serves as a

---

[4] This figure has been derived from polls of the German and Austrian population about the (effects of) the shadow economy. For further information, see Schneider (1998a, b, 2005). These polls also show that two-thirds of the value added produced in the shadow economy would not be produced in the official economy if the shadow economy did not exist.

[5] This definition is used, e.g., by Feige (1989, 1994), Frey and Pommerehne (1984), Schneider and Enste (2000), Pedersen (2003), Kazemier (2005), and Schneider (2001a, b, c, 2005, 2007).

[6] For a detailed discussion, see Frey and Pommerehne (1984), Feige (1989), Thomas (1992), Schneider (1986, 1994a, b, 1998a), and Schneider and Enste (2000).

Table 28.1 A Taxonomy of types of underground economic activities

Type of activity	Monetary transactions	Nonmonetary transactions
ILLEGAL ACTIVITIES	Trade in stolen goods; drug dealing and manufacturing; prostitution; gambling; smuggling, and fraud	Barter: drugs, stolen goods, smuggling etc. Produce or growing drugs for own use. Theft for own use.
	Tax Evasion	Tax avoidance
LEGAL ACTIVITIES	Unreported income from self-employment; Wages, salaries and assets from unreported work related to legal services and goods; Barter of legal services and goods	Employee discounts, fringe benefits; All do-it-yourself work and neighbor help

Structure of the table taken from Lippert and Walker (1997, p. 5), with additional remarks.

supplement to the recent survey of Andreoni et al. (1998, p. 819), who excluded shadow economy: "Unfortunately, there are many important issues that we do not have room to discuss, most notably the vast literature on the underground economy which exists in part as a means of evading taxes."[7]

Empirical Estimates

The Tables 28.2–28.4 serve to indicate approximate magnitudes of the size and development of the underground economy, defined as productive activities, i.e. using the narrow definition. Table 28.2 presents a rough comparison of the size of the underground economies relative to GNP for a selection of Western European countries, Japan and the United States for the early 2000s (2000–2003), using the DYMIMIC and currency demand approach.[8]

The South European countries (Greece, Italy) have an underground economy almost one third as large as the officially measured GNP, followed by Spain, Portugal and Belgium having a shadow economy between 19 and 23% of the (official) GNP. According to these estimates, the Scandinavian countries also have a sizeable unofficial economy (between 17 and 20% of GNP), which is attributed mainly to the high fiscal burden. The "central" European countries (Ireland, the Netherlands, France, Germany and Great Britain) have a smaller underground economy (between 12 and 15% of GNP) probably due to a lower fiscal burden and

[7] Compare also Feinstein (1999), who tries to close the gap between tax evasion and shadow economy research.

[8] The DYMIMIC (Dynamic Multiple Indicator and Multiple Causes) is a latent estimation approach used to estimate the coefficients driving the shadow economy (causes and indicators). The currency demand approach is used to overcome that disadvantage of the DYMIMIC approach to calculate the relative estimated size of the shadow economy in absolute ones.

Table 28.2 Size of the underground economy relative to GNP in various European countries, beginning of 2000s. Estimation based on the DYMIMIC and the currency demand approach

Country	Size of the shadow economy
Greece	26–29%
Italy	
Spain	
Portugal	19–23%
Belgium	
Sweden	
Norway	17–20%
Denmark	
Ireland	
France	
Netherlands	12–15%
Germany	
Great Britain	
Japan	
United States	8–11%
Austria	
Switzerland	

Source: Compiled from Schneider (2005, 2007).

moderate regulatory restrictions. Lower underground economies are estimated to exist in countries with relatively low public sectors (Japan, the United States and Switzerland), and comparatively high tax morale (United States, Switzerland).

Table 28.3 provides a rough comparison of the size of the underground economy relative to the official GNP for a selection of developing and transition economies for the beginning of the 2000s (2000–2003), using the DYMIMIC and currency demand approach. Some of these countries (Nigeria, Bolivia, Thailand) are estimated to have an underground sector greater than 50% of the size of the officially recorded GNP. In many countries the size is one quarter to one third of the "official" GNP. In Asian countries with a comparatively low public sector, high tax morale or high expected punishment (Hong Kong, Singapore) the underground economy is estimated to be similar to that in many "northern" European countries.

Often, transition economies are estimated to have substantial unofficial activities; the former Soviet Union countries especially have a shadow economy greater than 50% of the official GDP, and many former Eastern Block countries have around one quarter of the GNP. An exception is ex-Czechoslovakia (with the countries of Czech Republic and Slovakia) where according to these estimates the underground sector is between 7 and 11 percent of the GNP.

Table 28.4 reports estimates of the growth of the underground economy (relative to GNP) for selected Western countries and the United States over an extended period of time, using the DYMIMIC and currency demand approach.

The Scandinavian (Sweden, Norway, Denmark) and the German speaking countries (Germany, Austria) exhibit a sizeable increase of the underground economy within the 35 years (1960–1999) covered. But the countries with a low share in the

Table 28.3 Size of the underground economy relative to GNP in various developing and transition countries, beginning of the 2000s (2001–2003). Estimates based on the DYMIMIC and currency demand approach

Country developing countries	Size of the shadow economy
Africa	
Nigeria	50–65%
Egypt	
Tunisia	39–45%
Morocco	
Central and South America	
Peru	
Bolivia	40–60%
Panama	
Nicaragua	
Chile	
Costa Rica	
Venezuela	25–35%
Brazil	
Paraguay	
Columbia	
Asia	
Thailand	52%
Philippines	
Sri Lanka	38–50%
Malaysia	
South Korea	
Hong Kong	13%
Singapore	
Transition Economies	
Central Europe	
Hungary	24–28%
Bulgaria	
Poland	16–20%
Rumania	
Slovakia	7–11%
Czech Republic	
Former Soviet Union Countries	
Georgia	
Azerbaijan	50–65%
Ukraine	
Belarus	
Russia	
Lithunia	20–27%
Latvia	
Estonia	

Source: Compiled from Schneider (2005, 2007).

Table 28.4 Size and growth of the underground economy relative to GNP for selected West European countries and the United States, 1960–2005. Estimates based on the currency demand approach (rounded figures)

Country	Year 1960 (%)	Year 2005 (%)	Percentage point increase (%)
Sweden	2	16.3	14.3
Denmark	4.5	16.1	11.6
Norway	1.5	16.8	15.3
Germany	2	15.3	13.3
United States	3.5	7.9	4.4
Austria	0.5	9.3	8.8
Switzerland	1	8.5	7.5

Source: Compiled from Schneider (2005, 2007).

beginning (Switzerland, the United States) also show a significant increase; for the United States the share more than doubled. Sizeable increases have been estimated, with few exceptions, for all types of countries and all kinds of approaches: the increasing importance of the underground relative to the official economy is a robust phenomenon.

What are the Main Causes for the Increase in Shadow Economy?

Increase of the Tax and Social Security Contribution Burdens

In almost all studies,[9] the increase of the tax and social security contribution burdens is one of the main causes for the increase of the shadow economy. Since taxes affect labor—leisure choices and also stimulate labor supply in the shadow economy, or the untaxed sector of the economy, the distortion of this choice is a major concern of economists. The bigger the difference between the total cost of labor in the official economy and the after-tax earnings (from work), the greater is the incentive to avoid this difference and to participate in shadow economy. Since this difference depends broadly on the social security system and the overall tax burden, they are key features of the existence and the increase of shadow economy.

But even major tax reforms with major tax rate deductions will not lead to a substantial decrease of the shadow economy. They will only be able to stabilize the size of the shadow economy and avoid a further increase. Social networks and personal relationships, and the high profit from shadow economy activities and associated investments in real and human capital are strong ties which prevent people

---

[9] See Thomas (1992), Lippert and Walker (1997), Schneider (1994a, b, 1997, 1998a, b, 2001a, b, c, 2005, 2007), Schneider and Enste (2000, 2002), Johnson et al. (1998a, b), De Soto (1989); Zilberfarb (1986); Tanzi (1999), and Giles (1999a), just to quote a few recent ones.

from working in the shadow economy. For Canada, Spiro (1993) expected similar reactions from people facing an increase in indirect taxes (VAT, GST). After the introduction of the GST in 1991 in the midst of a recession, the individuals, suffering economic hardships because of the recession, turned to the shadow economy, leading to a substantial loss in tax revenue. "Unfortunately, once this habit develops, it is unlikely that it will be abandoned merely because economic growth resumes" (Spiro, 1993, p. 255). They may not return to the formal sector, even in the long run. This fact makes it even more difficult for politicians to carry out major reforms, because they may not gain a lot from them.[10]

The most important factor in neoclassical models is the marginal tax rate. The higher the marginal tax rate, the greater is the substitution effect and the bigger the distortion of the labor—leisure decision. Especially when taking into account that the individual can also receive income through the shadow economy, the substitution effect is definitely larger than the income effect[11] and, hence, the individual works less in the official sector. The overall efficiency of the economy is, therefore (*ceteris paribus*), lower and the distortion leads to a welfare loss (according to official GDP and taxation.) But the welfare might also be viewed as increasing, if the welfare of those who are working in the shadow economy were also taken into account.[12]

While there have been many theoretical studies on tax evasion in the last twenty years, empirical studies of tax evasion are hard to come by.[13] Most of them are based on tax compliance experiments and cover only some parts of the shadow economy.[14] Convincing empirical evidence for the theoretical hypothesis why people evade taxes is hard to find and the results are ambiguous (Pommerehne and Weck-Hannemann 1992). The results are more convincing for the shadow economy: for example, Schneider (1994a, b, 2007) and Johnson et al. (1998a, b) found strong evidence for the general influence of taxation on the shadow economy.

The strong influence of indirect and direct taxation on the shadow economy will be further demonstrated by showing empirical results in the case of Austria and the Scandinavian countries. In the case of Austria, Schneider (1994b) points out that as the driving force for shadow economy activities, the direct tax burden (including social security payments) has the biggest influence, followed by the intensity of regulation and complexity of the tax system.

---

[10] Schneider (1994b, 1998b), for a similar result of the effects of a major tax reform in Austria on the shadow economy. Schneider shows that a major reduction in the direct tax burden did not lead to a major reduction in the shadow economy. Because legal tax avoidance was abolished and other factors, like regulations, were not changed; hence, for a considerable part of the taxpayers, the actual tax and regulation burden remained unchanged.

[11] If leisure is assumed to be a normal good.

[12] See Thomas (1992), p. 134B7.

[13] For a broad survey, see Andreoni et al. (1998).

[14] See Alm (1996), for an overview of tax compliance explanations in different studies. The theoretical literature on tax evasion is summarized in Cowell (1990); see also Allingham and Sandmo (1972), for their path-breaking study in this area.

A similar result has been achieved by Schneider (1986) for the Scandinavian countries (Denmark, Norway, and Sweden). In all three countries, various tax variables (average direct tax rate, average total tax rate, indirect and direct tax rate and marginal tax rates) have the expected positive sign (on currency demand) and are highly statistically significant. Similar results are reached by Kirchgaessner (1983, 1984) for Germany, and by Klovland (1984) for Norway and Sweden.

Two other recent studies provide strong evidence of the influence of income taxes on shadow economy: Cebula (1997), using Feige's data for shadow economy, found evidence of the impact of government income tax rates, IRS audit probabilities, and IRS penalty policies on the relative size of the shadow economy in the United States. Cebula concludes that a restraint of any further increase of the top marginal income tax rate may at least not lead to a further increase of the shadow economy, while increased IRS audits and penalties might reduce the size of the shadow economy. His findings indicate that there is generally a strong influence of state activities on the size of the shadow economy: For example, if the marginal federal personal income tax rate increases by one percentage point, *ceteris paribus*, the shadow economy rose by 1.4 percentage points.

More detailed information about the labor supply decision in the underground economy is given by Lemieux et al. (1994), using micro data from a survey conducted in Quebec City, Canada. In particular, their study provides some economic insight into the size of the distortion caused by income taxation and the welfare system. The results of this study suggest that hours worked in the shadow economy are quite responsive to changes in the net wage in the regular (official) sector. It also provides some support for the existence of a Laffer curve. The Laffer curve suggests that an increase of the (marginal) tax rate leads to a decrease of tax revenue when the tax rate is too high. Their empirical results attribute this to an (mis-) allocation of work from the official to the informal sector, where it is not taxed. In this case, the substitution between labor market activities in the two sectors is quite high. These empirical findings clearly indicate that, "participation rates and hours worked in the underground sector also tend to be inversely related to the number of hours worked in the regular sector" (Lemieux et al., 1994, p. 235). The findings demonstrate a large negative elasticity of hours worked in the shadow economy with respect to the wage rate in the regular sector and also to a high mobility between the sectors.

In another investigation, Hill and Kabir (1996) found empirical evidence that marginal tax rates are more relevant than average tax rates, and that a substitution of direct taxes by indirect taxes seems unlikely to improve tax compliance. More evidence on the effect of taxation on the shadow economy is presented by Johnson et al. (1998b), who conclude that it is not higher tax rates *per se* that increase the size of the shadow economy but the ineffective and discretionary application of the tax system and the regulations by governments. Their finding that there is a *negative* correlation[15] between the size of the unofficial economy and the *top* (marginal) tax

---

[15] The higher the top marginal tax rate, the lower the size of the shadow economy.

rates might be unexpected. But since other factors like tax deductibility, tax reliefs, tax exemptions, the choice between different tax systems, and various other options for legal tax avoidance, were not taken into account, it is not all that surprising.[16] For example, hardly anybody is paying the top marginal tax rate in Germany, since there are many legal tax loopholes (of course, mostly used by wealthy people.)[17]

Johnson et al. (1998b) find a *positive* correlation between the size of the shadow economy and the corporate tax burden. They come to the overall conclusion that there is a large difference between the impact of direct taxes as compared to the corporate tax burden. Institutional aspects, like the efficiency of the administration, the extent of control rights held by politicians and bureaucrats, and the amount of bribery and especially corruption, therefore, play a major role in this "bargaining game" between the government and the taxpayers.

Intensity of Regulations

The increase of the intensity of regulations (often measured in the number of laws and regulations, like license requirements) is another important factor, which reduces the freedom (of choice) for individuals engaged in the official economy.[18] One can think of labor market regulations, trade barriers, and labor restrictions for foreigners. Although Johnson et al. (1998b) did not find overall significant empirical evidence of the influence of labor regulations on the shadow economy, the impact is clearly described and theoretically derived in other studies, for example, for Germany (Deregulation Commission 1990/91). Regulations lead to a substantial increase in labor costs in the official economy. But since most of these costs can be shifted to the employees, these costs provide another incentive to work in the shadow economy, where they can be avoided.

Empirical evidence supporting the model of Johnson et al. (1997), which predicts, *inter alia*, that countries with more general regulation of their economies tend to have a higher share of the unofficial economy in total GDP, is found in their analysis. A one point increase of the regulation index (ranging from 1 to 5, with 5 = the most regulation in a country), *ceteris paribus*, is associated with an 8.1

---

[16] Friedman et al. (1999) found a similar result in a cross-country analysis that higher tax rates are associated with less official activity as a percent of GDP. They argue entrepreneurs go underground not to avoid official taxes but they want to reduce the burden of bureaucracy and corruption. However, looking at their empirical (regression) results, the finding that higher tax rates are correlated with a lower share of the unofficial economy is not very robust and, in most cases, using different tax rates, they do not find a statistically significant result.

[17] See Enste (1997), for further details on the (postponed) major German tax reform.

[18] See, for a (social) psychological, theoretical foundation of this feature, Schmölders (1960, 1975), Brehm (1966, 1972); and for an (first) application to the shadow economy, see Pelzmann (1988).

percentage point increase in the share of the shadow economy, when controlled for GDP per capita (Johnson et al. (1998b), p. 18). They conclude that it is the enforcement of regulation that is the key factor for the burden levied on firms and individuals, and not the overall extent of regulation mostly not enforced that drive firms into the shadow economy. Friedman et al. (1999) reach a similar conclusion. In their study, every available measure of regulation is significantly correlated with the share of the unofficial economy and the sign of the relationship is unambiguous: more regulation is correlated with a larger shadow economy. A one point increase in an index of regulation (ranging from 1 to 5) is associated with a 10 percent increase in the shadow economy for 76 developing, transition, and developed countries.

These findings demonstrate that governments should put more emphasis on improving enforcement of laws and regulations, rather than increasing their number. Some governments, however, prefer this policy option (more regulations and laws), when trying to reduce the shadow economy, mostly because it leads to an increase in the power of the bureaucrats and to a higher rate of employment in the public sector.[19]

Social Transfers

The social welfare system leads to strong negative incentives for beneficiaries to work in the official economy since their marginal tax rate often equals or nearly reaches 100 percent. This can be derived either from the neoclassical leisure-income model or from empirical results.[20] Such a system provides major disincentives for individuals who are getting welfare payments to even search for work in the official economy, since their overall income is much higher when they are still receiving these transfers, while possibly working in the underground economy.

Labor Market

The numerous regulations in the official labor market and the total wage costs are also driving forces for the shadow economy. Two main aspects, the effects of the reduction in official working hours and the influence of the unemployment rate on the increase of the shadow economy are discussed quite often in this context:

- As in most OECD countries, unemployment is, to a large extent, caused by the fact that total labor costs are too high. This can be seen as a cause for an increase of the shadow economy.

---

[19] See, e.g., Frey (1989), for a first application of the Public Choice Theory to the shadow economy.

[20] See, e.g., Lemieux et al. (1994).

- The reduction in working hours in the official economy was introduced by governments (e.g., France) and/or labor unions (e.g., Germany) in order to reduce the unemployment rate. The idea behind this is that there is only a limited quantity of work, and that this quantity has to be "redistributed." But this idea neglects a key factor that especially a forced reduction (but an increase in flexibility of working hours, too) increases the potential of hours that can be worked in the shadow economy.[21] Early retirements can also lead to more unofficial activities and part-time work offers great opportunities to the individual to adopt another job in the untaxed, unregulated economy, as argued by de Gijsel (1984) and Riebel (1983, 1984).[22]

Public Sector Services

An increase in the shadow economy leads to reduced state revenues, which in turn reduces the quality and quantity of publicly provided goods and services. Ultimately, this can lead to an increase in the tax rates for firms and individuals in the official sector, quite often combined with a deterioration in the quality of public goods (such as the public infrastructure) and of the administration, with the consequence of even stronger incentives to participate in the shadow economy. Johnson et al. (1998b) present a simple model of this relationship. Their findings show that smaller shadow economies appear in countries with higher tax revenues, than those with lower tax rates, fewer laws and regulations, and less bribery facing enterprises. Countries with a better rule of the law, which is financed by tax revenues, also have smaller shadow economies. Transition countries have higher levels of regulation, leading to a significantly higher incidence of bribery, higher effective taxes on official activities, a large discretionary framework of regulations, and, consequently, to a higher shadow economy.

The overall conclusion is that "wealthier countries of the OECD, as well as some in Eastern Europe find themselves in the 'good equilibrium' of relatively low tax and regulatory burden, sizeable revenue mobilization, good rule of law and corruption control, and [relatively] small unofficial economy. By contrast, a number of countries in Latin America and the former Soviet Union exhibit characteristics consistent with a 'bad equilibrium': tax and regulatory discretion and burdens on firms are high, the rule of law is weak, and there is a high incidence of bribery and a relatively high share of activities in the unofficial economy." (Johnson et al., 1998a, p. I).

---

[21] After Volkswagen in Germany reduced the working hours considerably, there is some evidence that in the area around the firm, much more reconstruction and renovation of houses took place compared to similar other regions.

[22] See Becker (1965), Trockel (1987), and Werner (1990), for a more detailed analysis.

The Effects of the Shadow Economy on the Official Economy

In order to study the effects of the shadow economy on the official one, several studies integrate underground economies into macroeconomic models.[23] Houston (1987) develops a theoretical macro model of business cycle as well as tax and monetary policy linkages with the shadow economy. He concludes from his investigation of the growth of the shadow economy that, on one side its effect should be taken into account in setting tax and regulatory policies, and, on the other side, the existence of a shadow economy could lead to an overstatement of the inflationary effects of fiscal or monetary stimulus. Adam and Ginsburgh (1985) focus on the implications of the shadow economy on "official" growth in their study concerning Belgium. They find a positive relationship between the growth of the shadow economy and the "official" one and, under certain assumptions (i.e., very low entry costs into the shadow economy due to a low probability of enforcement), they conclude that an expansionary fiscal policy has a positive stimulus for both the formal and informal economies. A study of the United States by Fichtenbaum (1989) argues that the United States' productivity slowdown over the period 1970–89 was vastly overstated, as the underreporting of income due to the more rapid growth of the United States shadow economy during this period was *not* taken into account.[24]

Another hypothesis is that a substantial reduction of the shadow economy leads to a significant increase in tax revenues and therefore to a greater quantity and quality of public goods and services, which ultimately can stimulate economic growth. Some authors found evidence for this hypothesis. A recent study by Loayza (1996) presents a simple macroeconomic endogenous growth model whose production technology depends on congestible public services. The determinants and effects of the informal sector are studied, where excessive taxes and regulations are imposed by governments and where the capability to enforce compliance is low. The model concludes that in economies where (1) the statutory tax burden is larger than the optimal tax burden, and where (2) the enforcement of compliance is too weak, the increase of the relative size of the informal economy generates a reduction of economic growth. The reason for this correlation is the strongly negative correlation between the informal sector and public infrastructure indices, while public infrastructure is the key element for economic growth. For example, Loayza finds empirical evidence for Latin American countries where if the shadow economy increases by one percentage point of GDP *ceteris paribus*, the growth rate of official real GDP per capita decreases by 1.22 percentage points of GDP.

---

[23] For Austria, this was done by Schneider et al. (1989), and Neck et al. (1989). For further discussion of this aspect, see Quirk (1996), Giles (1999a), and Schneider (2005).

[24] Compare also the findings of Pommerehne and Schneider (1985) who come to similar conclusions.

This negative impact of informal sector activities on economic growth is not broadly accepted.[25] For example, the key feature of the model has been criticized, because the model is based on the assumption that the production technology essentially depends on tax-financed public services, which are subject to congestion. In addition, the informal sector is not paying any taxes but must pay penalties which are not used to finance public services. Based on these assumptions the negative correlation between the size of the informal sector and economic growth is therefore not very surprising.

Depending on the prevailing view of the informal sector, one might also come to the opposite conclusion. In the neoclassical view, the underground economy is optimal in the sense that it responds to the economic environment's demand for urban services and small-scale manufacturing. From this point of view, the informal sector provides the economy with a dynamic and entrepreneurial spirit and can lead to more competition, higher efficiency and strong boundaries and limits for government activities. The informal sector may offer great contributions "to the creation of markets, increase financial resources, enhance entrepreneurship, and transform the legal, social, and economic institutions necessary for accumulation" (Asea, 1996, p. 166). The voluntary self-selection between the formal and informal sectors, as described above in microeconomic models, may provide a higher potential for economic growth and, hence, a positive correlation between an increase of the informal sector and economic growth. The effects of an increase of the shadow economy on economic growth therefore remain considerably ambiguous.

The empirical evidence of these hypotheses is also not clear. Since many Latin American countries had or still have a tradition of excessive regulations and weak government institutions, Loayza (1996) finds some evidence of the implications of his growth model in the early 1990s in these countries: the increase in the size of the shadow economy negatively affects growth (1) by reducing the availability of public services for everyone in the economy, and (2) by using the existing public services less efficiently, or not at all.

On the other side, the positive "side effects" of shadow economy activities must be considered. Empirical findings of Schneider (1998b) show clearly that over 66 percent of the earnings in the shadow economy are rather immediately spent in the official sector. The positive effects of this expenditure for economic growth and for the (indirect) tax revenues must be taken into account as well. Bhattacharyya (1993, 1999) found clear evidence for the United Kingdom (1960–84) that the hidden economy has a significant effect on consumer expenditures. He points out that the hidden economy has a positive effect on consumer expenditures of nondurable goods and services, but an even stronger positive effect on consumer expenditures of durable goods and services.[26]

---

[25] See Asea (1996), for a more detailed criticism of the Loayza model. Also, compare the findings of Schneider (2005), who demonstrates that the negative effect prevails in developing countries, but a positive influence of the shadow economy on the official one for highly developed and transition countries shows up!

[26] A close interaction between official and unofficial economies is also emphasized in Giles (1999a), and in Tanzi (1999).

Corruption and the Shadow Economy: Substitutive or Complementary Effects?

Over the last 10 years, corruption has gained growing attention among scientists, politicians, and public officials regarding its origins, consequences, and ways to fight it.[27] Corruption has been defined in many different ways but "the most popular and simplest definition of corruption is that it is the abuse of public power for private benefit" (Tanzi, 1998, p. 8). From this definition the private sector seems to be excluded, which is, of course, not the case; a more general definition is "that corruption is the intentional non-compliance with arm's length relationship from this behavior for oneself or for related individuals" (Tanzi, 1998, p. 8). There are various kinds of corruption including cost reductions in response to bribes and cash payments, and there is an extensive literature about which factors stimulate corruption.[28] Activities, in which corruption is sometimes involved, include:

- regulations or licenses to engage in particular activities (e.g., opening a shop, a taxi license);
- land zoning and other similar official decisions;
- access to publicly provided goods and services;
- control over decision making regarding procurement of public investment contracts;
- control over the provision of tax incentives; and
- control over hiring and promotion within the public sector.

The effect of corruption on the official economy can be seen from different sides: Romer (1994) has suggested that corruption, as a tax on ex-post profits, may in general stimulate the entry of new goods or technologies, which require an initial fixed-cost investment. Mauro (1995) finds a significant negative correlation between the corruption index and the investment rate or rate of GDP growth. A one-standard-deviation improvement in the corruption index is estimated by Mauro to increase the investment rate by about 3 percent. Johnson et al. (1998b, p. 39) find a significant relationship between corruption and GDP growth (an increase in corruption on an indexed scale from 0 to 6 by only 1 point decreases GDP growth by 0.84 percentage points) but the relationship becomes insignificant if the shadow economy is entered as an independent variable. On the other side, Bardhan (1997, p. 1329) concludes that "it is probably correct to say that the process of economic growth ultimately generates enough forces to reduce corruption" a view supported by Rose-Ackermann (1997, 2006), who further argues that any reform that increases the competitiveness of the economy will help reduce incentives for corruption. Thus, policies that liberalize foreign trade

---

[27] The literature is quite large and only some of it (mostly more recent) is given here: Rose-Ackermann (1978, 1997, 1999, 2002, 2004, 2006), Shleifer and Vishny (1993), Tanzi (1994, 1997, 1998), Johnson et al. (1998a,b), and Kaufmann and Sachs (1998). For the latest survey, see Bardhan (1997), Jain (1998), and Rose-Ackermann (2006).

[28] See, e.g., Rose-Ackermann (1997, 1999, 2006), Jain (1998), Tanzi (1998), and Bardhan (1997).

and remove entry barriers for industry promote competition and reduce corruption. Such reforms will also encourage firms to move from the shadow economy into the official economy, where they can obtain access to capital at market rates. Rose-Ackermann (1997, p. 21) concludes that "going underground is a substitute for bribery, although sometimes firms bribe officials in order to avoid the official states." In a similar way Choi and Thum (2004) present a model where the option of entrepreneurs to go underground constrains a corrupt official's ability to ask for bribes. Dreher, Kotsogiannis and McCorriston (2005a/b) extend the model to the explicit specification of institutional quality. The model shows that corruption and shadow economy are substitutes in the sense that the existence of the shadow economy reduces the propensity of officials to demand graft.

Johnson et al. (1998b), on the contrary, model corruption and the shadow economy as complements. In their full-employment model, labour can be either employed in the official sector or in the underground economy. Consequently, an increase in the shadow economy always decreases the size of the official market. In their model, corruption increases the shadow economy, as corruption can be viewed as one particular form of taxation and regulation (driving entrepreneurs underground). Hindriks et al. (1999) also show that the shadow economy is a complement to corruption. This is because, in this case, the tax payer colludes with the inspector so the inspector under-reports the tax liability of the tax payer in exchange for a bribe[29]).

Theoretically, the relationship between corruption and the shadow economy is thus unsettled. There is, however, reason to believe that the relationship might differ among high and low income countries. In high income countries, the official sector provides public goods like the rule of law, enforcement of contracts, and protection by an efficient police. Usually, only craftsmen or very small firms have (or take) the option of going underground. In this case, the shadow economy is hidden from tax inspectors and other officials. In other words, there are no bribes necessary or possible to buy the way out of the official sector. In high income countries—typically showing comparably small levels of corruption—individuals confronted with a corrupt official always have the choice to bring the official to court. Moreover, in high income countries corruption quite often takes place, for example, to bribe officials to get a (huge) contract from the public sector (e.g. in the construction sector). This contract is then handled in the official economy and not in the shadow economy. Hence, corruption in high income countries can be a means to achieve certain benefits which make work in the official economy easier, e.g., winning a contract from a public authority, getting a license (e.g. for operating taxes or providing other services or getting the permission to convert land into "construction ready" land, etc.). In high income countries people thus bribe in order to be able to engage in more official economic activities. As Schneider and Enste (2000) point out, at least two thirds of the income earned in the shadow economy is immediately spent in the official sector. The shadow economy and the official sector might thus be

---

[29] See Dreher and Siemers (2005) for a formalization of this argument.

complements. The corresponding increase in government revenue and strengthened institutional quality is likely to decrease corruption. The prediction of a negative (substitutive) relation between corruption and the shadow economy is in line with the models of Choi and Thum (2004) and Dreher, Kotsogiannis and McCorriston (2005).[30]

In low income countries, on the contrary, I expect different mechanisms to prevail. Instead of working partly in the official sector and offering additional services underground as in high-income countries, enterprises completely engage in underground activity. Examples for enterprises operating completely underground are restaurants, bars, or haircutters—and even big production companies. One reason for this is that public goods provided by the official sector are, in many developing countries, less efficient compared to high income countries. Big companies, however, are comparably easy to detect and—in order to escape taxation and punishment— they have to bribe officials, thereby increasing corruption. Corruption often takes place in order to pay for activities in the shadow economy, so that the shadow economy entrepreneur can be sure not to be detected by public authorities. Here, shadow economy and corruption are likely to reinforce each other, as corruption is needed to expand shadow economy activities and—at the same time—underground activities require bribes and corruption. To get some additional income from the shadow economy entrepreneur, it is natural for public officials to ask for bribes and thus benefit from the shadow market. In low income countries, we therefore expect a positive (complementary) relationship between corruption and the shadow economy. This corresponds to the predictions of the models of Hindriks et al. (1999) and Johnson et al. (1997).

In summary, following Dreher and Schneider (2006), I expect:

Hypothesis 1: In low income countries, shadow economy activities and corruption are complements.
Hypothesis 2: In high income countries, shadow economy activities and corruption are substitutes.

These two hypotheses are tested for a cross-section of 120 countries and a panel of 70 countries for the period 1994 to 2002[31] in a paper by Dreher and Schneider (2006). Overall, they show that an increase in perceived corruption over time also increases the shadow economy. This confirms the models of Johnson et al. (1998) and Hindriks et al. (1999). Across countries, however, greater perceived corruption does not lead to a greater shadow economy. To some extent this also supports the results of Méon and Sekkat (2004) showing the within-country variation to be important in their analysis of corruption on foreign direct investment and exports.

---

[30] Consequently, Dreher, Kotsogiannis and McCorriston (2005a) test their model employing data for OECD countries only.

[31] For the description of the data, the estimation techniques used, and the various specification see Dreher and Schneider (2006, chapters 3 and 4).

Regarding the impact of the shadow economy on perceived corruption, these results for the overall sample are similar to those for the other way round. In the cross-country regressions, all coefficients are completely insignificant. An increase in the shadow economy over time increases corruption according to the fixed and random effects estimator, but not when the endogeneity of the shadow is controlled. Turning to the sub-samples, the results show that higher perceived corruption significantly reduces the shadow economy in high income countries, confirming the models of Choi and Thum (2004) and Dreher, Kotsogiannis and McCorriston (2005). In low income countries, on the contrary, corruption tends to increase with a higher shadow economy, again confirming the models of Johnson et al. (1998a) and Hindriks et al. (1999). This is true for the impact of perceived corruption in the within-groups specification and actual corruption in all specifications.

There are only a few other studies which empirically investigate the relationship between the shadow economy and corruption, either in a country or over a sample of countries.[32] Johnson et al. (1998b, p. 21) find, in their empirical investigation of 49 countries of Latin America, the OECD, and the post-communist countries of Eastern Europe and the former Soviet Union, a statistically highly significant relationship between the various measures of bribery or corruption and the shadow economy; a 1 point improvement (= less corruption) in the corruption index ICRG[33] leads to about an 8–11 percentage point decline in the shadow economy, ceteris paribus. Using another measure for corruption, the transparency International Corruption Index,[34] Johnson et al. found that a 1 point increase in this index (= less corruption) decreases the shadow economy by 5.1 percentage points, ceteris paribus. Friedman et al. (1999, p. 27) conclude: "... In summary, the relationship between the share of the unofficial economy and rule of law (including corruption) is strong and consistent across eight measures provided by six distinct organizations. All eight of the indices suggest that countries with more corruption have a higher share of the unofficial economy." In their investigation, they show that a one point increase in the index of corruption increased the share of the unofficial economy by 7.6 percentage points in the year 1997.

To summarize, Dreher and Schneider (2006) find that the relationship between the share (size) of the shadow economy and the amount of corruption is strong, as different measures show. Low (high) income countries with more corruption and briberies have a higher (lower) share (size) of the shadow economy. Whereas Rose-Ackermann concludes from her work that going underground is a *substitute* for corruption (bribery), the empirical results of Johnson et al. point more to a complementary process: *countries with more corruption, ceteris paribus, have higher shares of the shadow economy*. Hence, more research is needed to settle the so-far contradicting results.

---

[32] See, e.g., Johnson et al. (1998a, b), Johnson et al. (1997), and Kaufmann and Sachs (1998).

[33] This index ranks between 1 and 6 (best = no corruption), and was averaged by Johnson et al. (1998b, p. 21) for the 1990s.

[34] This index ranks between 0 and 10 (best = no corruption).

Summary and Conclusions

There are many obstacles to be overcome in measuring the size of the shadow economy and analyzing its consequences for the official economy. In this paper, it is shown that although it is difficult to estimate the size of the shadow economy, it is not impossible. I have demonstrated that with various methods (e.g., the currency demand, the physical input measure, and the model approach), some insights can be provided into the size and development of the shadow economy of developing, transition, and the OECD countries. The general impression from the results of these methods is that, for all countries investigated, the shadow economy has reached a remarkably large size. There is another common finding that the size of the shadow economy in most transition and all investigated OECD countries has been growing up to the end of the nineties and is staggering or shrinking since then. Furthermore, the results in this essay show that an increasing burden of taxation and social security payments, combined with rising state regulatory activities, are the major driving forces behind the size and growth of the shadow economy. Considering a public choice perspective, I draw the conclusion for highly developed OECD countries that a government may not have great interest to reduce the shadow economy due to the fact that:

   (i) tax losses my be moderate, as at least two-thirds of the income earned in the shadow economy is immediately spent in the official economy,
  (ii) income earned in the shadow economy increases the standard of living of at least one third of the working population, and
 (iii) between 40 and 50% of the shadow economy activities have a complementary character, which means that more value addition is created, which increases the official (overall) GDP, and
 (iv) people who work in the shadow economy have less time for other things like going to demonstrations, etc.

According to some studies, a growing shadow economy has a negative impact on official GDP growth and is linked to the amount of corruption.

   To conclude: shadow economies are a complex phenomenon, present to an important extent even in the most industrialized and developed economies. People engage in shadow economic activity for a variety of reasons; among the most important, as far as I can tell, are government actions, most notably taxation and regulation. Along with these considerations goes a third, no less important one: a government aiming to reduce shadow economic activity has to first and foremost analyze the complex and frequently contradictory relationships that are among the consequences of its own policy decisions.

References

Adam, M. C. and V. Ginsburgh. (1985). "The effects of irregular markets on macroeconomic policy: Some estimates for Belgium." *European Economic Review*, 29(1), 15–33.

Allingham, M. G. and A. Sandmo. (1972). "Income tax evasion: A theoretical analysis." *Journal of Public Economics*, 1(3–4), 323–338.

Alm, J. (1996). "Explaining tax compliance," in Pozo, Susan (eds.) *Exploring the Underground Economy*. Michigan: Kalamazoo, pp. 103–127.

Andreoni, J., E. Brian, and J. Feinstein. (1998). "Tax compliance." *Journal of Economic Literature*, 36, 818–860.

Asea, P. K. (1996). "The informal sector: Baby or bath water?" *Carnegie-Rochester Conference Series on Public Policy*, 45, 163–171.

Bardhan, P. (1997). "Corruption and development: a review of issues." *Journal of Economic Literature*, 35, 1320–1346.

Becker, G. S. (1965). "A theory of the allocation of time." *The Economic Journal*, 75(299), 493–517.

Bhattacharyya, D. K. (1993). *How Does the "Hidden Economy" Affect Consumers Expenditure? An Econometric Study of the U.K. (1960–1984)*. Berlin: International Institute of Public Finance (IIPF).

Bhattacharyya, D. K. (1999). "On the economic rationale of estimating the hidden economy." *The Economic Journal*, 109, 348–359.

Brehm, J. W. (1966). *A Theory of Psychological Reactance*. New York: Academic Press.

Brehm, J. W. (1972). Responses to Loss of Freedom- A Theory of Psychological Reactance. Morristown: General Learning Press.

Cebula, R. J. (1997). "An empirical analysis of the impact of government tax and auditing policies on the size of the underground economy: The case of the United States, 1993–94." *American Journal of Economics and Sociology*, 56(2), 173–185.

Choi, J. and M. Thum. (2004). "Corruption and the shadow economy", *International Economic Review*, 12(4), 308–342.

Cowell, F. (1990). *Cheating the Government*. Cambridge, MA: MIT Press.

de Gijsel, P. (1984). "Ökonomische Theorie des schwarzarbeitsangebots und der Mehrfach-beschäftigung," in Gretschmann, K., Heinze, R. G., Mettelsiefen, B. (Hrsg.) *Schattenwirtschaft. Wirtschafts- und sozialwissenschaftliche Aspekte, internationale Erfahrungen*. Göttingen: Vandenhoeck und Rubrecht, pp. 76–96.

Dreher, A., C. Kotsogiannis, and S. McCorriston. (2005). "How do institutions affect corruption and the shadow economy?" *Discussion Paper*, Department of Economics, University of Konstanz, Germany.

Dreher, A. and F. Schneider. (2006). "Corruption and the shadow economy: An empirical analysis", *Discussion Paper*, Department of Economics, University of Linz, Linz, Austria.

Dreher, A. and L. H. R. Siemers. (2005). "The intriguing nexus between corruption and capital account restrictions", *KOF Working Paper 113*, Swiss Institute of Technology (ETH Zürich), Switzerland.

Enste, D. H. (1997). "Ökonomische Wirkungsanalyse der Einkommensteuerreform. Grundlegende Ziele, Wirkungen und Gegenfinanzierungsmöglichkeiten," in Institut für Wohnungsrecht und Wohnungswirtschaft an der Universität zu Köln (Hrsg.) *Ökonomische Wirkungsanalyse der Einkommensteuerreform*. Köln. INWO 14, 9–25.

Feige, E. L. (ed.) (1989). *The Underground Economies. Tax Evasion and Information Distortion*. Cambridge: Cambridge University Press.

Feige, E. L. (1994). "The underground economy and the currency enigma." *Supplement to Public Finance/Finances Publiques*, 49, 119–136.

Feinstein, J. S. (1999). "Approaches for estimating non-compliance: Examples from federal taxation in the United States." *The Economic Journal*, 109(456), 360–369.

Fichtenbaum, R. (1989), "The productivity slowdown and the underground economy." *Quarterly Journal of Business and Economics*, 28(3), 78–90.

Frey, B. S. (1989). "How large (or small) should the underground economy be?" in Feige, E. (ed.) *The Underground Economies: Tax Evasion and Information Distortion*. New York: Cambridge University Press, pp. 133–149.

Frey, B. S. and W. Pommerehne. (1984). "The hidden economy: state and prospect for measurement." *Review of Income and Wealth*, 30(1), 1–23.

Friedman, E., S. Johnson, D. Kaufmann, and P. Zoido-Lobaton. (1999). "Dodging the grabbing hand: The determinants of unofficial activity in 69 countries." Discussion Paper, Washington, DC, World Bank.

Giles, D. E. A. (1999a). "Measuring the hidden economy: implications for econometric modeling." *The Economic Journal*, 109(456), 370–380.

Giles, D. E. A. (1999b). "Modeling the hidden economy in the tax-gap in New Zealand." *Working Paper*, Department of Economics, University of Victoria, Canada.

Hill, R. and M. Kabir. (1996). "Tax rates, the tax mix, and the growth of the underground economy in Canada: what can we infer?" *Canadian Tax Journal/Revue Fiscale Canadienne*, 44(6), 1552–1583.

Houston, J. F. (1987). "Estimating the size and implications of the underground economy." *Working Paper* 87–9, Federal Reserve Bank of Philadelphia, Philadelphia (N.J.).

Jain, A. K. (ed.) (1998). *Economics of Corruption*. Boston: Kluwer.

Johnson, S., D. Kaufmann, and A. Shleifer. (1997). "*The unofficial economy in transition.*" Brookings Papers on Economic Activity, Fall, Washington D.C.

Johnson, S., D. Kaufmann, and P. Zoido-Lobatón. (1998a). "Regulatory discretion and the unofficial economy." *The American Economic Review*, 88(2), 387–392.

Johnson, S., D. Kaufmann, and P. Zoido-Lobatón. (1998b). "Corruption, public finances and the unofficial economy." *Discussion Paper*, The World Bank, Washington, DC.

Kaufmann, D. and J. Sachs. (1998). "*Determinants of Corruption.*" Harvard University (Unpublished manuscript).

Kazemier, B. (2005). "The undergroud economy: A survey of methods and estimates", *Discussion Paper*, Statistics Netherlands, Voorburg, Netherlands.

Kirchgaessner, G. (1983). "Size and development of the West German shadow economy, 1955–1980." *Zeitschrift für die gesamte Staatswissenschaft*, 139(2), 197–214.

Kirchgaessner, G. (1984). "Verfahren zur Erfassung des in der Schattenwirtschaft erarbeiteten Sozialprodukts." *Allgemeines Statistisches Archiv*, 68(4), 378–405.

Klovland, J. (1984). "Tax evasion and the demand for currency in Norway and Sweden: is there a hidden relationship?" *Scandinavian Journal of Economics*, 86(4), 423–439.

Lemieux, T., F. Bernard, and P. Fréchette. (1994). "The effect of taxes on labor supply in the underground economy." *The American Economic Review*, 84(1), 231–254.

Lippert, O. and M. Walker. (eds.) (1997). *The Underground Economy: Global Evidences of its Size and Impact*. Vancouver, B C: The Frazer Institute.

Loayza, N. V. (1996). "The economics of the informal sector: A simple model and some empirical evidence from Latin America." *Carnegie-Rochester Conference Series on Public Policy*, 45, 129–162.

Mauro, P. (1995). "Corruption and growth." *Quarterly Journal of Economics*, 11093, 681–712.

Mogensen, G. V., H. K. Kvist, , E. Körmendi, and P. Soren. (1995). *The Shadow Economy in Denmark 1994: Measurement and Results*. Study No. 3, Copenhagen, The Rockwool Foundation Research Unit.

Neck, R., H. Markus., and F. Schneider. (1989). "The consequences of progressive income taxation for the shadow economy: Some theoretical considerations," in Boes, D. and Felderer, B. (eds.) *The Political Economy of Progressive Taxation*. Berlin Heidelberg New York: Springer, pp. 149–176.

Pedersen, S. (2003). "The Shadow Economy in Germany, Great Britain and Scandinavia: A Measurement Based on Questionnaire Service", *Study No. 10*, The Rockwoll Foundation Research Unit, Copenhagen.

Pelzmann, L. (1988). *Wirtschaftspsychologie*. Arbeitslosenforschung, Schattenwirtschaft, Steuerpsychologie. Berlin Heidelberg New York: Springer.

Pommerrehne, W. W. and F. Schneider. (1985). "The decline of productivity growth and the rise of the shadow economy in the U.S." Working Paper 85–89, University of Aarhus, Aarhus, Denmark.

Quirk, P. J. (1996). "Macroeconomic implications of money laundering." *IMF Working Paper* 96/66, Washington, DC.

Riebel, V. (1983). *Die Schwarzarbeit als Problem der Zeitallokation*. Frankfurt am Main, Bern: Lang.

Riebel, V. (1984). "Arbeitszeitverkürzung und Schwarzarbeit. Auswirkungen einer Verkürzung der Wochenarbeitszeit auf das individuelle Arbeitsangebot." *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 104(5), 515–538.

Romer, P. (1994). "New goods, old theory, and the welfare costs of trade restrictions." *Journal of Development Economics*, 43(1), 5–38.

Rose-Ackermann, S. (1978). *Corruption: A study in Political Economy*. New York: Academic Press.

Rose-Ackermann, S. (1997). *Corruption and Development*. Annual Bank Conference on Development Economics. Washington DC, The World Bank.

Rose-Ackermann, S. (1999). *Corruption and Government: Causes, Consequences and Reforms*. Cambridge, MA: Cambridge University Press.

Rose-Ackermann, S. (2002). 'Grand' corruption and the ethics of global business," *Journal of Banking and Finance* 26(4), 1889–1918.

Rose-Ackermann, S. (2004). "Governance and corruption," in Lomborg, B. (ed.) *Global Crises, Global Solutions*. Cambridge UK: Cambridge University Press, pp. 301–344.

Rose-Ackermann, S. (ed.) (2006). *International Handbook on the Economics of Corruption*. Cheltenham UK: Edward Elgar Publishing.

Schmölders, G. (1960). Das Irrationale in der öffentlichen Finanzwirtschaft. Probleme der Finanzpsychologie. Hamburg: Rowohlt.

Schmölders, G. (1975). *Einführung in die Geld- und Finanzpsychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Schneider, F. (1986). "Estimating the size of the Danish shadow economy using the currency demand approach: an attempt." *The Scandinavian Journal of Economics*, 88(4), 643–668.

Schneider, F. (1994a). "Measuring the size and development of the shadow economy. Can the causes be found and the obstacles be overcome?" in Brandstaetter, H., and Güth, W. (eds.) *Essays on Economic Psychology*. Berlin Heidelberg New York: Springer, pp. 193–212.

Schneider, F. (1994b). "Can the shadow economy be reduced through major tax reforms? An empirical investigation for Austria." *Supplement to Public Finance/Finances Publiques*, 49, 137–152.

Schneider, F. (1997). "The shadow economies of western Europe." *Journal of the Institute of Economic Affairs*, 17(3), 42–48.

Schneider, F. (1998a). "Further empirical results of the size of the shadow economy of 17 OECD countries over time." Paper to be presented at the 54 Congress of the IIPF Cordowa, Argentina, and discussion paper, Department of Economics, University of Linz, Linz, Austria.

Schneider, F. (1998b). Stellt das Anwachsen der Schwarzarbeit eine wirtschaftspolitische Herausforderung dar? Einige Gedanken aus volkswirtschaftlicher Sicht. Linz, *Mitteilungen des Instituts für angewandte Wirtschaftsforschung* (IAW), I/98, S. 4–13.

Schneider, F. (2001a). "Die Schattenwirtschaft—Tatbestand, Ursachen, Auswirkungen," in Anton Rauscher (ed.) *Die Arbeitswelt im Wandel,* Mönchengladbacher Gespräche, Band 21, Köln, 2001.

Schneider, F. (2001b). Arbeit im Schatten: Einige theoretische und empirische Überlegungen über die Schattenwirtschaft, *Perspektiven der Wirtschaftspolitik*, Blackwell, Band 2, Heft 4, 2001.

Schneider, F. (2001c). "What do we know about the shadow economy?" Evidence from 21 OECD countries." *World Economics*, 2(4), 19–32.

Schneider, F. (2005). "Shadow Economies around the World: What do we really know?", *European Journal of Political Economy*, 21(3), 598–642.

Schneider, F. (2007). "Shadow economies and corruption all over the world: New estimates for 145 countries", *Economics: The Open Access, Open Assessment* E-journal No. 2007/9, July 24, 2007.

Schneider, F. and D. Enste. (2000). "Shadow economies: Size, causes, and consequences." *Journal of Economic Literature*, 38(1), 114–177.

Schneider, F. and D. Enste. (2002). "The Shadow Economy: Theoretical approaches, empirical studies, and political implications", Cambridge: Cambridge University Press.

Schneider, F., M. F. Hofreither, and R. Neck. (1989). "The consequences of a changing shadow economy for the official economy: Some empirical results for Austria," in Boes, D. and Felderer, B. (eds.) *The Political Economy of Progressive Taxation*. Berlin Heidelberg New York: Springer, 181–211.

Shleifer, A., and R. W. Vishny. (1993). "Corruption." *Quarterly Journal of Economics*, 108(4), 559–617.

Smith, P. (1994). "Assessing the size of the underground economy: The Canadian statistical perspectives." *Canadian Economic Observer*, Catalogue No. 11–010, 3.16–33, at 3.18.

Spiro, P. S. (1993). "Evidence of a post-GST increase in the underground economy." *Canadian Tax Journal/Revue Fiscale Canadienne*, 41(2), 247–258.

Tanzi, V. (1994). "Corruption, governmental activities, and markets." IMF Working Paper, 99, 1–20.

Tanzi, V. (1998). "Corruption around the world: Causes, consequences, scope, and cures." IMF Working Paper, 63, 1–39.

Tanzi, V. (1999). "Uses and abuses of estimates of the underground economy." *The Economic Journal*, 109(456), 338–340.

Thomas, J. J. (1992). "Informal economic activity." *LSE, Handbooks in Economics*. London: Harvester Wheatsheaf.

Thomas, J. J. (1999). "Quantifying the Black economy: 'Measurement without theory' yet again?" *The Economic Journal*, 109(456), 381–389.

Trockel, J. (1987). "Die Schattenwirtschaft in der Bundesrepublik Deutschland." *Eine ökonomische Analyze am Beispiel der Bauwirtschaft*. Köln, Eul: Bergisch-Gladbach.

Werner, C. (1990). *Die Beschäftigungswirkungen der Schattenwirtschaft*. Pfaffenweiler: Centaurus

Zilberfarb, B. Z. (1986). "Estimates of the underground economy in the United States, 1930–80." *Staff Papers* (International Monetary Fund), 33(4), 790–798.

Chapter 29
Vote and Popularity Functions[1]

Martin Paldam

Introduction

During the last 30 years about 300 papers on Vote and Popularity functions (VP-functions) (defined in Table 29.1) have been written.[2] Most of the research is empirical. The purpose of this article is to survey this literature and discuss how the empirical results fit into economic theory.

It is my experience that when academic economists are confronted with the results of the VP- research, they frown, as they go against "our" main beliefs. Voters do not behave like the *economic man* of standard theory. In other words, the results are not "economically correct"—as defined in Table 29.2. Political scientists have other problems depending on their school, so this essay is written to the typical mainstream economist (as is the author).

From bedrock theory follows a remarkable amount of nice, sound theory, and everything can be generalized into the general equilibrium, growing along a steady-state path maximizing consumption. Politics convert the demand for public good into the optimal production of such goods and minimizing economic fluctuations. The past is relevant only as it allows the agents to predict the future, markets are efficient, etc. This nice theory is well known, and it is a wonderful frame of reference. Especially in the 1980s, a strong movement in economics argued that the world was really much closer to the bedrock than hitherto believed. If the noise terms are carefully formulated, the world is loglinear and everybody maximizes from now to infinity.

Bedrock theory suffers from two related problems. The first is that it is a bit dull. So, many models are "set into motion" by some (small) deviation from perfection—for

---

Table 29.1 Defining the VP-function

Vote function: $V_t = F_t^e + F_t^p,$	Explaining the vote for the government, at the election in time t, by economic, F_t^e, and political variables, F_t^p
Popularity function:	Is a formally similar function explaining the polled popularity of the government
VP-function:	Vote and Popularity functions are formally alike and closely related

Table 29.2 Characterizing the economically correct model

Bedrock theory: Models are built symmetrically around a *central case* where *rational* agents maximize their utility from now to infinity given perfect foresight. The key agent is termed *economic man*.

example, an observation that seems to contradict the theory.[3] It is almost like a good old crime story. A criminal is needed for the story to be interesting. However, in the end the criminal is caught and all is well once again.

The second problem is *the theodicy problem of economics*.[4] With such rational agents, we expect that economic outcomes are rational too. How come we see so many crazy outcomes when we look at the world? Average GDP differs between countries by 100 times. Some countries have pursued policies that have reduced their wealth by 2%–3% per year for several decades (think of Zambia). All countries have irrational institutions such as rent control and (at least some) trade regulation. Discrimination based upon ethnic differences is common. Debt crises have frequently occurred. The reader will probably agree that nobody can be closer to *economic man* than the bankers of the island of Manhattan: How come that even they managed to lend so much to Bolivia that at one point in time the debt burden reached 145% of GDP? We will return to the theodicy problem of economics at the end, but let us return to the subject matter.

Economic data tend to be much better—and easily accessible—than political data, so most of the literature on VP-functions concentrates on the economic part of the function. The present essay follows this tradition and uses the setup listed in Table 29.3.

Many experiments have been made with the lag structure, plenty of economic variables have been tried and sometimes more genuine political variables have been included,[5] nonlinear functional forms have been used, etc., but we shall mainly discuss the simple linear expressions (29.3) and (29.4) from Table 29.3.

---

[3] Also, a lot of papers in economics seem to be theory-driven in the sense that they change old models using unfashionable assumptions into new models using more appropriate assumptions.

[4] The reader probably knows that theologists have struggled with their theodicy problem for the last 2000 years: If God is good and omnipotent, how come we see so much random and human cruelty in the world? Theologists most of the time manage to convince themselves that it is bad theology even to recognize that there is a problem, but then they may encounter a child suffering from terminal cancer.

[5] Dummies for special events are common, but some quantitative variables have been tried, too. The most famous is the *rally-around-the-flag* variable for foreign policy crisis constructed from political almanacs, where the size of the spikes are assessed by media volume and the speed of the decay after the event is estimated.

Table 29.3 The basic quarterly macro VP-function

(29.1) $VP_t = F_t^e + F_t^e$	Model from Table 29.1
(29.2) $F_t^e = \alpha_g + \gamma t_g + e_{2t}$	Political model (formal)
(29.3) $F_t^e = \beta_1 u_{t-\text{lag}} + \beta_2 p_{t-\text{lag}} + \cdots + e_{1t}$	Economic model (measured variable)
(29.4) $VP_t.!10.6u_{t-(1/4)} 10.6p_{t-(1/4)} + \cdots + \alpha_g 10.15 t_g + e_t$	Typical coefficients estimated

Greek letters are coefficients to be estimated. e's are residuals $u_t, p_t, \ldots$, are economic variables as unemployment (u), inflation (p), etc. α_g and t_g, the political part is reduced to a government specific constant and trend

Main Results in the Literature

The literature on VP-functions is large, but most of the findings can be concentrated as in Table 29.4. The starting point is a simple hypothesis.

The responsibility hypothesis: Voters hold the government responsible for the economy. From this hypothesis follows the reward/punishment mechanism: Voters punish the government in votes and polls if the economy goes badly, and they reward it if the economy goes well.

The hypothesis is not without problems. Governments may not have a majority, or external shocks may occur, which no sane person can ascribe to the government. A variable giving the *clarity of responsibility* may consequently enter the function. This is referred to as "content" in Table 29.4—a subject that will not be discussed at present.

The literature was started by Kramer (1971) writing about vote functions (in the United States),[6] while Mueller (1970) and Goodhart and Bhansali (1970) presented the first popularity function—for the United States and United Kingdom—almost simultaneously.

The most important contributions since then are much more difficult to point out, but the following may be mentioned: Frey and Schneider (1978a, b) and Hibbs (1982) generated a wave of papers, both due to their lively fight and to new developments: Frey and Schneider integrated the VP-function into a set of policy reaction functions, while Hibbs mainly developed the modeling techniques.

The micro-based literature was started by Kinder and Kiewiet (1979). It was further pushed by Lewis-Beck (1988), while the cross-country discussion was started by Paldam (1991). A good sample of papers giving the present stage of the arts can be found in two recent volumes (both results from conferences trying to collect the main researchers in the field): Lewis-Beck and Paldam (2000) and Dorussen and Taylor (2002). This effort has generated the results listed in Table 29.4.

Each of the items of the table except the last will be discussed in a short section. We will argue that most are contrary to economic correctness, but that they are all possible to rationalize. However, they are at the *opposite extreme* of the *rationality spectrum*

---

[6] It appears that Kramer's paper was around as a working paper longer than the other two, even when it was published a year later.

Table 29.4 Main results in the literature

Section	Finding	Empirical status
3.	*The big two:* Voters react to mainly unemployment and inflation	Uncontroversial
4.	*Myopia:* The time horizon of voters is short – events more than 1 year from an election have small effects only	Uncontroversial
5.	*Retrospective:* Voters react to past events more than to expected future ones, but the difference is small as expectations are stationary	Controversial
6.	*Sociotropic:* In most countries voters are both sociotropic and egotropic[a]	Controversial
7.	*Low knowledge:* Voters know little about the (macro) economy	Uncontroversial
9.	*Grievance asymmetry:* Voters punish the government more for a bad economic situation than they reward it for a similarly sized good one	Controversial
10.	*Cost of ruling:* The average government ruling a normal four-year period loses 2.5% of the votes. This result is independent of party system, voting law, country size, etc.	Uncontroversial
Not covered	*Context:* The VP-function only generalizes if set in the same context. In particular, the responsibility pattern generalizes if the government is *clearly* visible to the voter	Only explored in a dozen papers

Note: The status line indicates if the result is *controversial*, i.e., if a minority of the researchers in the field disagree. The article only considers the responsibility pattern and thus assumes a simple setting in which both the government and the opposition are well defined. Complex, changing coalitions and minority governments are not discussed.

[a] Sociotropic: voters care about the national economy. Egotropic: voters care about their own economy.

from the one normally considered by economists, see Fig. 29.1: on one side is *economic man* and on the other, the actual voter. This essay is written to say that this other side of the spectrum is getting far too little attention in standard theory.

In Table 29.3 the models are written as simple relations between levels, assuming all variables to be stationary. The present will largely disregard the substantial problems of estimation. That is, (i) should the model be formulated in levels or in first differences? (ii) should it contain an error correction term? (iii) should the series be pre-whitened? Popularity series are known to have complex and shifting structures in the residuals when modeled. So there are plenty of estimation problems. Many of the papers are from before the cointegration revolution, but new papers keep coming up using state-of-the-art techniques, though they rarely find new results.

Politologists have found that only 20% of the voters in the typical election are *swing voters*, and the net swing is considerably smaller (see Section 29.10). It means that 80% vote as they always do. The permanent part of voting is termed the *party identification, Id*. It is not 100% constant, and it needs an explanation, but it should be differently explained from the swing vote. The VP-function concentrates on the swing voters, but it may be formulated with terms handling the more permanent part of the vote as well.
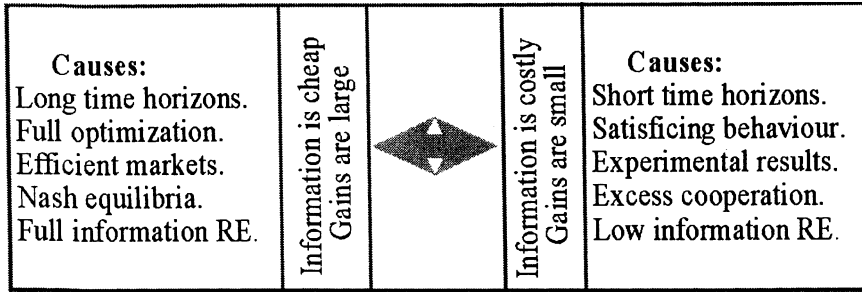
Fig. 29.1 The two ends of the rationality spectrum

In the more simple formulations, one may work with *level estimates* where the committed voters enter into the constants, and with the *first difference estimates* where the committed voters are "differenced out". The choice between the two formulations is then a question of estimation efficiency to be determined by the structure of cointegration between the series, and the (resulting) structure of the error terms.

The Big Two: Unemployment (Income) and Inflation

The two variables that are found in most VP- functions are the rates of unemployment and inflation, u_t and p_t. Both normally get negative coefficients, often about -0.6 as listed in model (29.4) of Table 29.3. Unemployment is sometimes replaced with income, thereby confirming Okun's law. The Phillips curve is sufficiently weak so that unemployment and inflation have little colinearity in these functions.

The data for the rate of unemployment and the vote share for the government have roughly the same statistical structure so that it is possible that they can be connected as per the linear version of model (29.1). Data for the rate of inflation are upward skewed. That is, inflation can explode and go as high as the capacity of the printing press allows. Also, people pay little interest to low inflation rates, but once it reaches a certain limit it becomes the key economic problem.[7] Hence, inflation cannot enter linearly in model (29.1), except of course, if we consider a narrow interval for inflation rates. Fortunately, inflation is often within a narrow interval in Western countries.

An interesting controversy deals with the role of unemployment. It was started by Stigler (1973) commenting on Kramer (1971). Stigler remarked that a change of unemployment of 1 percentage point affected 1% of the workers only—that is 0.5% of the population. Approximately 80% of those vote for the Left anyhow. The potential for affected swing voters is therefore only 0.1% of the voters. How can

---

[7] This tallies nicely with the findings in the literature dealing with effects of inflation. High inflation is harmful, but whether inflation is 2% or 5% seems to have no effects on the real economy.

this influence the vote by 0.6%? Note that Stigler automatically assumes that voting is egotropic and retrospective. You change the vote if you—yourself—are affected by the said variable.[8] This point deals with the micro/macro experiences and the observability of the variables. It has often reappeared in the literature. Table 29.5 shows some of the key points.

The table lists the important variables and some other variables that have often been tried, but with little success. Unemployment and income affect individuals differently and are observable both at the micro and the macro levels. Inflation is more difficult to observe for the individual at the macro level. We see prices go up, but the individual cannot observe if they rise by 2% or 3%. However, this is covered by the media.

The other variables listed—the balance of payments and the budget deficit—are much discussed in the media and are important in political debates. They are important *predictors for policy changes*, and indicators of *government competence*. However, they have no micro observability. It is interesting that they are rarely found to work in VP- functions.

Refined models with competency signaling and full information rational expectations, where voters react to predicted effects of, e.g., budget deficits, are contrary to the findings in the VP- function literature. In fact, when we look at what people know about the economy—see Section 29.7—it is no wonder that they do not react to changes in the balance of payments and the budget.

Under the responsibility hypothesis model (29.1) is an estimate of the *social welfare function*. It comes out remarkably simple. Basically, it is linear and looks like Model (29.4) in Table 29.3. The main problem with such estimates is that they are unstable. Many highly significant functions looking like (29.4) have been estimated, but they frequently "break down" and significance evaporates.

If (29.4) is stable, it appears inconceivable that it cannot be exploited politically. And, in fact a whole literature on political business cycles has been written—since Nordhaus (1975) exploring the possibility for creating *election cycles* in the

Table 29.5 The character of the variables entering in the VP-functions

	Micro-experience	Observability	Significant
Unemployment Income	Very different for individuals	Personal and media	Mostly[a]
Inflation Balance of	Different for individuals	Personal and media	Mostly
payment Budget deficit	Similar for individuals	Mostly media	Rarely
	None	Only media	Never

[a] The two variables have strong colinearity in VP-functions.

_____

[8] It is easy to reach a much higher number than did Stiegler. (i) Imagine that the average household has three voters, who have an income loss if one member becomes unemployed. (ii) An increase of 1% in the rate of unemployment means that three people suffer a spell of unemployment. Hence, nine people experience a loss when unemployment rises by 1%. The 0.1% in the text now is 0.9% and that is larger than the 0.6% that has to be explained. Also, maybe your welfare is affected by the way your friends fare.

economy. However, most studies have shown that such cycles do not exist in practice. What may exist is rather the reverse, governments may steer the economy as per their ideology and create *partisan cycles*.[9]

Voters are Myopic

The voter's myopia result deals with the duration of the effect of a sudden economic change. Imagine a short and sharp economic crisis—such as a drop in real GDP lasting one year—how long will this influence the popularity of the government?

One of the most consistently found results in the VP-function literature is the voter's myopia result. Only the events of the past year seem to count. A few researchers (Hibbs 1982) have found that as much as one-third of the effect remained after one year, but most researchers have been unable to find any effect after one year. The myopia result has been found even for political crises, which are often sharply defined in time. Consequently, the model looks as follows (equation (29.4) in Table 29.3):

$$VP_t = I_{-4}^t \ f(u_t, p_t)e^{rt}dt \cdot I_{t-1}^t f(u_t, p_t)e^{rt}dt \qquad (29.5)$$

The subscript $(t-1)$ represents a lag of one quarter (or perhaps one year) as before. The welfare maximization of the variable u (say unemployment) leading to the vote is made from $t-1$ to t, where the time unit "1" is a year, and the "discounting" expression e^{rt} has a high discount rate so that everything before $t-1$ is irrelevant.

Formula (29.5) is surely not how such expressions look in economic textbooks. A key part of economic correctness is that *economic man has a long time horizon and looks forward*. The common formulation of the closest corresponding models is:

$$W_t = I_t^4 f(u_t, p_t)e^{-rt}dt \qquad (29.6)$$

The welfare to be maximized is a function of the relevant economic variable from now to infinity, with a small discount rate, r, perhaps even as small as the long-run real rate of interest or the real growth rate. It gradually reduces the weight of future values, but events 20 years into the future count significantly.

Expressions (29.5) and (29.6) are hard to reconcile. First, the maximization is *retrospective* in (29.5) and *prospective* in (29.6), as will be discussed in Section 29.5. Second, the time horizons are dramatically different. None of the 300 studies has ever found evidence suggesting that events as far back as two years earlier have a measurable impact on the popularity of the government!

---

[9] The main articles in the field are reprinted and surveyed in Frey (1996). See Gärtner (1994) and Paldam (1997) for recent surveys.

Many descriptions have been made of the political decision process by participants in the process and by the keen students of current affairs found among historians, political scientists, and journalists. A common finding is that the decision process tends to have a short time horizon. The political life of a decision maker is uncertain and pressures are high. How decisions are made has little in common with the description of "benevolent dictators maximizing social welfare" still found in economic textbooks and many theoretical models.

The outcomes of "benevolent dictator calculations" have some value as a comparative "benchmark," and as ideal recipes for economic policy making.[10] However, some theorists present such exercises as realistic descriptions of policy making, deceiving young economists into believing that this is the way political decisions are made.

Voters are Retrospective/Expectations are Static

One of the key facts about economic theory is that it is largely theory driven. One of the main areas over the last four decades has been the area of expectation formation. It has been subjected to a huge theoretical research effort. Less interest has been given to research in the actual formation of inflationary expectations where real people are actually polled, as the results have typically been embarrassing. I think that we all know in our heart of hearts that real people cannot live up to our beautiful theories about *economic man*.

In the field of VP- functions about 50 papers have looked at the existing data and found that in many countries retrospective/prospective (RP) pairs (equation (29.4) in Table 29.3)—defined in Table 29.6—have been collected for such series as unemployment, inflation, and real income. The papers have then tried to determine which of the two variables in the pair are the most powerful ones for predicting the vote/popularity of the governments.

Most of the analysis is done on micro-data of individual respondents, so many thousands of observations have been used to determine this controversy. The many papers do not fully agree, so the results of the efforts have to be summarized as follows:

RP1: The two series in the RP pair normally give almost the same results.
RP2: Most results show that the retrospective series are marginally more powerful.[11]

---

[10] The reader should consult Tinbergen (1956) and Johansen (1978, 1979) to find the classical version of how benevolent dictators maximize social welfare, just in case one such ruler ever happened. When reading such descriptions, it is sad to contemplate that we are dealing with some of the most brilliant minds in our profession.

[11] Recent research from Portugal has found a RP-set that differs considerably, and here the retrospective series works significantly better, see Veiga and Veiga (2004).

Table 29.6 A polled RP-pair

Economics	Politology	Question in poll
Past experience	Retrospective	How has X developed during the last Z-period?
Expectations	Prospective	How do you expect X will develop during the next Z-period?

Note: X is an economic variable and Z is a time period like a quarter, a year or a couple of years.

I think that virtually all economists will agree that the correct result in the RP controversy is that the prospective twin should beat the retrospective one by a long margin. But by a rough count the retrospective twin wins in two of three cases in the studies made.

Some of the main discussants in this research are Helmut Norpoth for the majority retrospective view and Robert S. Erikson for the minority prospective view. They are working with the same data for the United States. Erikson terms the controversy *bankers or peasants*. Bankers work professionally with the economy and have prospective expectations. Peasants are interested in matters of farming mainly, and hence are retrospective, when it comes to the economy. The question therefore is, whether the average voter behaves mostly as a banker or as a peasant.[12] Once the question is asked, it appears that the obvious answer must be that the average person is a peasant. However, Erikson finds that voters behave as bankers.[13]

When the results of Erikson and Norpoth are compared, the difference is small. The most disgraceful result actually is (RP1) that the two series in the existing RP-pairs are as similar as identical twins. The only conclusion one can draw is that people form largely static expectations.

The author's own poll of 4788 Danes asking about the RP-pair for inflation found a net difference in the answers of 34 cases, i.e., 0.7% of the respondents (Nannestad and Paldam, 2000).[14] With such a tiny difference it is no wonder that we were unable to find any difference in the fit of the VP- function if we used the prospective or retrospective series. This is typical also for the British and the German results.

This brings us back to the large gap separating formulas (29.5) and (29.6). It does solve the apparent contradiction between the direction of the maximization if voters have static expectations. But then, surely it is much easier to use the past as in (29.5). When we look at the vast literature building highly refined theory of inflationary expectations and analyzing the dynamic consequences of the different assumptions, it is hard to reconcile with the findings of the VP- literature. Here is surely a field where facts are much duller than fiction.

---

[12] Compare Nickelsburg and Norpoth (2000) and Erikson et al. (2000)—using almost the same data. It is probably unfair to single out two of the authors from their coauthors, but they are the ones I have heard defending their views most eloquently.

[13] Even the farsightedness of bankers can be doubted as mentioned in the introduction.

[14] The polls were done quarterly for the years 1990–93, when inflation had a clear downward trend. Nevertheless, only 0.7% of the respondents (net) predicted a fall.

Voters are Mainly Sociotropic—Or Perhaps Not

Once the analysis of VP-functions moved into micro research, data increased dramatically to allow much stronger tests, but new interesting problems came up. The most intriguing was probably the egotropic/sociotropic (ES) controversy, where the two terms are defined in Table 29.7.[15]

Like the RP-pairs also ES-pairs exist in various data sets: The egotropic question is "how has your own economic situation developed in the last Z-period?" The corresponding sociotropic question is: "How has the economic situation of your country developed in the last Z-period?" The "economic situation" is sometimes replaced with "unemployment," and once more Z is typically either a quarter or a year.

The economically correct answer is surely that *economic man* is egotropic. Therefore, it was shocking when Kinder and Kiewiet (1979) demonstrated that the U.S. voter was sociotropic. Several other studies—notably Lewis-Beck (1988)—have confirmed the result also for other countries, even after considerable refinement of the questions. So, for more than a decade there was no doubt that the main result in the literature was that voters are sociotropic, contrary to economic correctness.

Kinder and Kiewiet's model was remarkably simple:

$$VP_i = (_0 + \forall E_i + \exists S_i + 8Id_i + u_i) \tag{29.7a}$$

Here i is an index for the individual, E_i is the egotropic variable, S_i is the sociotropic variable, and Id_i is a party identification variable.

The model of Kinder and Kiewiet was estimated on a cross-section. In an unusually aggressive article, Kramer (1983) pointed out that this was not the right approach for the problem. In a cross-section, the true sociotropic variable is constant and, hence, unable to explain anything. What is estimated as sociotropic can only be different perceptions, N_i, of the same objective variable, Y:

$$VP_i = (_0 + \forall E_i + \exists N_i(Y) + 8Id_i + u_i) \tag{29.7b}$$

Kramer did not see that the reformulation makes the economic correctness problem larger, not smaller! Surely, the rational voter perceives Y unbiased—that is, the perception error is white noise—and hence the coefficient to $N_i(Y)$ should be zero. Kinder and Kiewiet's finding that the sociotropic term, $\exists$, dominates the egotropic term, $\forall$, becomes even more mysterious. The most reasonable interpretation of the finding is, therefore, that the different perceptions estimated must be due to different personal experiences, and hence that what is estimated as a sociotropic effect is really egotropic.

---

[15] Both words are created for the purpose and not in most dictionaries: sometimes the term egotropic is replaced by more loaded terms as "egocentric" or "egoistic."

Table 29.7 Defining egotropic and sociotropic

The economic factor in the VP-function is:	
Egotropic	what matters is the personal economy
Sociotropic	what matters is the national economy

The next generation of models trying to come to grips with the egotropic/sociotropic distinction was introduced by Markus (1988). He reformulated (29.7b) into a mixed cross-section time series model, which includes a time index, t:

$$\text{VP}_{i,t} = (_0 + \forall E_{i,t} + \exists Y_t + 8Id_{i,t} + u_{i,t}) \tag{29.8a}$$

$$\text{VP}_{i,t} = (_0 + \forall E_{i,t} + \exists N_{i,t}(Y_t) + 8Id_{i,t} + u_{i,t}) \tag{29.8b}$$

In Model (29.8a), the polled sociotropic variable is replaced with the "objective" one from the national statistical office. It can be compared with (29.8b), which should be almost the same as (29.7).

Model (29.8a) gives rather different results than (29.7). Now $\forall$ and $\exists$ become approximately the same on U.S. data. Several studies have now estimated the various models for more countries. United Kingdom is covered by Price and Sanders (1994),[16] Denmark by Nannestad and Paldam (1997a), and Sweden by Johrdahl (2001). It appears that both $\forall$ and $\exists$ become significant, though not in a predictable mixture.

Thus the old agreement that voting is *only* sociotropic has given way to a more unclear picture, where voting is a *mixture* of egotropic and sociotropic factors. This is less economically incorrect than the old view, but a sizable sociotropic factor still appears in voting in most countries.

Voters are Uninformed about the Economy

Many polls have asked people about their knowledge of the economy. It appears that nobody has collected such polls systematically, and few studies have therefore been made comparing results.[17] However, the results show rather decisively that voters know little about the economy.

In our Danish polls (Nannestad and Paldam, 2000) most voters knew the number of unemployed within a few percentage points. They tended to know that the country had a balance of payments deficit and a budget deficit, *both* when the country had such deficits, and when it did not. When the two balances had the reverse sign, many mixed them up. Virtually nobody knew the sizes of the balances. Also, about

---

[16] The studies of the United Kingdom are rather numerous as a search in a good library database under the names Harold Clarke, Davis Sanders, and Paul Whitley will show.

[17] Aidt (2000) is an attempt to find and summarize such polls.

two-thirds of the voters could not give an assessment of the inflation rate within 2 percentage points (that is, if inflation was 3% they were either outside the range of 1–5% or answered "don't know").

However, there was one time during the four years covered, where knowledge increased substantially. This was around a general election. All of a sudden people knew that the balance of payments had changed from red to black figures.

The same type of result has been found about the EU: In European countries with no referenda institution, people know little about the European Union, but in countries with frequent referenda people know a lot more (Paldam 2001).

So, either people do seek information in connection with elections or they learn from watching the big show of an election/referendum campaign on TV. The fact that the information level goes up is worth to note as it explains why vote functions have a lower fit than popularity functions. It also explains why party popularities are normally much more volatile around elections than else.

Economic theory predicts that voters know what they need to know. The marginal benefits of information should be equal to the marginal costs:

$$MB(I) = MC(I) \text{ (the condition for rationality of information)} \qquad (29.9)$$

One way to *define rational expectations* is to demand that (29.9) is fulfilled.[18] This is the definition used in the present paper.

The big problem surrounding (29.9) is what the benefits MB(I) are. Is it possible to argue convincingly that MB(I) is significantly larger than zero?

It is sometimes argued that people do need information on the macro level when they buy and sell shares and bonds (i.e., in connection with financing a house or a business) and when they buy and sell foreign exchange in connection with journeys abroad. But the theory of *market efficiency*—where many professional dealers are present—does effectively cut the link to information here. A (small) dealer in these markets can find no better information than the price.

The one occasion where people need information about the macro economy is when they vote at national elections, their union, etc. Hence, we write:

$$MB(I) = 1MB(V)/MI \text{ (marginal benefit of economic information should be}$$
$$\text{equal to the marginal improvement in the benefits derived from voting better)}$$
$$(29.10)$$

This brings us to the problems surrounding MB(V), the marginal benefit of voting. It is the problem known in the paradox of voting, where almost the same equation as (29.10) can be written:

$$MB(V) = MC(V) \text{ (the condition for rationality of voting)} \qquad (29.11)$$

---

[18] Two other definitions are: (i) Data contain no exploitable regularity. (ii) Expectations are model-consistent.

Much has been written about (29.11), but nobody has been able to argue convincingly that MB(V) differs significantly from zero. That makes all terms in (29.10) and (29.11) zero! However, we know that people do vote even if they do not know much about the economy. Surely, our friend *economic man* is in trouble. He should know nothing about the economy, and never vote. In short, the average voter behaves differently from *economic man*.

Two Ways to Rescue *Economic Man*

Two lines of arguments are often used to rationalize this apparent mess described in the last section. One is to argue that the cost of information, C(I), is small too. The second is to show that the cost of voting in an election, C(V), is large, and belongs to the class of social capital observations. Unfortunately, the two attempts are rather in opposite directions.

C(I) is small too. This argument has been used by researchers such as Wittman (1995), Erikson et al. (2000), and Sanders (2000).[19] They argue the voters know what they need to know, and that this is not much. What they need is to "grasp" what is going on, i.e., to have some *feel* for the way the economy is going. This *feel* is acquired from the media without really trying. If most experts look gloomy, then things are going badly, and if they look relaxed, then things are going well. If a feel is enough, then it is possible to argue that C(I) is almost zero as well. There is something in this argument, and Sanders does show that soft "feel" questions about the economy can be scaled to track the actual economy reasonably well.

If everything goes on at the extreme low-cost end of the scale, so that MB(I) = MC(I) = MC(V) = MB(V) ≈ 0 which is almost zero, then things become a bit wooly. Also, it is well known that people are unable to make utility assessments involving small numbers.

The other approach is to start from *the large size of C(V)*. People do spend a lot of time looking at election campaigns on TV. Many go to some meetings, and pay membership fees to political parties. Nearly everybody spends an hour driving to the election location, waiting in line, and voting. In short, they have considerable costs *participating in the national democratic process*. Also, we know that parties find it worthwhile to spend millions on their campaigns. The total costs in time and money of a national election are likely to be something like 1% of GDP in the typical democracy.[20] So we have to explain why so much is spent by the individual and the political system.

---

[19] All the above are politologists, of which Donald Wittman is known as the one who is making the rational expectations revolution in political science.

[20] Assume that people use on average 15 hours on the election, and that only half of this is leisure, this is 0.5% of the normal working year to start with. Then it is normal to declare a school holiday, and the advertisement budget and the opportunity costs of the media coverage, etc. If one also adds the cost of a bad government for about half a year before the election and a bit after, then surely 1% of GDP is a low estimate of the costs of a national election.

Voting is a social capital phenomenon. It is a well-known observation that people in experiments play the cooperative solution much more than they should if they solve the game as good little *economic men*. The excess frequency with which the cooperative solution is played is a measure of the *social capital* or the *mutual trust* of the players. The key point to note is that trust does not need to be specific. It can be general. Many attempts have been made measuring general trust, and we know that it is much higher in some—generally more successful societies—than in others.[21]

In infinitely repeated games it is possible to uphold cooperative solutions, and it is arguable that society is a large number of games played with many players in different combinations over and over, seemingly without end. We do not bother solving all games in which we participate—and some of the everyday games are too complex to solve. So we develop rules-of-thumb standards, which can be termed trust. If a country succeeds in reaching a high level of trust, then it is an advantage to everyone, as the cooperative solution becomes the one automatically reached in many situations.

The attempts to integrate these observations in standard theory are still going on. It seems likely that they may succeed, and then the paradox of voting may be solved at long last.

Given that this solution works, then we know that people do undertake considerable cost to follow the political scene and participate in the process. Given that these costs are so high, why not imagine that people also try to follow the economy? The fact that they are not so knowledgeable may simply be that it is difficult as all those of us who teach the subject know.

It is interesting that the two attempts to save *economic man* are so different. In my judgment, the second approach is the most promising, but it is not yet integrated into standard theory. However, both approaches agree that *economic man* is the wrong model for the job at hand.

Voters have a Grievance Asymmetry

Already, the first popularity function study (Mueller 1970) discovered that voters react more to a negative economic event than to a corresponding positive one. It was also found by Bloom and Price (1975) commenting on Kramer (1971). Then, for some time nobody looked for a grievance asymmetry, and the effect was forgotten.

Once the analysis of micro data became organized, it became possible to look for more effects. And, in the 1990s several studies looked carefully and found a rather strong asymmetry. See, e.g., Price and Sanders (1994) and Nannestad and Paldam (1997b). It appears that if, e.g., unemployment increases by 1% (point), it has *twice* as large a negative effect as the positive effect of a fall in unemployment of 1% (point). The effect is thus large, and it is highly significant.

---

[21] In ongoing work I have measured how much larger social capital is in Denmark than in Russia by a whole battery of questions using the same questionnaire. It is between 2.5 and 4 (times) by *all* measures. The literature on social capital has recently been surveyed by several authors, see e.g., Paldam (2000).

This has the consequence that if the economy moves from A to B in a straight line, then the voter is more content than if it moves in a more roundabout way. In short, the variation around the growth path causes the government popularity to fall. Consider the average (Avr) and variance (Var) of a positive variable, i.e., a variable where

$$MVP/MX > 0 : MVP/MAvr(X) > 0 \text{ and } MVP/MVar(X) < 0 \qquad (29.12)$$

This is the standard formulation of *risk aversion*, so this is well integrated into economics, as long as it is prospective.

If the variables are retrospective, it is different. Relation (29.12) now changes from risk aversion to *loss aversion*. This is, in principle, an important change entering into one of the *economic-man* problems discussed by Kahneman (1994) and other critiques of standard theory. The problematic aspect is that utility becomes path dependent. Costs are not sunk at all.

However, if expectations are stationary, then it becomes unclear if we are dealing with risk aversion—which is perfectly nice theory—or loss aversion—which is a bad anomaly!

It Costs Governments Votes to Rule

The average government in a mature western democracy loses $2\frac{1}{2}\%$ of the votes just by ruling (see Paldam, 1991; and Nannestad and Paldam, 2002). There is no difference between the average outcome and its standard deviation in such countries as the Netherlands, Sweden, Italy, and Japan, so it is a surprising fact. Also, it is heavily under-researched.[22]

Per definition, the average government rules exactly as the rational voter should expect. So the economically correct prediction must surely be that the voter votes as before. It is hence a strange result that the average government loses a highly significant fraction of the votes. Three theories try to provide an explanation:

The oldest (Mueller, 1970) is the *coalition of minorities theory*. It essentially says that the governments are formed when parties before elections manage to put together inconsistent coalitions. However, when they rule, the inconsistencies are reviled. So, essentially the theory is that you can get away with unrealistic promises when in opposition. On average, $2\frac{1}{2}\%$ of the voters form unrealistic expectations at every election. This is a small irrationality, but it is a long-run fault, so it is not a "nice" theory, and it appears that no other evidence exists for the theory.

The second theory is the *median gap model*. It starts from a slightly amended version of the median voter theory. The pure form of the model suffers from the well-known problem that if both parties accept the policy of the median voter, they

---

[22] The research on the cost of ruling is surveyed and discussed in Nannestad and Paldam (2002), so I am brief at present. The second and third theory are from Paldam and Skott (1995) and Nannestad and Paldam (1997b).

become perfectly alike and there is no reason to vote. To be distinguishable, there has to be a gap around the median position between the voters. If there is a gap of size (some of the voters will get as close as they can come to their ideal policy if the government changes at every election. This has all been worked out as a perfectly respectable model, and it is even possible to calibrate the model with reasonably looking parameters so as to explain the observed fact. Unfortunately, little corroborating evidence exists for the theory.

Finally, the third theory builds upon the *grievance asymmetry* just discussed. Imagine that the outcome is symmetrically distributed around the expected outcome. So, some variables improve and some deteriorate, but the gains and losses are equally large. If the reaction is asymmetric, then there must be a loss of popularity in average. It is easy to show that the model produces the typical cost of ruling for reasonably sized VP-functions with the typical grievance asymmetry and realistically noisy economic outcomes.

Let us for a moment assume that the last theory is true. The average government loses votes because it rules, and this causes governments to change. This destroys the evidence for the median voter theorem. In a two-party (two-block) system, both parties (blocks) converge to 50% of the votes just because they rule when they are larger than 50%. The simplest explanation of the cost of ruling thus undercuts one of the key theorems of political economy.

Conclusion: Look at the Other End of the RE-Spectrum

Throughout the above survey it has appeared that the findings in the large VP-literature contradict the notions that the average voter behaves as does *economic man* of standard theory. That is, the symmetric, forward-looking agent who takes *everything relevant* into consideration for his decisions. This is the fellow we constantly meet in the theoretical literature. However, the voters we meet "out there in the real world" do not optimize forward, but backward, and they have a short time horizon. Also, they have a strong grievance asymmetry, so that it cost votes to rule.

Section 29.7 argued that it would be irrational if the voter behaved as *economic man*! Voting is a decision where ignorance is rational, even if it is accepted that a "feel" for the economy is enough, we are still faced with complex and unsolved questions.

However, once you start from the notion that politics is a field with much uncertainty, and hence a short time horizon, things start to fall into some order. Under the circumstances, it would be inconsistent if the voters had long memories. This means the election cycles are out of the question, but it is consistent with the notion of partisan cycles. The shorter the time horizon, the better does the past predict the future. Surely, within a few quarters nothing is likely to change very much.

The sizeable fraction of sociotropic voting—it is probably 25%–50%—also makes a lot of sense when you ask what elections are all about. It does deal with the whole economy, not with the economy of Mrs Voter herself! So perhaps it is not

so surprising that people give some consideration to the way they feel the government handles the economy when they vote.

Also, we do have a lot of results showing that the average citizen in all countries has considerable risk aversion (or, for that matter, loss aversion). The whole of the financial sector makes a perfectly good living out of turning risk pooling into a negative sum game for everybody else. So it is natural to expect a clear grievance asymmetry.

Once that is accepted, the cost of ruling follows. It is an important finding, as it causes parties to change in power and government coalitions to converge to 50% of the vote. This happens irrespective of the median voter theorem and the minimum winning coalition theorem. Or rather, it produces exactly the same observable facts in a much simpler way.

In short, it is worthwhile to take the findings in the VP-function literature seriously and use these findings for the development of a more realistic theory. Such a theory will also make it easier to explain the many suboptimal—or even crazy—outcomes we observe.

References

Aidt, T. (2000). "Economic voting and information." *Electoral Studies*, 19, 349–362.

Bloom, H.S. and Price, H.D. (1975). "The effects of aggregate economic variables on congressional elections." *American Political Science Review*, 69, 1232–1269.

Dorussen, H. and Taylor, M. (eds.) (2002). *Economic Voting*. London: Routledge.

Erikson, R.S., MacKuen, M.B., and Stimson, J.A. (2000). "Bankers or peasants revisited: economic expectations and presidential approval." *Electoral Studies*, 19, 295–312.

Frey, B. (1996). *Political Business Cycles*. Aldershot: Edward Elgar.

Frey, B.S. and Schneider, F. (1978a). "An empirical study of politico-economic interaction in the US." *Review of Economics and Statistics*, 60, 174–183.

Frey, B.S. and Schneider, F. (1978b). "A politico-economic model of the United Kingdom." *Economic Journal*, 88, 243–253.

Gärtner, M. (1994). "Democracy, elections, and macroeconomic policy: two decades of progress." *European Journal of Political Economy*, 10, 85–110.

Goodhart, C.A.E. and Bhansali, R.J. (1970). "Political economy." *Political Studies*, 18, 43–106.

Hibbs, D.A., Jr. (1982). "On the demand for economic outcomes: macroeconomic outcomes and mass political support in the United States, Great Britain, and Germany." *Journal of Politics*, 44, 426–462.

Johansen, L. (1978, 1979). *Lectures on Macroeconomic Planning*, 2 vols. Amsterdam: North-Holland.

Johrdahl, H. (2001). An economic analysis of voting in Sweden. Paper for the EPCS, Econ Department, Uppasala University, Paris.

Kahneman, D. (1994). "New challenges to the rationality assumptions." *Journal of Institutional and Theoretical Economics (JITE)*, 150, 18–36.

Kinder, D.R. and Kiewiet, D.R. (1979). "Economic discontent and political behavior: the role of personal grievances and collective economic judgement in congressional voting." *American Journal of Political Science*, 23, 495–527.

Kramer, G.H. (1971). "Short-term fluctuations in U.S. voting behavior, 1896–1964." *American Political Science Review*, 65, 131–143.

Kramer, G.H. (1983). "The ecological fallacy revisited: aggregate versus individual-level findings on economics and elections and sociotropic voting." *American Political Science Review*, 77, 92–111.

Lewis-Beck, M.S. and Paldam, M. (eds.) (2000). "Economic voting: an introduction." *Electoral Studies (special issue Economics and Elections)*, 19, 113–121.

Lewis-Beck, M.S. (1988). *Economics and Elections: The Major Western Democracies*. Ann Arbor: Michigan University Press.

Markus, G.B. (1988). "The impact of personal and national economic conditions on presidential voting: A pooled cross-section analysis." *American Journal of Political Science*, 36, 829–834.

Markus, G.B. (1988). "The impact of personal and national economic conditions on the presidential vote: a pooled cross-sectional analysis." *American Political Science Review*, 32, 137–154.

Mueller, J.E. (1970). "Presidential popularity from Truman to Johnson." *American Political Science Review*, 64, 18–34.

Nannestad, P. and Paldam, M. (1994). "The VP–function: a survey of the literature on vote and popularity functions after 25 years." *Public Choice*, 79, 213–245.

Nannestad, P. and Paldam, M. (1997a). "From the pocketbook of the welfare man: a pooled cross-section study of economic voting in Denmark, 1986–92." *British Journal of Political Science*, 27, 119–136.

Nannestad, P. and Paldam, M. (1997b). "The grievance asymmetry revisited: a micro study of economic voting in Denmark, 1986–92." *European Journal of Political Economy*, 13, 81–99.

Nannestad, P. and Paldam, M. (2000). "What do voters know about the economy? A study of Danish data, 1990–1993." *Electoral Studies*, 19, 363–392.

Nannestad, P. and Paldam, M. (2002). "The cost of ruling. A foundation stone for two theories," in Dorussen and Taylor (eds.) *Economic Voting*. London: Routledge.

Nickelsburg, M. and Norpoth, H. (2000). "Commander-in-chief or chief economist? The president in the eye of the public." *Electoral Studies*, 19, 313–332.

Nordhaus, W.D. (1975). "The political business cycle." *Review of Economic Studies*, 42, 169–190.

Paldam, M. (1981). "A preliminary survey of the theories and findings on vote and popularity functions." *European Journal of Political Research*, 9, 181–199.

Paldam, M. (1991). "How robust is the vote function?" in H. Norpoth, M.S. Lewis-Beck, and J.-D. Lafay (eds.) *Economics and Politics: The Calculus of Support*. Ann Arbor: University of Michigan Press.

Paldam, M. (1997). "Political business cycles," in D.C. Mueller (ed.) *Perspectives on Public Choice- A Handbook. Cambridge*, UK: Cambridge University Press, Chap. 16.

Paldam, M. (2000). "Social capital: one or many? Definition and measurement." *Journal of Economic Surveys*, 14, 629–653.

Paldam, M. "A small country in Europe's integration. Generalizing the political economy of the Danish case." *Public Choice* (forthcoming).

Paldam, M. and Skott, P. (1995). "A rational-voter explanation of the cost of ruling." *Public Choice*, 83, 159–172.

Price, S. and Sanders, D. (1994). "Party support and economic perceptions in the UK: a two-level approach," in D. Brouhton et al. (eds.) *British Elections and Parties Yearbook 1994*. London: Frank Cass, pp. 46–72.

Sanders, D. (2000). "The real economy and the perceived economy in popularity functions: how much do the voters need to know? A study of British data, 1974–97." *Electoral Studies*, 19, 275–294.

Stigler, G.J. (1973). "General economic conditions and national elections." *American Economic Review (Papers and Proceedings)*, 63, 160–180. Special panel on Kramer (1971).

Tinbergen, J. (1956, 1964). Economic Policy: Principles and Design. Amsterdam: North-Holland.

Veiga, F.J. and Veiga, L.G. (2004). "The determinants of vote intentions in Portugal." *Public Choice*, 118(3–4), 341–364.

Wittman, D. (1995). *The Myth of Democratic Failure. Why Political Institutions are Efficient*. Chicago: Chicago University Press.

Chapter 30
Corruption[1]

Susan Rose-Ackerman

Corruption is an archetypal topic for students of public choice. It brings together the private search for economic gain with the government's efforts to supply public goods, correct market failures, and aid the needy. Public choice's insistence on viewing politicians and government bureaucrats as motivated by the same economic interests as private individuals and firms provides a background for understanding why corruption occurs and why it is difficult to combat.

Corruption in my formulation is the *misuse* of public office for private gain. This definition leaves open the issue of just what constitutes misuse, but it recognizes that sometimes public office can legitimately provide private benefits to politicians and bureaucrats. Thus, targeted "pork barrel" projects and special interest legislation are not corrupt. They result from the day-to-day operation of a representative political system. If a legislator works to pass a statute that is favored by his or her legal campaign donors, this is not corrupt even if it violates democratic ideals. Those who seek to discredit government across the board often put the "corruption" label on all kinds of government actions. Although many of these phenomena are indeed proper subjects of study and the loci of reform efforts, it will not help the analysis of democracy to put them all into the corruption pot.

There are several reasons for maintaining a distinction between bribery, fraud, and self-dealing, on the one hand, and quid pro quo politics, on the other. First, a political system that encourages legislators to "bring home the bacon" for their constituents may also be one that encourages voters to monitor their representatives to be sure they are not benefiting personally from their position. Voting systems that limit constituency-based politics may encourage corruption (Kunicová and Rose-Ackerman 2005). Second, strict rules on legal campaign donations may simply drive contributions underground into a corrupt netherworld. Thus, it is valuable to maintain a distinction between legal donations from wealthy interests and illegal, secret gifts. Third, some reform proposals designed to deal with bureaucratic corruption involve the use of legal incentive payments. Mixing financial incentives

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 67–76.

with the provision of public services is not invariably corrupt. Often it is an efficient method of service delivery.

This chapter concentrates on corruption that involves a public official, either a politician or a bureaucrat. However, corrupt incentives can also arise in purely private interactions. Corruption is, in essence, an agency\principal problem. An agent violates the trust of his or her principal through self-enrichment or through illegally enriching a political party. A public official may take a bribe in return for a favorable decision or may simply steal from the state's coffers. Clearly, corporate managers can face similar incentives, and with the growing privatization of former state enterprises, the locus of some forms of corruption will shift into the private sector. Private-to-private corruption has been little studied, but ought to be the object of future work (for one example see Andvig, 1995).

I proceed as follows. The first section outlines the underlying causes of corruption and its consequences from a political-economic point of view. The second section discusses reform options in the light of the discussion in the first section and the broader literature behind the summary presented here. This note provides only a brief overview of both topics. Readers who want to pursue these issues further should consult my two books—Rose-Ackerman (1978) and (1999), review articles by Pranab Bardhan (1997) and Rose-Ackerman (2004), the frameworks presented in Shleifer and Vishny (1993) and Svensson (2005), Klitgaard's illustrative case studies (1988), della Porta and Vannucci's reflections on the Italian case (1999), Manion's study of China (2004), and Glaeser and Goldin's (2006) edited book on American history. These references also include extensive references to the literature. To access current work, the World Bank Institute maintains a website [http://www.worldbank.org/wbi/governance] as does Transparency International (TI), an international nongovernmental organization committed to fighting international corruption, [http://www.transparency.org], the U4 Utstein Anti-Corruption Resource Center in Bergen, Norway [http://www.u4.no].

The Incentives and Consequences of Bribery

I focus on bribery. Ordinary fraud is relatively uninteresting as an analytic matter, and few would argue that stealing from the state is to be encouraged. However, with bribery the story is different. Some economists observe money changing hands and assume that something efficient must be occurring. Some public-choice scholars who favor a minimal state and who view most state actions as illegitimate exercises of power, interpret bribes as a desirable way to avoid the exercise of government power. I want to argue that both these tolerant views are, as a general matter, mistaken, but to do so requires one to understand the incentives for paying and accepting bribes. My basic message is that even if an individual bribe seems to further efficiency or get around an irrational rule, the systemic effects of widespread tolerance are invariably harmful both for the efficient operation of the economy and for the legitimacy of the state.

The government allocates scarce benefits and imposes costs. Individuals and firms may be willing to pay government agents to gain the former and to avoid the latter. Opportunities for corruption arise whenever the officials' actions involve the exercise of discretion and are impossible to monitor perfectly (Klitgaard 1988). The level of benefits under official control can vary from the allocation of a driver's license to the award of a major public works contract. The costs avoided can be a traffic ticket, a multi-million dollar tax bill, or a prison sentence. Bribes can also improve quality, notably by speeding up service delivery or jumping someone ahead in a queue.

The potential bribe revenues available to any individual politician or bureaucrat depend upon his or her monopoly power. If potential bribe payers have noncorrupt alternatives, bribes, if they are paid at all, will be low. If the chance of being caught and punished is high, corruption may be deterred. Thus, one can think of corruption in cost\benefit terms where payoffs will be deterred if at least one side of the potential deal faces costs that exceed the benefits. If expected penalties increase more than in proportion to the size of the bribe, only small bribes may be paid and accepted. Conversely, if penalties do not rise in proportion to benefits, small bribes are deterred, and large bribes are unaffected (Rose-Ackerman 1978; 1999).

The mere existence of corrupt opportunities, however, says nothing about their welfare implications. In discussing this issue, it is important to recognize that the level of bribe payments is likely to be a poor measure of their social cost. Sometimes very small bribes have large consequences. Bribes may be low, not because the value of the quid pro quo is low, but because the bribe payer has bargaining power relative to the official. For example, if a majority-rule legislature with weak parties is bribed to approve a law favored by a particular firm, no individual politician has much bargaining power; he or she can easily be replaced by another person formerly outside the corrupt coalition (Rose-Ackerman 1978, Rasmusen and Ramseyer 1994). Thus, my focus is not on situations where bribes are high but on those cases where the social costs are severe.

One might suppose that if a government has scarce benefits to distribute, say a number of business permits, then corruption will distribute them to those with the highest willingness-to-pay, and the winners will be the most efficient restaurateurs. There are several responses to this claim. First, corrupt markets are inefficient compared with the aboveboard sale of licenses. Bribe-prices are secret, and entry may be blocked. Thus, the government should simply legally sell the scarce rights if its goal is to allocate the service to those who value it the most in dollar terms. Second, the basic purposes of some public programs would be violated by sales to the highest bidders. For example, selling places in public universities and in subsidized housing would undermine the basic goals of those programs. Third, toleration of corruption gives officials an incentive to engage in the creation of more scarce benefits in order to create more corrupt opportunities. For example, corrupt contracting officials have an incentive to support wasteful public projects designed to make payoffs easy to hide.

Similar points can be made about bribes paid to avoid the imposition of costs. Clearly, if a regulation is onerous and inefficient, then paying for an exemption seems efficient. However, permitting such individualized law compliance can be

very harmful. First, profit-maximizing firms and individuals will not distinguish between socially efficient and socially inefficient rules. They will want to be exempted from all of them. The rules will only be enforced against those with a low willingness to pay. This includes not just those for whom the rule is not costly but also poor households and marginal businesses. In the case of tax collection, those exempted from taxes generate higher bills or lower services for others. Selective exemption on the basis of willingness to pay is inefficient and unfair. Second, officials will seek to create even more restrictive rules so that they can be paid to decline to enforce them. Empirical work suggests that in countries where corruption is high, red tape is high, and managers spend considerable time dealing with public officials (Kaufmann 1997). In a survey in the countries undergoing a transition from socialism, those business people who paid bribes generally spent *more* time dealing with officials than those who paid no bribes (Fries, Lysenko and Polanec 2003). Thus, even if each individual corrupt decision is rational for the bribing firm, the overall costs of doing business in society are high. Investment and entrepreneurship are discouraged.

The costs of corruption are not limited to its impact on the efficacy of public programs taken one by one. In addition, endemic corruption has implications for the legitimacy of the state in the eyes of its citizens. In highly corrupt states, where both day-to-day interactions with officials and high-level deals are riddled with payoffs, people often express great cynicism about political life (Anderson and Tverdova, 2003).

This can lead to vicious spirals. The theoretical work on corruption has produced a number of multiple-equilibria models where both high corruption and low corruption solutions exist (Bardhan, 1997; Rose-Ackerman, 1999, pp 107–108, 124–125). Some countries, particularly a number of the former socialist countries, illustrate these pathologies. To give a flavor of these models consider two variants. First, suppose that there is a fixed supply of law enforcement personnel. If very few transactions are corrupt, the enforcers can catch most of the illegal deals, thereby encouraging more people to be honest in the next round and so forth. If most are corrupt, the law enforcement authorities are spread very thin and only catch a few wrongdoers. This encourages more to enter the corrupt arena next period and so on in a vicious spiral. Similar results occur if we assume that the moral stigma of corruption is a function of the number of others who engage in it. If most are corrupt, the stigma is low, and next period more shift to the corrupt side, and so forth. Second, another kind of spiral can affect the character of those who become politicians or bureaucrats. If most officials are corrupt, this will discourage honest people from working for the government and encourage the dishonest to apply, making the government even more corrupt. If government work makes one rich, those who want to get wealthy choose the public sector and do not become entrepreneurs. Their corruption creates a costly environment for business that further discourages private business activities. This self-selection mechanism can produce an equilibrium in which the dishonest and the greedy have disproportionately chosen public sector employment.

Empirical work has begun to shed light on some of the costs of corruption outlined above. Research on corruption is difficult because the perpetrators seek to

keep their transactions secret. Nevertheless, scholars have begun to analyze and measure the impact of corruption on economic and political phenomena and to explain how political and economic conditions contribute to corruption. Two types of research are important: studies based on cross-country data and those that concentrate on detailed micro-level experiences. A recent collection includes both type but emphasizes the latter (Rose-Ackerman 2006). The collection reflects my view that future research should concentrate more heavily at the micro level to try to find out exactly how corrupt systems operate so that policy can be designed to disrupt them.

Nevertheless, the cross-country research has been important in raising the salience of the issue and showing that high levels of corruption are harmful to growth and development and that corruption is the result of weak economic and political institutions. The cross-country research uses data that measure perceptions of corruption, such as the composite Transparency International index, developed by Johann Graf Lambsdorff, or the World Bank Institute's recalculation using similar data. The perceptions are mostly those of international business people and country experts. Studies using these data have found that high levels of corruption are associated with lower levels of investment and growth, and that foreign direct investment is discouraged (Mauro 1995; Wei 2000). Highly corrupt countries tend to underinvest in human capital by spending less on education and to overinvest in public infrastructure relative to private investment (Mauro, 1997; Tanzi and Davoodi 1997). Corrupt governments lack political legitimacy and hence tend to be smaller than more honest governments, everything else equal (Johnson, Kaufmann, McMillan, and Woodruff 2000). Corruption reduces the effectiveness of industrial policies and encourages business to operate in the unofficial sector in violation of tax and regulatory laws (Kaufmann 1997). Turning the causal story around, recent research suggests that autocracies tend to be more corrupt than democracies, but that democracy is not a simple cure. Within the universe of democracies, corruption is facilitated by features of government structure such as presidentialism, closed-list proportional representation, and federalism (Kunicová and Rose-Ackerman, 2005; Treisman 2000).

These are important findings, but they are limited by the aggregated nature of the data. Each country is treated as a single data point that is more or less "corrupt." This work shows that corruption is harmful but says little about the precise mechanisms. To counter this weakness, two new types of research are underway: detailed questionnaires that target households, businesses, and public officials; and what might be called "econometric case studies." The questionnaires permit researchers to explore people's actual experiences. The case studies help one understand how corrupt sectors operate and how malfeasance might be controlled.

Here are some examples of the research I have in mind. Several studies questioned small- and medium-sized businesses about the costs of corruption and red tape. Other researchers have used questionnaires and focus groups to examine household attitudes and behavior. Researchers have studied countries as diverse as those in sub-Saharan Africa and in Central and Eastern Europe. Some of the most comprehensive are a study of four countries in Central and Eastern Europe by William Miller, Åse Grødeland, and Tatyana Koshechkina (2001) and work that

focuses on the business environment in the same region by Simon Johnson, Daniel Kaufmann, John McMillan, and Christopher Woodruff (2000). This research complements the World Bank Institute's work on "state capture" and administrative corruption in post-socialist countries (Hellman, Jones, and Kaufmann, 2000).

Sectoral studies are represented by work on how corruption limits the performance of the judiciary in Latin America (Buscaglia 2001; Buscaglia and Dakolias 1996). Other examples are Dean Yang's studies of preshipment inspection as a tool to control corruption in customs, Benjamin Olken's (2007) study of corruption in the provision of subsidized rice in Indonesia, and research by Rafael di Tella and Ernesto Schagrodsky (2002) on the benchmarking of product prices in the hospital sector in Argentina that shows how monitoring and civil service pay reform can go hand in hand. As an example of research that can make a difference, consider Ritva Reinikka and Jakob Svensson's (2006) documentation of the severe leakage of federal funds meant for local schools in Uganda. Their study led to a simple, information-based reform that had positive results.

These contributions are very diverse in topic and methodology, but they all share an interest in using detailed data to understand both how corrupt systems operate and which policies have promise. Only if one looks at the fine structure of political and economic systems, can one go beyond a showing that corruption is harmful to an understanding of the way it operates in different contexts. Given that knowledge, reform programs can attack corruption where it has the worst effects.

Reform

Reform strategies influenced by the public-choice perspective attack the problem of corruption from several directions: program redesign, law enforcement, and improved government performance and accountability. Before presenting this mixture of reform options, however, I begin with a solution that is favored by some public-choice scholars. Many such analysts accept the claim that corruption is harmful. However, they argue that the solution should be not the reform of public programs, but a reduction in the size of government. They argue that the best way to avoid corruption is to shrink government and rely on the market. Of course, this will sometimes be true, but it is not a general solution and would be risky if employed across the board. The most obvious problem with this recommendation is that it misses the benefits of some, even poorly operating, public programs. Programs to limit external costs, correct for information failures, produce public goods, or aid the needy have no effective private market counterparts. Free rider problems plague efforts at private provision. Furthermore, if a program is reduced in size but not eliminated, corruption may increase instead of decreasing. To see this, consider a program to provide public housing to the needy. A cut in the program by half creates scarcity and hence the competition for places. Bribes may increase.

Another form of government "load shedding" has similar difficulties. Privatization is justified as a way of introducing market discipline into the operation of formerly

state-owned firms. Competitive pressures and the need to raise capital in the private market will squeeze out waste and encourage a focus on consumer satisfaction. Unfortunately, privatization does not always imply the creation of competitive markets. Sometimes the process of turning over assets has itself been corrupted by collusion between powerful private and public interests. This sometimes implies that public firms are sold too cheaply to insiders and that the terms of the deal give the new owners access to monopoly rents. Corruption in the privatization process in some countries is analogous to corruption in large-scale public procurements—powerful politicians and business interests gain at the cost of ordinary citizens (Manzetti 1999). Citizens lose both because the benefits to the state coffers are lower than they should be and because the benefits of expanding the role of competitive markets are lost. Thus, overenthusiastic efforts to limit the role of government should be avoided, and the cutbacks that are carried out should be carefully designed to avoid the problems outlined here.

If a country faces a vicious spiral of corruption, such as outlined above, this would seem the best case for the "load shedding" solution. The government is in a dysfunctional low-level trap where piecemeal reform will be ineffective. The state needs a major overhaul in law enforcement and in the recruitment of personnel. However, a simple attempt to shrink the state is unlikely to be effective because it can create a chaotic situation in which a lawless free-for-all replaces the corruption that went before. A new kind of corruption and self-dealing may arise, that is based on the attempt to establish some kind of certainty in a situation of fluidity and chaos.

If corruption cannot be countered by single-minded efforts to limit the size of government, then one must consider ways to reform government from within and to limit the willingness of citizens and firms to pay bribes. Any actual program needs to be adapted to the conditions in a particular country, but the broad outlines can be identified.

Anticorruption policies can increase the benefits of being honest, increase the probability of detection and the level of punishment, reduce the corrupt opportunities under the control of public officials, and increase the accountability of government to its citizens. The incentives for corruption are influenced by:

- the level of benefits and costs under the discretionary control of officials,
- the formal laws designed to combat defining corruption, bribery, and conflicts of interest, and to regulate finance spending,
- the credibility of law enforcement against both those who pay and those who accept bribes,
- the conditions of civil service employment, and the performance incentives officials face,
- the extent of auditing and monitoring within government,
- the ability of citizens to learn about government activities, file complaints, and obtain redress, and
- the level of press freedom and the freedom of individuals to form nongovernmental organizations.

I focus on four broad categories: reductions in the discretion and monopoly power of government officials, enforcement of anticorruption laws, civil service reform, and increased accountability to citizens.

Reducing the Incentives for Payoffs

The most basic reforms are those that reduce the level of benefits under the control of public officials. As I noted above, the most obvious option is simply to eliminate laws and programs that are permeated with corruption. If the state has no authority to restrict exports or license businesses, no one will pay bribes in those areas. If a subsidy program is eliminated, the bribes that accompanied it will disappear as well. If price controls are lifted, market prices will express scarcity values, not bribes.

In general, any reform that increases the competitiveness of the economy will help reduce corrupt incentives. Thus policies that lower the controls on foreign trade, remove entry barriers for private industry, and privatize state firms in a way that ensures competition will all contribute to the fight against corruption. The main problem here is that the removal of barriers can create a backlash with those who benefited from controls using corruption to seek favors through the backdoor. For example, a firm might pay off officials to harass a competitor even if the rules are meant to limit such behavior.

Furthermore, many regulatory and spending programs have strong justifications and ought to be reformed, not eliminated. Corruption in the collection of taxes obviously cannot be solved by failing to collect revenue. One solution is to clarify and streamline the necessary laws in ways that reduce official discretion. Rules could be made more transparent with publicly provided justifications. Governments might favor simple nondiscretionary tax, spending, and regulatory laws as a way of limiting corrupt opportunities. Clear rules of proper behavior could be established, so violations can be noticed even if the bribery itself is not. Where possible, procurement decisions could favor standard off-the-shelf items to provide a benchmark and to lower the cost of submitting a bid. Obviously, the value of such reforms depends upon the costs of limiting the flexibility of public officials (Anechiarico and Jacobs, 1996). Sometimes a certain risk of corruption will need to be tolerated because of the benefits of a case-by-case approach to program administration. Transparency and publicity can help overcome corrupt incentives even in such cases, but only if the systems of accountability discussed below exist. If they do not, simple, clear rules can simply permit a top ruler more effectively to extract payoffs. This is just one example of the importance of viewing reform in the context of the entire political-economic environment.

Economists have long recommended reforming regulatory laws in such areas as environmental protection by introducing market-based schemes that limit the discretion of regulators. Analysts also recommend user fees for scarce government services. These reforms have the additional advantage of removing corrupt incentives by replacing bribes with legal payments. The sale of water and grazing

rights, traceable pollution rights, and the sale of import and export licenses can improve the efficiency of government operations while limiting corruption.

Finally, administrative reforms may lower corrupt incentives. Corruption is often embedded in the hierarchical structure of the bureaucracy. Low level officials collect bribes and pass a share on to higher level officials perhaps in the form of an up-front payment for the job itself. Conversely, higher ups may organize and rationalize the corrupt system to avoid wasteful competition between low-level officials. The top officials may then share the gains of their organizational ability with subordinates, perhaps using them to run errands, transfer funds, and do other risky jobs that expose them to arrest. To break such patterns may require a fundamental reorganization effort.

One possibility is the introduction of competitive pressures within government to lower the bargaining power of individual officials. If bribes are paid for such benefits as licenses and permits, which are not constrained by budgetary limits, overlapping, competitive bureaucratic jurisdictions can reduce corruption. Because clients can apply to any of a number officials and can go to a second one if the first turns him down, no one official has much monopoly power. Thus no one can extract a very large payoff. For qualified clients, bribes will be no larger than the cost of reapplication. Unqualified clients will still pay bribes, but even they will not pay much so long as they too can try another official (Rose-Ackerman 1978). If all officials are corrupt, the outcome is stable. However, if some establish an honest reputation, applicants will prefer those officials, thereby reducing the gains to the corrupt. This reduction in benefits may induce some marginal officials to shift to being honest, further reducing the benefits to the remaining corrupt officials and so on. A small number of honest officials can overturn a corrupt system if congestion is not a serious problem. Honesty may drive out dishonesty even if only a few officials are honest on principle (Rose-Ackerman 1978). If, instead, those who pay bribes are unqualified, the honesty of some officials increases the gains to those who are corrupt, inducing more to become corrupt.

When officials, such as police officers, can impose costs, another type of overlapping jurisdiction model should be considered. Police officers seeking to control illegal businesses can be given overlapping enforcement areas. That way gamblers and drug dealers will not pay much to an individual policeman since a second one may come along later and also demand a payoff. The first one is simply unable to supply protection. Bribes may fall so low that it is not worthwhile for police officers to risk taking them. This system may work better if the law enforcement officers belong to different police forces—state or federal, for example. Then collusion between officers to defeat the system will be less likely (Rose-Ackerman, 1978).

Alternatively, consider the losers in corrupt transactions. The state could introduce ways for the potential losers to appeal unsatisfactory decisions. Sometimes bribe payers view themselves as losers who would be better off in an honest world. They feel themselves to be the victims of extortion. Such bribe payers are potential allies in an anti-corruption effort, who will cooperate in efforts to eliminate payoffs. Conversely, in other cases bribery makes both payer and receiver better off with respect to a no-bribery world. Thus control incentives must rest with outsiders not in on the corrupt deal (for example, disappointed bidders, taxpayers, consumers).

The existence of losers, such as disappointed bidders, with a large stake in the outcome can facilitate efforts to limit corruption.

Anticorruption Laws and Credible Law Enforcement

A basic condition for corruption control is a viable legal framework that enforces the law without political favoritism or arbitrariness. The goal is both to deter those tempted to engage in corrupt acts and to educate the public to resist criminal conduct by officials. Tough laws are not sufficient. Many highly corrupt countries have exemplary formal statutes that have no real meaning because they are seldom enforced. A country serious about reform must have effective investigation and prosecution bodies and a well-functioning judicial system that is not itself corrupt. Because corruption is a two-sided offense, the law must specify the status of both those who make payments and those who receive them. If just one of the parties can be deterred, that is sufficient to prevent the deal from going through.

Designing an optimal deterrence strategy raises a seeming paradox. The more severe the penalties for corruption faced by officials, the lower the incidence of corruption, but the higher the bribes. If the risk of detection is high, officials must receive a high return in order to be willing to engage in bribery. One way around such a result is an expected penalty function that is an increasing function of the size of the bribe (Rose-Ackerman 1978, pp 109–135). Conversely, if penalties on bribe payers have deterrent effects, this will lower the demand for corrupt services and the level of bribes at the same time.

An independent judiciary or some other kind of independent tribunal is a necessary condition for the use of law enforcement to check official malfeasance. This is a serious problem in many countries where the judicial system is backlogged and some judges are corrupt. Prosecutors, whether they are formally in the executive branch, as in the United States, or part of the judiciary, as in Italy, must be able have the independence to pursue corruption allegations and need to be able to reward those who report on corrupt deals with lowered charges and penalties. Some countries have had success with independent anticorruption commissions or inspector generals reporting only to the chief executive or the parliament. These can be useful responses, but a single-minded focus on law enforcement is unlikely to be sufficient if the incentives for corruption are deeply imbedded in the structure of public programs and if law enforcement efforts can be diverted to harass political opponents.

The Civil Service

Many developing countries have very poorly paid civil servants. Although at independence most former colonies inherited civil service pay scales that exceeded private sector wages, this advantage has eroded over time. Wages relative to private sector wages have fallen in countries in transition in Eastern Europe and the former Soviet Union. The pattern varies across countries and over time. In some parts of

the developing world public sector pay is so low that officials must supplement their pay with second jobs or payoffs. Some work suggests that there is a negative correlation between civil service wages (relative to private sector wages) and the level of corruption (Van Rijckegham and Weder 2001).

If officials are paid much less than people with similar training elsewhere in the economy, only those willing to accept bribes will be attracted to the public sector. Civil service pay should be set at least equal to equivalent positions in the private sector in order to make it possible to recruit based on merit and to permit those selected to serve without resorting to corruption. If the benefits under the control of officials are very valuable, however, parity may not be sufficient. Instead, civil service wages may need to be set above the going private sector wage with generous benefits, such as pensions, that will be received only if the worker retires in good order. This strategy, however, must be combined with an effective monitoring system. There must be a transparent, merit-based system of selecting civil servants or else people will pay the powerful to be allotted desirable government jobs.

Pay reform is necessary, but not sufficient. Penalties must be tied to the marginal benefits of accepting payoffs. In cases where corruption's only efficiency cost stems from its illegality, the payments should be legalized. In the design of such systems, however, it is important to avoid giving monopoly power to bureaucrats that they can use to extract increased levels of rents.

Public Accountability

Corruption can be checked by structures that create independent sources of power and information inside and outside the government. Although not sufficient, taken by themselves, these options complement other reform strategies by reducing corrupt opportunities and increasing the risks of paying and accepting payoffs. There are several linked aspects in a system of public accountability over and above the checks provided by periodic democratic elections.

- Outsiders, such as ordinary citizens or the media, can obtain *information* about how the government is operating and have a way of expressing their displeasure about general policies. Nongovernmental organizations can organize easily and face few legal hurdles. They may even be subsidized.
- The structure of government includes *guarantees that protect the individual against the state*. Government actions may be checked by a specific Bill of Rights that limits state power, and individuals can appeal attempts to extort bribes. The legal system provides protection and perhaps rewards to individuals who come forward to "blow the whistle" on corrupt practices, but the state is also constrained by legal rules that protect the accused.
- *Higher level governments and international organizations* can use what leverage they have to constrain the behavior of individual governments.
- The threat of *exit* can be a powerful constraint on governments, reducing corrupt opportunities and limiting the scope for waste.

First, the private sector, particularly an independent media, can be an important check on the arbitrary exercise of power by government (McMillan and Zoido, 2004), but only if the government provides information, if the press is not controlled, and if people can organize into associations. Accountability to the public requires both that individuals can find out what the state is doing and that they can use this information to hold public actors accountable. Governments must publish budgets, revenue collections, statutes and rules, and the proceedings of legislative bodies. Financial data should be independently audited. Secret funds available to chief executives and top ministers are an invitation to corruption. Procurement regulations must keep the process open and fair. Scandals frequently occur because top officials overrule tender boards or because lower level officials operate without formal controls on their purchasing decisions.

Freedom of information acts in the United States and in a number of European countries are an important precondition for effective public oversight. These laws permit citizens to request information as members of the public without showing that their own personal situation will be affected (Ackerman and Sandoval 2006).

Finding out what is happening is of little value however, unless people can use their knowledge to influence government. Individuals face a familiar free rider problem in seeking to control political and bureaucratic processes and to limit malfeasance. Information may be, in principle, available, but no one may have an incentive to look at it. Laws that make it easy to establish associations and nonprofits will help. For example, Transparency International has local chapters that carry out a range of activities including participation in Integrity Workshops, sometimes organized with the help of aid agencies. These workshops bring together concerned people from both the public and the private sectors to discuss the problem of corruption. Nonprofit organizations can carry out and publish public opinion surveys that reveal public attitudes toward government services. An alternative to NGO surveys of service users is the creation of "hot lines" so that citizens can complain directly to the government. The information from such complaint mechanisms will be less systematic than a survey and may well be self-serving, but hotlines provide a means of making a complaint without the necessity of establishing an organization. This method will only be successful, however, if those who complain can either do so anonymously or are not fearful of reprisals. Furthermore, if the complaints concern individuals, they must have a credible way of defending themselves against false accusations.

The second aspect of accountability is the way the government structure protects individuals against the state. The forms of administrative law and the protection they provide to individuals are of critical importance. If an official tries to extort a bribe from individuals or firms, do they have any recourse? Obviously, if the bribe is to be paid to permit illegal activities or to soften a legal regulation or tax assessment, the answer is no. Corruption of this type is unlikely to be revealed by the parties to the deal unless they have been arrested and are seeking to mitigate their punishment. However, those who face bribe demands as a condition for obtaining a legal benefit may not go along with the demand if they can appeal to an honest forum, such as an appeals board within the agency or the courts. In order to make appeals worthwhile, however, the processes must not only be honest, but also speedy and efficient.

The Ombudsman represents one route for citizen complaints. Many countries have established Ombudsmen to hear complaints of all kinds, not just those related to malfeasance. These offices can help increase the accountability of government agencies to ordinary citizens, but they are seldom a way to uncover large-scale systemic corruption and most have no authority to initiate lawsuits.

Ombudsmen and other complaint mechanisms are insufficient if people are unwilling to complain. Reporting the peculations of others can be dangerous. Thus, governments should consider promulgating whistleblower statutes that protect and reward those in the public and the private sector who report malfeasance. However, whistleblower protection is obviously pointless unless the prosecutorial system follows up, the courts are incorruptible and relatively efficiently run, and the penalties are severe enough to deter potential offenders.

The third check on corruption can arise from intergovernmental relations. In a federal system, the national government can constrain the states, and the states, the localities. Similarly, institutions operating internationally may provide a check on national governments. This kind of leverage has problematic aspects since those who exercise it can make no straightforward claim to represent the interests of the affected citizens. There are two cases, however, in which such actions may be justified. First, corruption and waste frequently have cross-border consequences. Corrupt politicians or those engaged in legal joint ventures with private firms may try to use their political power to restrict commerce across state borders. Internationally, officials working in collaboration with corrupt business firms harm the prospects of honest businesses. Second, state and local governments may be under the control of narrow elites that use the apparatus of government for personal gain. Although both oversight from above and competition between jurisdictions for investment resources limit corrupt possibilities at the local level, they do not eliminate them. In fact, cross-country empirical work suggests that federal states are, on balance, more corrupt than unitary states suggesting that the negative effects outweigh the positive (Treisman 2000).

Exit, the final constraint on corruption, has the advantage of not requiring a concerted organizational effort. In a world with many coequal governments, the corruption and ineffectiveness of government officials is limited by the ability of constituents and business firms to go elsewhere. Multinational firms trying to decide where to locate a manufacturing plant can limit bribe demands by locating several feasible sites. Residents of a village whose officials extract large payoffs for routine services can move elsewhere. The mobility of people and businesses clearly limits the ability of officials to extract payoffs for services to which one is entitled.

Mobility, however, is not always helpful. It will make it more difficult for an individual jurisdiction to control undesirable behavior. Suppose, for example, that a city government has installed an honest police force that cracks down on illegal gambling. The gamblers may simply move to a friendly suburb that they can control and establish their business there. Several examples of such phenomena exist in U.S. urban areas. The ease with which funds can cross national borders, coming to rest in various "financial paradises" is another example of how multiple, competing jurisdictions can make control of corruption, fraud, and tax evasion more, not less, difficult. Thus interjurisdictional competition should be encouraged when it reduces

the economic rents available for corrupt distribution and helps control waste, but should be limited when it facilitates the illegal behavior that corruption often makes possible or requires.

A system of public accountability implies that once a law or regulation is put in place, individuals and groups both inside and outside government have the ability to find out how it is being administered, to complain, and to set in motion a legal or political enforcement process. To be a meaningful anticorruption check, however, knowledge must be combined with the existence of institutions that can take effective action both to promulgate new laws and to enforce existing ones.

Conclusions

Corruption has a moral dimension, but it can be understood and combated through the application of political-economic principles. A first step in the understanding of corruption is the documentation of the incentives for private gain built into political and bureaucratic processes. Next is an evaluation of the social costs when officials and private citizens succumb to these incentives. Part of the reform agenda involves explaining the social harm of corruption and trying to change a culture of tolerance both within government and in the citizenry and the business community (Rose-Ackerman, 2002). Moral suasion may work if backed up by concrete arguments for why corruption is harmful to society. Reformers do not simply point to corruption and appeal for people to change their behavior; rather they demonstrate that reducing corruption provides real gains, not just symbolic victories. The key point is to encourage people to look beyond the net gains from any particular corrupt deal to see how tolerance of corruption has negative systemic effects.

However, as public-choice theory teaches, most people will not behave well simply because they are told that such actions are in the public interest. A change in behavior needs to be in their interest as well. A political-economic approach can go beyond documenting the costs of corruption to suggest ways to lower its incidence and impact. Although reforms in law enforcement and in internal monitoring are part of the story, the most important lessons of a political-economic approach are its recommendations to turn attention to the redesign of individual public programs, on the one hand, and to ways to increase government transparency and accountability on the other. That strategy both reduces the corrupt incentives facing bribe payers and recipients and facilitates effective public oversight by the population.

References

Ackerman, J. M., and I. E. Sandoval-Ballesteros. 2006. "The Global Explosion of Freedom of Information Laws," *Administrative Law Revew* 58, 85–130.
Anderson, C. J., and Y. V. Tverdova. 2003. "Corruption, Political Allegiances, and Attitudes Toward Government in Contemporary Democracies," *American Journal of Political Science* 47, 91–109.

Andvig, J. C. 1995. "Corruption in the North Sea Oil Industry: Issues and Assessments," *Crime, Law, and Social Change* 28, 289–313.

Anechiarico, F. and J. B. Jacobs. 1996. *The Pursuit of Absolute Integrity: How Corruption Control Makes Government Ineffective*, Chicago: University of Chicago Press.

Bardhan, P. 1997. "Corruption and Development: A Review of Issues," *Journal of Economic Literature* 35, 1320–1346.

Buscaglia, Jr., E. 2001. "An Analysis of Judicial Corruption and Its Causes: An Objective Governance-Based Approach," *International Review of Law and Economics* 21, 233–240.

Buscaglia, Jr., E. and M. Dakolias. 1996. *Judicial Reform in Latin American Courts: The Experience in Argentina and Ecuador*, World Bank Technical Paper No. 350, Washington DC: World Bank.

Della Porta, D. and A. Vannucci. 1999. *Corrupt Exchange*, New York: Aldine.

Di Tella, R. and E. Schargrodsky. 2002. "The Role of Wages and Auditing During a Crackdown on Corruption in the City of Buenos Aires," in D. della Porta and S. Rose-Ackerman, eds. *Corrupt Exchanges: Empirical Themes in the Politics and Political Economy of Corruption*, Frankfurt: Nomos Verlag.

Fries, S., T. Lysenko, and S. Polanec. 2003. *The 2002 Business Enterprise and Enterprise Performance Survey: Results from a Survey of 6,100 Firms*, Working Paper 84, European Bank for Reconstruction and Development, London, http://www.ebrd.com/pubs/econ/workingp/84.pdf.

Glaeser, E. L. and C. Goldin, eds. 2006. *Corruption and Reform: Lessons from America's Economic History*, Chicago: NBER and Chicago University Press,

Hellman, J. S., G. Jones, and D. Kaufmann. 2000. *"Seize the State, Seize the Day": State Capture, Corruption, and Influence in Transition*, Policy Research Working Paper 2444, World Bank: Washington DC.

Johnson, S., D. Kaufmann, J. McMillan, and C. Woodruff. 2000. "Why Do Firms Hide? Bribes and Unofficial Activity after Communism", *Journal of Public Economics* 76, 495–520.

Kaufmann, D. 1997. "The Missing Pillar of Growth Strategy for Ukraine: Institutional and Policy Reforms for Private Sector Development," in P. K. Cornelius and P. Lenain, eds. *Ukraine: Accelerating the Transition to Market*, Washington: IMF, pp. 234–275.

Klitgaard, R. 1988. *Controlling Corruption*, Berkeley CA: University of California Press.

Kunicová, J., and S. Rose-Ackerman. 2005. "Electoral Rules and Constitutional Structures as Constraints on Corruption," *British Journal of Political Science* 35, 573–606.

Manion, M. 2004. Corruption by Design: Building Clean Government in Mainland China and Hong Kong, Cambridge MA: Harvard University Press.

Manzetti, L. 1999. *Privatization South American Style*, Oxford: Oxford University Press.

Mauro, P. 1995. "Corruption and Growth," *Quarterly Journal of Economics*, 110, 681–712.

Mauro, P. 1997. "The Effects of Corruption on Growth, Investment, and Government Expenditure: A Cross-Country Analysis" in K. A. Elliott, ed., *Corruption and the Global Economy*, Washington: DC: Institute for International Economics, pp. 83–108.

McMillan, J. and P. Zoido. 2004. "How to Subvert Democracy: Montesinos in Peru," *Journal of Economic Perspectives* 18(4), 69–92.

Miller, W. L., Å. Grødeland, and T. Y. Koshechkina. 2001. *A Culture of Corruption: Coping with Government in Post-Communist Europe*, Budapest: Central European University Press.

Olken, B. A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy* 115, 200–249.

Rasmusen, E., and M. Ramseyer. 1994. "Cheap Bribes and the Corruption Ban: A Coordination Game among Rational Legislators," *Public Choice* 78, 305–327.

Reinikka, R., and J. Svensson, 2006. "How Corruption Affects Service Delivery and What Can Be Done About It," in S. Rose-Ackerman, eds. pp. 441–456.

Rose-Ackerman, S. 1978. *Corruption: A Study in Political Economy*. New York: Academic Press.

Rose-Ackerman, S. 1999. *Corruption and Government: Causes, Consequences and Reform*. Cambridge: Cambridge University Press.

Rose-Ackerman, S. 2002. "'Grand' Corruption and the Ethics of Global Business," *Journal of Banking and Finance* 26, 1889–1918.

Rose-Ackerman, S. 2004. "Governance and Corruption," in B. Lomborg, ed., *Global Crises, Global Solutions*. Cambridge: Cambridge University Press, pp. 301–344.

Rose-Ackerman, S. ed. 2006. *International Handbook on the Economics of Corruption*. Cheltenham: Edward Elgar Publishing.

Shleifer, A., and R. Vishny. 1993. "Corruption," *Quarterly Journal of Economics* 108, 599–617.

Svensson, J. 2005. "Eight Questions about Corruption," *Journal of Economic Perspectives* 19, 19–42.

Tanzi, V., and H. Davoodi. 1997. "Corruption, Public Investment and Growth," IMF Working Paper WP/97/139, Washington DC: International Monetary Fund.

Treisman, D. 2000. "The Causes of Corruption: A Cross-National Study," *Journal of Public Economics* 76, 399–457.

Van Rijckeghem, C., and B. Weder. 2001. "Bureaucratic Corruption and the Rate of Temptation: Do Wages in the Civil Service Affect Corruption, and by How Much?," *Journal of Development Economics* 65, 307–331.

Wei, S. 2000. "How Taxing is Corruption on International Investors?", *Review of Economics and Statistics*, 82, 1–11.

Chapter 31
Environmental Politics[1]

Hannelore Weck-Hannemann

Building on the seminal contributions by Pigou (1920), Coase (1960) and Baumol and Oates (1971), economists have extensively explored the role that economic incentives might play in bringing a more efficient allocation of natural resources. The theory of environmental economics suggests that pricing instruments are an adequate means to internalize external costs. More specifically, there is widespread agreement within the scientific community that from a theoretical point of view pricing instruments are preferable to alternative measures, owing to their efficiency advantages (Frey et al., 1985). However, though economists see pricing instruments as an attractive policy tool, most attempts to introduce economic incentives in environmental policy have failed, and the acceptance of these mechanisms in the political debate is still rather limited (Hahn, 1989; Schneider and Weck-Hannemann, 2005).

There are many possible reasons why incentive instruments as a means to internalize external costs have been rarely applied in the past. It certainly would be too simple just to refer to imperfect information on the part of decision-makers about the advantages of incentive-based instruments. On the contrary, there seem to be good reasons why politicians, voters, bureaucrats, and/or representatives of interest groups are rather reluctant to favor price instruments on a large scale in environmental politics.

It is the purpose of the political economy of environmental policy to point out these reasons by concentrating on the process of political decision-making and the incentives of the political agents to implement alternative environmental instruments. Public-choice methodology can be used to explain the discrepancy between economic theory and political reality also in environmental politics. Though public-choice theory has been applied extensively in politico-economic modeling of popularity and voting functions, in analyzing political business cycles and in explaining rent-seeking behavior and the persistence of protectionism, for example, it is relatively less developed in environmental economics. Originated by the seminal

---

[1] This paper is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 91–96.

study of Buchanan and Tullock (1975), the literature on the political economy of environmental policy has mainly focused on the comparative analysis of alternative policy measures and their chances for implementation, respectively (Dewees 1983; Hahn 1990; Downing 1991; Pearson 1995; Congleton 1996; Dijkstra 1999; Kirchgässner and Schneider 2003; Frey 2003; Schneider and Weck-Hannemann 2005). Besides, the public-choice approach has been applied to analyze international environmental problems (Schulze and Ursprung 2001; Bommer 1998; Kirchgässner 1999) as well as more specific topics as, e.g., the failure to cope with global warming and natural disasters (see, e.g., Congleton 2006; Schwarze and Wagner 2007).

In their initial study, Buchanan and Tullock (1975) argue that direct control measures have better chances to be favored and implemented in the political process than incentive-based instruments like taxes on pollution. More generally, it is argued in the public-choice literature on environmental politics that incentive oriented instruments are neither in the interest of the decision-makers on the supply side nor are they favored by the most influential groups of voters on the demand side in the political market. It is hypothesized that if any instrument of environmental policy is used at all, the main actors of environmental policy have a strong interest to apply command and control measures instead of incentive-based instruments.

More recently, however, ecological taxes as well as tradable permits have become more popular and voluntary agreements have been implemented. According to the Kyoto protocol, market-based instruments are intended to play a more prominent role also in international environmental policy. Kirchgässner and Schneider (2003, p. 372), therefore, conclude that "while we are still far away from general acceptance and widespread application of market based environmental instruments, the situation has changed at least somewhat". Consequently, it has to be asked whether the old diagnosis by Robert Hahn (1989) and the papers in the public-choice tradition still holds, i.e., that the patients do not follow the doctor's orders in that environmental policy is dominated by command and control measures and, if applied at all, market-based instruments deviate from the therapy that economists typically prescribe.

Generally, public-choice theory not only intends to analyze how the agents in the political sector (i.e., in particular, politicians and public bureaucrats) influence the state of the economy but also how the state of the economy in turn influences voters' preferences and thereby the evaluation of policies and parties. The level and structure of public interventions are determined endogenously in the political market for state interventions. In order to analyze the process of environmental policy, it is important to identify the various actors involved and their interests and impact in the political decision-making process, respectively. The usual way is to single out four groups of actors which are examined in more detail, i.e., voters, politicians, public bureaucrats, and interest groups representing the private sector.

Political economists view the policy measures that governments and parliaments adopt as outcomes of an exchange process. Elected officials supply the policies that voters and interest groups demand. In exchange for regulation, politicians receive

votes, money, and information. From a political economy perspective, it is useful to think about the negative externalities of private production and consumption as transfers to specific groups which are allowed to make use of resources without bearing the full opportunity costs. The introduction of alternative environmental policies then increases transfers to some groups and decreases the transfers to others. Whether or not it is possible to devise a pricing scheme that will find political acceptance not only depends on the changes in welfare brought about by pricing but also on the relative influence of groups in the political game.

In highly stylized models of political competition with two parties and a single policy dimension, the preferences of the median voter determine policy (Downs 1957). In practice, however, elected officials are not this tightly bound to citizen preferences for a number of reasons: First, voters are rationally ignorant in the sense that they acquire political information up to the point where the marginal cost of acquiring additional knowledge equals marginal benefits. These benefits are low because an individual has only a miniscule impact on policy-making. If voters are unaware of what elected officials do, the latter can deviate from citizen preferences. Second, in representative democracies, voters simultaneously decide a large number of issues when electing their representatives. In contrast to unemployment or general tax policy, environmental issues are not particularly salient during general election campaigns. As a consequence, the influence of voter preferences on policy-making is weaker in the area of environmental policy. Third, the lack of political information on the part of voters allows interest groups to influence policy-making. Even in a competitive political environment, elected officials are willing to distort policies in favor of organized interests because the campaign contributions from these interests allow candidates to increase their popularity with voters. And finally, as voters have little political information, it is often simplest for them to evaluate the relative performance of their elected officials. The resulting "yardstick competition" implies that there is little pressure on politicians to implement effective environmental instruments as long as other jurisdictions do not have successful programs of their own.

Once rational ignorance and the influence of groups are taken into account, the set of environmental policy instruments that is employed in political equilibrium can deviate significantly from the instruments citizens as voters (or, all the more, a social planner) would use. Nevertheless, there is little doubt that voter preferences constitute a significant constraint on political decision making and public opinion is influential in setting policy. Though the sensitivity of voters to environmental issues has increased over the last decades resulting in environmental issues being considered as fairly important by many voters, there is also ample evidence that voters are less than enthusiastic about bearing high costs for better environmental quality. Faced with the trade-off between higher real individual income and the production of better environmental quality that largely is a public good, it is reasonable that in many cases voters care more about their economic short-term well-being than the prospective environmental situation.

Voters also seem to prefer a policy of direct regulations and command and control measures to price incentives. There is evidence that pricing is not considered to

be a fair allocation mechanism neither as a mechanism to eliminate excess demand (Kahneman et al. 1986) nor in public good contexts. As regards the latter, Frey and Oberholzer-Gee (1996) document that willingness-to-pay is seen as the least fair of seven allocation mechanisms using a locally unwanted, but socially beneficial facility as their example. Moreover, there is considerable evidence that the introduction of economic incentives in one area can have negative consequences in others (Lepper and Greene 1978; Frey 1997). Such negative spillovers exist if pricing crowds out intrinsic motivation. This does not imply that price incentives fail to work but they become less effective, and there may be negative spillovers to other areas where no incentives for environmental protection exist. Altogether, these arguments contribute to explain why voters may be reluctant to accept effective environmental policies in general terms and market-based instruments particularly.

According to the public-choice approach, alternative policy measures are supplied by politicians in the political market pursuing their own goals subject to various constraints. Politicians are hypothesized to have self-interest in implementing specific instruments being either in line with their ideology or increasing their discretionary power or their personal income. However, in order to be re-elected they have to take into account voters' interests. The more binding the re-election constraint is, the less discretionary power the politicians have at their disposal in order to pursue their self-interest and the more they are linked to the demand side of the political process.

Given competition among alternative political parties and the re-election constraint being restrictive, politicians have to trade off benefits and costs (in terms of gains and losses in votes) when evaluating alternative policy measures. In political equilibrium, policies match the preferences of well-organized interests better than the preferences of more dispersed groups. In general, smaller groups are easier to organize than larger groups, and associations that find it less difficult to produce a mix of private and public goods ("selective incentives") are more likely to overcome the free-rider problem associated with interest group activities (Olson 1965). If groups are not already organized, it is unlikely that they will exercise decisive influence in any policy debate, whereas existing organizations can be counted on to exert considerable influence. In particular, producer interests (i.e., employers and employees) are better organized than consumers, and industry and business associations are more important players in the political game compared to environmental interest groups. By making campaign contributions and information available to politicians using them in order to attract additional voters, special interest groups can afford to be successful although their preferred policies are not in line with the preferences of the majority of voters.

The ability of groups to overcome free-rider problems is one of the determinants of the level of transfers to different groups. Another is the cost of transfers. The Chicago school of political economy emphasizes that political competition will ensure that the most efficient method of redistribution is chosen (Becker 1983). If ecological taxes or tradable permits are in fact the most efficient means to allocate environmental resources, the Chicago school suggests that interest groups will

prefer this form of transfers to other forms. Thus, given the will to reduce negative external effects with environmental policy, pricing schemes should be a politically attractive policy instrument.

However, the Chicago view of political economy, which emphasizes that law-makers and interest groups seek efficient ways to make transfers, stands in stark contrast to the Virginia school, which emphasizes that politicians will use ineffi-cient means of transfer if this allows them to hide the cost of redistribution. Tullock (1983), and Coate and Morris (1995) show that inefficient transfers will occur if voters have *ex post* difficulty distinguishing efficient from inefficient policies and if they are uncertain if the elected officials work in their best interest. In many political situations, these assumptions appear to be fairly realistic. Thus, politicians favor policies whose costs are difficult to see. Benefits, on the other hand, should be highly visible. Consequently, it can be stated that environmental policies are less promising than alternative policy issues (as, e.g., employment policies), and regu-lation policies are more attractive than pricing instruments. Charging drivers, for example, the prices for road usage directly keeps the costs of using roads highly visible, reminding voters of the policy every time they stop at toll booths or look at their electronically generated charges. While the costs remain highly visible, the benefits of the policy—reduced road congestion and better environmental quality—are much less salient (Oberholzer-Gee and Weck-Hannemann 2002).

Public-choice theory applied to environmental politics generally suggests that direct control measures have better chances to be realized than incentive-based instruments though the latter are more efficient. Both, a policy of command and control and incentive-based instruments, involve costs for reducing the emissions. In the case of taxes or tradable permits, however, the polluters have to pay for remaining emissions, which under a policy of command and control is avoidable, resulting in an additional rent (Buchanan and Tullock 1975). Moreover, polluting industries may consider that with direct control measures there is some leeway for negotiations with the environmental protection agency. Polluting industries can make use of their informational advantage in arguing for less strict regulations and exceptions from the rule. Thus, taken together, there seem to be good reasons why regulated industries prefer command and control measures to pricing instruments provided that they are not successful to avoid any environmental regulation at all.

Besides politicians, officials in the public bureaucracy have a considerable influence in the political market by preparing and formulating alternative policy proposals. They also have to implement and to examine the policy measures adop-ted. According to public-choice theory, public bureaucrats aim to increase their discretionary power and to weaken the budget constraint. In contrast to politicians, they are not faced with a re-election constraint. Their discretionary power arises out of the specific principal–agent relationship between the representatives in the political sector and public bureaucracy. They are expected to favor policy measures that have to be administered explicitly (providing them with discretionary power vis-à-vis government and the private sector) and as a result, they generally prefer direct control instruments and oppose the application of market-based instruments in environmental policy.

Nevertheless, environmental taxes and tradable permits might be attractive means to seek individual rents on the part of the relevant actors in the political debate. Generally, policy makers favor instruments that weaken the government's budget constraint. In this respect, environmental taxes recommend themselves because they generate additional funding. Thus, besides regulatory measures, also pricing instruments may well serve the self-interest of policy-makers provided that the additional resources are at the disposal of policy-makers themselves.

In recent years, economists and lawmakers have considered the option of linking the phasing in of environmental taxes to reductions in taxes on labor, a reform project that is often referred to as an *ecological tax reform*. If the revenues from environmental taxes were used to lower other taxes, it is theoretically possible to reduce the overall cost of transfers in an economy, thereby making such a pricing scheme politically more attractive. While there is little disagreement about the existence of a "green" dividend—ecological taxes are generally expected to increase environmental quality—it is less clear if a "blue" dividend exists, where "blue" refers to a reduction in the overall distortions in the tax system and a subsequent increase in employment (for a survey of the double dividend debate see Goulder 1995).

Bovenberg and DeMooij (1994) show that environmental taxation can in fact reduce employment and economic welfare. Their argument, based on optimal taxation theory, is that taxing a broad base (e.g. labor) will induce lesser distortions than taxing a narrow base (as, e.g., energy or CO_2). If the environmental tax is ultimately borne by labor, this narrow-based tax will finally lead to larger distortions than the broad-based labor tax.

Altogether, theoretical and empirical work does not support the idea that an ecological tax reform will bring about notable efficiency gains that help establish environmental taxes. Keeping in mind a political economy perspective, however, an ecological tax reform may still bring about additional benefits for two reasons. First, by definition, a narrower tax base allows citizens to more easily substitute away from the taxed activities, making tax increases less attractive from the perspective of a revenue-maximizing politician and keeping the size of government more limited (Brennan and Buchanan 1980). Secondly, unlike taxes on labor, proportional (indirect) taxes have the advantage of not automatically increasing with labor productivity (Kirchgässner 1998).

While these arguments may be appealing for voters, politicians are not attracted by ecological taxes for these reasons. Their concern is neither to tame *Leviathan* nor primarily to improve the natural environment. Rather, they may be concerned about the situation on the labor market and the reduction of the unemployment rate in order to weaken their re-election constraint, or they are interested in taxes creating additional revenue at their discretionary disposal. Thus, in contrast to the previously dominant view in public-choice theory, governments may argue in favor of environmental taxes and by this way aim at improving the environment but "for the wrong reasons" (Kirchgässner and Schneider 2003, p. 383).

In addition, if pricing revenues are returned to citizens, politicians can try to channel these funds toward their own constituencies. Pricing revenues could also be

used to compensate those who lose when economic incentives are introduced. Well-organized groups can be expected to support pricing measures provided that the revenues are used to finance infrastructure and services being in their own interest. On the other hand, they are assumed to be less in favor of pricing measures given that the purpose is explicitly and exclusively to internalize external costs combined with lump sum transfers or a reduction of other taxes. In effect, this is an argument to target revenues from environmental taxes to projects that benefit polluters. There is some empirical evidence that taxes can be introduced if they are channeled back to those opposing the price measure. Kimenyi et al. (1990), for example, show for the United States in general that, in comparison to general fund financing, earmarking leads to increased tax revenues. Hence, given the re-election constraint to be decisive, pricing instruments may even so have a chance if they are introduced in such a way that well-organized groups are benefited most and the costs are spread to less influential and latent interest groups. Earmarking of revenues in this case may be an essential feature to achieve the respective aim on the part of politicians and most powerful interest groups.

Beyond that, the opposition to environmental taxes by main polluters may be mitigated by accepting exceptions and tax allowances (Hahn 1989). If emission taxes are fixed at a relatively low rate and therefore avoidance costs in the case of emission standards exceed the tax burden, this solution is in effect favorable for polluters. If, likewise, exemptions are made for the most polluting sectors, e.g., the energy-intensive producing industries in the case of CO_2 taxes (Ekins and Speck 1999), this implies that the resistance of those producers who produce most emissions can be weakened. However, this also reduces the environmental impact of such a policy significantly.

Likewise, tradable permits may be implemented in such a way that those groups mostly affected get an additional rent (Hahn 1989; Kirchgässner and Schneider 2003). If the permits are auctioned, there is additional revenue for the government, which can be used either in their own interest or to the benefit of taxpayers or to the advantage of effectively lobbying interest groups. If, on the other hand, grandfathering is used, the existing firms get the pollution rights for free and are put in a position to sell them. Moreover, grandfathering creates a barrier to entry against new firms because these have to pay for all the permits they need or the permit market may be so much restricted that no significant trade occurs and newcomers are kept away by this way. It follows that existing firms may well favor the grandfathering of tradable permits. And, indeed, according to Svendsen (1999), the position of private business interest groups seems to have changed in the United States from less advocating a command and control policy in favor of a grandfathered permit market. Likewise, for reasons of political feasibility, the implementation of tradable allowance systems (as, e.g., the SO_2 allowance trading scheme under the Clean Air Act in the United States, or the EU trading scheme for carbon emissions from energy-intensive installations) has been linked to a free initial allocation of emission allowances (Böhringer and Lange 2005).

Thus, all in all, the dominant interest groups are expected to orient their lobbying activities toward preventing any effective policy measures. As far as alternative

environmental instruments are concerned, they most likely accept direct control measures but, nevertheless, incentive-based instruments may also have a chance to be implemented if the following conditions hold: the less pronounced the incentive effect of the pricing measures turns out (i.e., moderate changes in prices with only a limited incentive effect); the more likely it is for special interest groups to realize exceptions from the rule (e.g., when those groups particularly affected by these measures are exempted or at least admitted a reduced rate or a transitional arrangement); the more likely it is to shift the burden on to latent interest groups or groups without voting rights (as, e.g., foreigners); and if the link with a rebating scheme (like grandfathering of tradable permits) or earmarking of emission tax revenues ensures that there are not only costs but also benefits (e.g., when revenues from pricing instruments are earmarked to the use of maintaining and improving the infrastructure of services which benefits the producers, operators, and users of the corresponding services).

Recently, another instrument that is also often labeled as a market-based instrument is increasingly used. Yet, voluntary agreements are in no way such an instrument but instead the main purpose of their support is to prevent the use of effective instruments of environmental policy. As Kirchgässner and Schneider (2003) emphasize, the only possibility to make voluntary agreements effective is to combine them from the beginning with the threat that the government will intervene if the negotiated results will not be reached. But, in this case the voluntary agreement is actually superfluous and just a kind of symbolic policy.

Also, in international environmental policy the willingness to introduce market-based instruments, such as internationally tradable permits or "joint implementation" or "clean development mechanism" projects, might be of a more symbolic nature: in demanding to introduce such instruments, of which it is obvious that the distributional problems bring about that their implementation has no real political chance, may be an effective way to prevent the implementation of more effective and enforceable policy measures (Kirchgässner and Schneider 2003). In the case of the Kyoto Protocol, Böhringer and Vogt (2004) argue that the concessions made essentially reduce it to a symbolic treaty that codifies business-as-usual emissions and makes compliance a rather cheap deal.

Altogether, considering these new developments, the moderate increase in the use of economic instruments of environmental policy does not invalidate the arguments put forward by the public-choice approach. There is still only limited support of the use of incentive-based instruments, and their application in many respects deviates from the ideal therapy. The synopsis given by Kirchgässner and Schneider (2003, p. 384) seems to be well targeted when they state that economic instruments, at best, "will be introduced for other (non-environmental) reasons and/or in a way which is not very helpful for the environment. But, on the other hand, it is a step in this direction and one might hope that over time citizens become more familiar with such instruments and their advantages which might—in the long run—increase their acceptance in the electorate."

One might also think about adequate institutional conditions contributing to improve the chance that incentive-based instruments as the most efficient means

in environmental policy have a better chance to be implemented in the political decision-making process. Referring to a process-oriented approach, it can be argued that the political process itself has to ensure that all relevant arguments have an equal chance to enter into the discussion resulting in efficiency to be reached endogenously, i.e., via the process and not via the evaluation of alternative outcomes (Weck-Hannemann 2006). Ideally, all the pros and cons have to enter in the political process without distortion. This is best guaranteed if voters have a direct say in political matters and can act as agenda setters, as well as if the principle of fiscal equivalence and institutional congruence is realized. With the institutions of direct democracy and the right of initiative and institutional congruence, it can be expected that politicians are forced to be more responsive to voters' interests than in a system of representative democracy with spillovers of external effects.

At the constitutional level, the decision makers do not know their specific individual position but the social consequences of alternative policy programs. This "veil of uncertainty" enables that fair and efficient rules are adopted. However, in order to elicit such fair and efficient rules, the "veil of uncertainty" has to be sufficiently strong. This might be approximated in the following ways (Kirchgässner 1994): if rules are discussed with respect to uncertain future events, if individuals decide for their descendants, and if the time span is long enough between the decision about the rules and the coming into force of these rules. Consequently, the acceptance and implementation of pricing instruments in environmental policy might be furthered by assigning them as long-term general measures instead of discussing the issue in a predominantly short-term and concrete context.

References

Baumol, W.J. and Oates, W.E. (1971). "The use of standards and prices for the protection of the environment." Swedish Journal of Economics, 73, 42–54.

Becker, G.S. (1983). "A theory of competition among pressure groups for political influence." The Quarterly Journal of Economics, 98(3), 371–400.

Böhringer, C. and Lange, A. (2005). "Economic implications of alternative allocation schemes for emission allowances." Scandinavian Journal of Economics, 107(3), 563–581.

Böhringer, C. and Vogt, C. (2004). "The dismantling of a breakthrough: The Kyoto protocol as symbolic policy." European Journal of Political Economy, 20, 597–617.

Bommer, R. (1998). Economic Integration and the Environment. A Political-Economic Perspective. Cheltenham: Edward Elgar Publishing.

Bovenberg, A.L. and DeMooij, R.A. (1994). "Environmental levies and distortionary taxation." American Economic Review, 84, 1085–1089.

Brennan, G. and Buchanan, J.M. (1980). The Power to Tax: Analytical Foundations of a Fiscal Constitution. Cambridge: Cambridge University Press.

Buchanan, J.M. and Tullock, G. (1975). "Polluters' profits and political response: direct controls versus taxes." American Economic Review, 65, 139–147.

Coase, R.H. (1960). "The problem of social cost." Journal of Law and Economics, 3, 1–44.

Coate, S. and Morris, S. (1995). "On the form of transfers in special interests." Journal of Political Economy, 103(6), 1210–1235.

Congleton, R.D. (ed.) (1996). The Political Economy of Environmental Protection: Analysis and Evidence. Ann Arbor: University of Michigan Press.

Congleton, R.D. (2006). "The Story of Katrina: New Orleans and the political economy of catastrophe." Public Choice, 127, 5–30.

Dewees, D.N. (1983). "Instrument choice in environmental policy." Economic Inquiry, 21, 53–71.

Dijkstra, B.R. (1999). The Political Economy of Environmental Policy. Cheltenham: Edward Elgar Publishing.

Downing, P.B. (1991). "A political economy model of implementing pollution laws." Journal of Environmental Economics and Management, 8, 255–271.

Downs, A. (1957). An Economic Theory of Democracy. New York: Harper & Row.

Ekins, P. and Speck, S. (1999). "Competitiveness and exemptions from environmental taxes in Europe." Environmental and Resource Economics, 13, 369–396.

Frey, B.S. (1997). Not Just for the Money. An Economic Theory of Personal Motivation. Cheltenham: Edward Elgar Publishing.

Frey, B.S. (2003). "Why are efficient transport policy instruments so seldom used?," in J. Schade and B. Schlag (eds.) Acceptability of Transport Pricing Strategies. Oxford: Elsevier, 63–75.

Frey, B.S. and Oberholzer-Gee, F. (1996). "Fair siting procedures: An empirical analysis of their importance and characteristics." Journal of Policy Analysis and Management, 15, 353–376.

Frey, B.S., Schneider, F., and Pommerehne, W.W. (1985). "Economists' opinions on environmental policy instruments: Analysis of a survey." Journal of Environmental Economics and Management, 12(1), 62–71.

Goulder, L.H. (1995). "Environmental taxation and the double dividend — A reader's guide." International Tax and Public Finance, 2, 157–183.

Hahn, R.W. (1989). "Economic prescriptions for environmental problems: how the patient followed the doctor's orders." Journal of Economic Perspectives, 3(2), 95–114.

Hahn, R.W. (1990). "The political economy of environmental regulation: Towards a unifying framework." Public Choice, 65, 21–47.

Kahneman, D., Knetsch, J., and Thaler, R. (1986). "Fairness as a constraint on profit seeking: Entitlements in the market." American Economic Review, 76, 728–741.

Kimenyi, M.S., Lee, D.R., and Tollison, R.D. (1990). "Efficient lobbying and earmarked taxes." Public Finance Quarterly, 18(1), 104–113.

Kirchgässner, G. (1994). "Constitutional economics and its relevance for the evolution of rules." Kyklos, 47(3), 321–339.

Kirchgässner, G. (1998). "The double dividend debate: Some comments from a politico-economic perspective." Empirica, 25(1), 37–49.

Kirchgässner, G. (1999). "International environmental policy: Problems and possible solutions." Aussenwirtschaft, 54(2), 269–290.

Kirchgässner, G. and Schneider, F. (2003). "On the political economy of environmental policy." Public Choice, 116(3–4), 369–396.

Lepper, M. and Greene, D. (eds.) (1978). The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation. New York: Erlbaum.

Oberholzer-Gee, F. and Weck-Hannemann, H. (2002). "Pricing road use: Politico-economic and fairness considerations." Transportation Research Part D: Transport and Environment, 7, 357–371.

Olson, M. (1965). The Logic of Collective Action. Cambridge: Harvard University Press.

Pearson, M. (1995). "The political economy of implementing environmental taxes," in L. Bovenberg and S. Cnossen (eds.) Public Economics and the Environment in an Imperfect World. Boston; Kluwer, pp. 37–57.

Pigou, A.C. (1920). The Economics of Welfare. London: MacMillan.

Schneider, F. and Weck-Hannemann, H. (2005). "Why is Economic Theory Ignored in Environmental Policy Practice?" in C. Böhringer and A. Lange (eds.) Applied Research in Environmental Economics. Heidelberg: Physika, pp. 257–275.

Schulze, G.G. and Ursprung, H.W. (eds.) (2001). International Environmental Economics — A Survey of the Issues. Oxford: Oxford University Press.

Schwarze, R. and Wagner, G.G. (2007). "The political economy of natural disaster insurance: Lessons from the failure of a proposed compulsory insurance scheme in Germany." European Environment, 17(6), 403–415.

Svendsen, G.T. (1999). "U.S. interest groups prefer emission trading: A new perspective." Public Choice, 101, 109–128.

Tullock, G. (1983). Economics of Income Redistribution. Boston: Kluwer.

Weck-Hannemann, H. (2006). "Efficiency of Protection Measures," in W.J. Ammann et al. (eds.) Risk 21 - Coping with Risks Due to Natural Hazards in the 21st Century. London: Taylor & Francis, pp. 147–154.

Chapter 32
Experimental Public Choice[1]

Arthur J.H.C. Schram

Introduction

A few decades ago, most economists believed that their discipline was non-experimental. Economic phenomena should be studied theoretically or empirically. The ideal paper was one where rigorous theory was tested using advanced econometric methods. The fact that the empirics were usually based on (often incomplete) field data only remotely related to the problem at hand was no problem: this is why we had econometrics.

However, over the past decades, economics has rapidly become an experimental science. It has become obvious to many that laboratory experiments provide the means to control conditions in a way that allows for a systematic test of economic theories. Contrary to econometric testing, laboratory testing allows one to systematically test essential elements of a theory. In addition, experiments can be used to explore potential paths of new research, in situations where no theory exists or where existing theory is shown to be inadequate. Finally, experiments have the advantage that they can be replicated, allowing for a systematic analysis of the robustness of the findings.

The use of laboratory experiments in public choice research has also increased rapidly in the last 30 years or so. At meetings of the various public choice societies, it has become very common to encounter experimental papers. This is no coincidence but has been actively solicited by the societies themselves. For example, for more than a decade the North American Public Choice Society organized its yearly spring conference together with the Economic Science Association (the international society of experimental economists). The bylaws of the European Public Choice Society even explicitly state that, "*The Society's interest is in theoretical rigor, empirical and experimental testing, and real world applications.*"

---

[1] This chapter is a revised and updated version of an essay that first appeared in *The Encyclopedia of Public Choice* edited by Charles K. Rowley and Friedrich Schneider and published in 2004 by Kluwer Academic Publishers, Volume I, 96–104.

The increased use of experiments in public choice is, definitely, an enrichment to this literature. Much of the literature on non-market decision-making is based on theoretical assumptions about individual behavior (Schram 2000) or on field data from elections or surveys that are not particularly tailored to answer the questions raised by the theory. In both cases, experiments provide a method that is complementary to the existing methods. Together with theory and empirics based on field data, experiments allow us to understand public choice phenomena in more depth.

Two types of experimental studies can be important for public choice. One group is concerned with individual behavior and motivations. Its conclusions with respect to individuals' motivations and preferences (Schram 2000), or the role of emotions and bounded rationality (Bosman 2001; Bosman and van Winden 2001), for example, can have important consequences for the assumptions made in many public choice theories. However, this type of study is not discussed in detail in this essay. A brief evaluation of their importance is given in the concluding section. A detailed discussion of their relevance to public choice can be found in Schram (2000). Instead, this essay focuses on a second group of studies: those where experiments are used to analyze a number of traditional public choice topics.[2]

This essay is organized as follows. The next section briefly describes the experimental methodology. This is followed by four sections on experiments in public choice: public goods, voter turnout and participation games, rent seeking and lobbying, and spatial voting. A concluding discussion is presented at the end.

Experimental Economics

In a laboratory experiment, behavior is studied in a controlled environment. Participants (in most cases university students) are invited to a computer laboratory, where they are asked to make decisions in a framework designed by the experimenter. Decisions are 'real,' e.g., in the sense that they have monetary consequences for the subjects. At the end of the experiment, they are paid in cash, an amount that depends on their own decisions and (in many cases) on the decisions of other participants. Excellent descriptions of what an experiment in economics entails and how one can set up an experiment are provided in Davis and Holt (1993) and Holt (2007).

Traditionally (Smith 1994), one distinguishes the environment, institutions and behavior in an experiment. The environment refers to the structural characteristics of an economic problem, such as the number of agents, the information structure, preferences, endowments, cost structure, etc. According to Davis and Holt (1993), economists traditionally viewed economic problems almost exclusively in terms of these characteristics. Institutions refer to the rules governing the interaction of the economic agents, such as the market or auction rules, or the government decision-making procedures. For a long time, it was argued that it is possible to control the

---

[2] A third group of studies observed is that of 'political engineering:' experiments are used to help design political systems. See Riedl and van Winden (2001) for an example.

environment and institutions in an experiment and to study behavior. By varying institutions, for example, one could investigate how they affect behavior.

However, two caveats can be made with respect to this traditional distinction. First of all, one can argue that many non-experimental economists have considered the importance of institutions as well. The boom in institutional economics and game theory has highlighted the important effects that they may have on behavior. Second, it is not obvious that one can control the environment completely. Especially, preferences might be difficult to control in a laboratory. Though one tries to induce preferences by offering a payoff scheme, one cannot control individual preferences for other things than their own private earnings. Nevertheless, it is obvious that the laboratory allows for a much higher level of control than was possible before.

Experimental results can therefore carry much weight. The control in a laboratory allows one to address very specific research questions. For example, if we are interested in studying the phenomenon of committee voting on two proposals, under two different voting rules (see section on **Spatial Voting Experiments**), there is no better setting to study this than in an experiment where the only treatment variable is the voting rule. Keeping all other aspects of the problem constant (e.g., number of members, payoff to each member if either proposal is accepted, etc.) the environment is stripped of all the confounding elements that we typically observe in the outside world. What remains is exactly what we want to study; the effect of the voting rule. If we combine an analysis along these lines with a theoretical analysis and an empirical analysis using field data, this is likely to lead to a much more complete understanding of the problem at hand than we would be able to achieve without the laboratory data.

Of course there are also disadvantages related to using the experimental method. Many of theses are discussed in the standard texts in this field (Davis and Holt 1993; Kagel and Roth 1995; Holt 2007). Plott (1982) and Schram (2005) systematically discuss questions raised by economists about the validity of laboratory experiments. Here, we briefly discuss the issue of external validity, i.e., is the evidence obtained in a laboratory relevant for the 'outside world?'

Naturally, the external validity of an experiment depends on the experimental design. There is no reason why the external validity of all laboratory experiments per se should be doubted. Subjects participating in an experiment are real people. They are facing real monetary incentives that (in a carefully designed experiment) are salient. Hence, if we observe certain behavior in an experiment it is economic behavior.[3] Nevertheless, every experimental design should be critically assessed with respect to the structure and its relationship with the problem being studied. In general, a thorough theoretical analysis of the problem at hand is useful in this assessment.

---

[3] It is, of course, possible that the behavior observed is specific to the particular kind of subject in the experiment (usually students). However, This is a common problem related to experimental procedures as opposed to the experimental method as such.

Public Good Experiments

Public good experiments usually study voluntary, individual, contributions to a public good. Given the role that government plays in providing public good and the possibility that government provision crowds out individual contributions, this is of obvious importance in public choice. It is therefore no surprise that one of the first major papers on this topic was published in the journal Public Choice (Isaac et al. 1984). Since then, studies on voluntary contributions to public good have been a major part of the experimental literature.

The typical setup of a public good experiment is as follows. Subjects are allocated into groups of size N (typically, N is equal to 4 or 5). Each is given an endowment of 'tokens.' These must each be invested in either a 'private account' or a 'public account.' Each token in the private account gives a payoff A to the subject alone. Each token in the public account gives a payoff B to every participant in the group. Hence, an investment in the public account is a voluntary contribution to a pure public good. The interesting cases are where $B < A < NB$, because this implies that contributing nothing to the public good is a dominant strategy, whereas contributing all tokens is efficient. The relative payoff to the two accounts, A/B, gives the marginal rate of substitution (hereafter, mrs) between private and public account.[4]

Ledyard (1995) presents one of the first extensive surveys of experiments of this type. Some important regularities listed in his survey are:

 i) contributions to the public good start at a relatively high level (typically 40–60% of the endowments);
 ii) with repetition, these contributions decrease to 0–30% however, very often they do not decrease to zero(Isaac et al. 1994);
 iii) the contribution levels are a decreasing function of the mrs.

To these regularities, one can add:

 iv) contributions increase with group size (Isaac et al. 1994);
 v) many subjects split their tokens across the two accounts, i.e., they don't contribute everything to one of the two accounts (Palfrey and Prisbrey 1997; Brandts and Schram 2001);
 vi) there is considerable subject heterogeneity: systematic differences across subjects exist; some consistently contribute, some never do; others switch from contributing to not contributing (Palfrey and Prisbrey 1997; Brandts and Schram 2001);
 vii) if group composition is held constant across periods, contributions get more concentrated in groups as the experiment moves on (Brandts and Schram 2001).

The type of public good experiments described is by far the type most often studied. The linear production function for public good is easy for subjects to understand and

---

[4] It is also quite common to refer to B/A (i.e., 1/mrs) as the marginal per capita return (mpcr); see Isaac et al. (1984).

provides an interesting tradeoff between private earnings and group welfare.[5] Over the last few years many extensions to the setup have been studied, often in an attempt to study other preferences or reciprocity (Schram 2000; Fehr and Gächter 2000a; Fehr and Fischbacher 2003). These extensions include the study of situations where there is no efficiency gain to be made from cooperation (Palfrey and Prisbrey 1997; Brandts and Schram 2001); cross-cultural comparisons (Cason et al. 2002; Brandts et al. 2004); framing (Andreoni 1995a; Sonnemans et al. 1998); and the effect of allowing costly punishment for free riders (Fehr and Gächter 2000b).

The bottom line in this whole body of literature is that subjects in public good experiments contribute voluntarily to the public good, to a much larger extent than the selfish individually rational prediction of free riding would have them do. In addition, contributions cannot simply be attributed to erratic behavior of the subjects (Andreoni 1995b; Brandts et al. 2004). On the other hand, contributions tend to be lower than the efficient level, many subjects do free ride and contributions decrease with repetition. There is an ongoing discussion about what motivates subjects to behave in this way.

From a public choice point of view, the relationship between group size and free riding is of interest. Mancur Olsen's idea that free riding will increase with group size is not supported by experimental data. On the contrary, contributions increase with group size (see regularity iv, above).[6] Isaac et al. (1994) show that it is not group size per se that matters, but the interaction between group size and mrs. Keeping the mrs constant, the 'pure' group size effect is positive (if the mrs is large enough). At this stage, it is difficult to understand this apparent anomaly. One possibility is that it is related to the gains from cooperation (Brandts and Schram 2001). For any given mrs, a specific number k of contributors is needed to make them better off (as a subgroup) than if none of them would contribute. The larger the group, the more likely it is that there will be k contributors. From an individual's point of view, in a large group it is less likely that he will be a 'sucker', for whom the cooperative gain is smaller than the individual contribution to the public good.

Voter Turnout as a Participation Game

The paradox of voter turnout has been the subject of academic debate for decades (for an early survey, see Schram 1991). The debate probably started with Downs' (1957) formulation of the problem. He notes that, due to the low probability of being decisive, the expected benefits from voting in a large-scale election are generally

---

[5] Another widely studied case is where the production technology uses a threshold: a minimum amount of contributions needs to be collected for the public good to be produced. These 'step-level public good' are extensively studied in Offerman et al. (1996). A meta-analysis is given by Croson and Marks (2000).

[6] On the other hand, Offerman et al. (1996) show that voluntary contributions decrease with group size in a step-level public good game.

outweighed by the cost of the act. Nevertheless a very large number of voters actually turn out to vote in general elections. Many theoretical and empirical papers have been published trying to explain the paradox, but only few rational choice models have been developed that show that voter turnout might sometimes be rational in an instrumental sense (Ledyard 1984; Schram 1991, and the references given there).

Palfrey and Rosenthal (1983) model the turnout problem as a participation game and study it with respect to game-theory. In this game, there are two or more teams. Everyone has to make a private decision that is beneficial to every member in one's own team and harmful to the members of other teams. The decision is whether or not to 'participate' in an action, where participation is costly. Palfrey and Rosenthal show that in many cases, Nash-equilibria with positive levels of participation exist. Note that there are two types of conflicts in a participation game. Within the group, there is an incentive to free-ride on the costly participation of other group members. Between groups, there is an incentive to compete and out-vote the other group. Note the difference with the incentives in the public good games of the previous section, where only the free-riding incentive exists.

It is difficult to study voter turnout using field data.[7] However, Participation games provide a structure to study this decision experimentally. This was first done by Bornstein (1992) and Schram and Sonnemans (1996a, b). Here, we shall describe the experiment used in the latter two papers. In the experiments, subjects are split in two groups of 6 individuals. Each subject had to decide whether or not to buy a token at a given price. The number of tokens bought in each group determines the payoffs. There were two payoff schedules, representing a winner-takes-all election (*WIN*) and proportional representation (*PR*). In *WIN*, each member of the group that bought the most tokens (won the elections) received a fixed sum and the payoff for the other group was zero (with ties broken randomly). In *PR* the payoff to any group-member was proportional to the relative turnout of the groups. In addition to these experiments, two *WIN* sessions were run with groups of 14 subjects and two where one group consisted of 8 and the other of 6 participants. The results obtained show that:

   i) Nash equilibrium is a poor predictor of turnout;
  ii) participation is higher in winner-takes-all than in proportional representation; this is in line with the comparative static of the pure strategy equilibria;
 iii) participation is higher with repetition in fixed groups;
  iv) participation increases substantially after five minutes of free communication;
   v) there is no significant effect of group size on relative turnout;
  vi) when the groups' size is unequal, relative turnout is higher in the smaller group.

---

[7] Güth and Weck-Hannemann (1997) and Blais and Young (1999) use field experiments to study the paradox of voter turnout. In both cases, the turnout decision of a group of students in a real election is monitored. Güth and Weck-Hannemann study the value of a vote by offering a payment in return for abstention. Blais and Young study the effect of being exposed to a presentation about the turnout paradox.

Cason and Mui (2004) study the role of participation costs and uncertainty with respect to the benefits of voting in the participation game framework. They find that increasing costs cause decreasing participation rates. Uncertainty has mixed effects, depending on which of the subjects are facing it. Cost uncertainty is also an element in Levine and Palfrey (2007). Their study supports result (vi) and finds support to two additional comparative static predictions derived from the Palfrey and Rosenthal model:

vii) turnout decreases with the size of the electorate;
viii) turnout increases with expected closeness.

With co-authors, Großer studies three extensions to the participation games: (1) the effect of polls on turnout (Großer and Schram 2007); (2) the endogenization of policy (and group) formation (Großer et al. 2002); and (3) the introduction of information about the turnout decision of some other voter (Großer and Schram 2006). The results in these papers include:

ix) turnout increases after the release of polls if elections are predicted to be close;
x) a mix of allied and floating voters in a group yields higher turnout rates than a situation without allied voters;
xi) endogenous political ties between voters and candidates are observed, i.e., over a series of elections, candidates design policies for specific groups of voters who reward them with their votes;
xii) higher turnout rates are observed when subjects are informed about the decision of other participants.

The bottom line in the research on participation games is that (as in elections) traditional theory is a bad predictor of turnout. However, participants often do react to a change of incentives (e.g., an increase in the costs of voting) in a predictable way. Moreover, the results of the recent experiments on group size uncertainty and information about other voters show how experimentation might be useful in a further analysis of the turnout paradox. Another interesting development is that some of the results observed in participation games can be explained in a quantal response framework McKelvey and Palfrey (1995). Goeree and Holt (2000) show that the results reported in Schram and Sonnemans (1996a, b) are consistent with a quantal response equilibrium. Cason and Mui (2004), Großer and Schram (2007) and Levine and Pafrey (2007) also observe that a quantal response equilibrium organizes the data relatively well.

Rent-Seeking and Lobbying

van Winden (1999, 2003) provides detailed surveys of the experimental literature on lobbying. He distinguishes between three types of studies that are relevant in this respect: experiments on common agency problems, signaling models and rent seeking.

We will briefly discuss the first two categories (see van Winden's surveys for more details) and elaborate a bit on the rent seeking experiments.

Common agency experiments (Kirchsteiger and Prat 2001) study the effect of campaign contributions in exchange for favorable policies. Lobbyists compete in offering a politician a 'contribution schedule' which depicts the (possible) campaign contribution as a function of potential policies. The politician chooses a policy and collects the contribution. One interesting experimental result is that lobbyists tend to focus on the most preferred result and do not offer significant contributions for less preferred alternatives.

In signaling games, the focus is on the strategic transmission of (relevant) information from a lobbyist to a politician. This (credibly) assumes that the lobbyist may have information that is relevant to the politician, when making his decision. Of course, there may be an incentive for the lobbyist to transmit untruthful information. The experiments of Potters and van Winden (1992) study this environment. Though their results cast doubt on the predictive power of the theoretical literature in this field, they do find that mutual beneficial (costly) transmission of information takes place from the lobbyist to the politician.

In the experimental rent seeking literature the role of the policy maker is replaced by some commonly known mechanism. The focus is on the competition between lobbyists. They typically compete for a 'prize' (rent) by placing bids. The mechanisms used are typically that the highest bid wins the prize or that the probability of winning the prize is proportional to the (relative) bid. An important characteristic of this setup is that all bids are irreversible (like in an all-pay auction). This yields possible inefficiencies (overdissipation) in the lobbying process, because the sum of the bids may be higher than the value of the prize.

The main experimental studies on rent-seeking are Millner and Pratt (1989), Davis and Reilly (1998), Potters et al. (1998), and Weimann et al. (2000). Here, we will describe the experiments in Potters et al. The mechanism used in this study to determine a winner is part of the experimental design, which distinguishes between perfectly and imperfectly discriminating contests. This distinction allows the authors to compare a situation where the equilibrium strategies yield positive probabilities of overdissipation (i.e., inefficiency), with one where the probability of overdissipation is zero. In other words, the equilibrium in one experimental treatment attributes a positive probability to the event that the sum of the bids is higher than the prize, whereas another attributes zero probability to this event. The main results are:

 i) the (Nash) equilibrium predictions are not supported;
 ii) overdissipation is more likely when theory predicts that it will be. In other words, the point predictions derived from game theory are rejected but the predicted comparative static are supported;
iii) subjects 'learn' to play more according to theory as they gain more experience through repetition over rounds.

The bottom line in this line of research is that only limited support for the theoretical literature on lobbying is found. This may be due to the fact that some

subjects do play in line with theory, whereas others do not (Potters et al. 1998), yielding an aggregate outcome that does not provide support.

Spatial Voting Experiments

Two types of voting experiments can be observed in the literature, both starting in the late seventies and early eighties at the California Institute of Technology. Both are basically experiments on spatial voting models, one focusing on committee voting, and the other on the median voter model.[8] An early survey of these experiments is presented in McKelvey and Ordeshook (1990). Because many of these studies were undertaken more than a decade ago, this survey still covers many of the important experiments in this field. A more recent survey is included in van Winden and Bosman (1996).

In the committee voting experiments, the typical setup is one, where a committee of n members has to choose a point (x,y) on some two dimensional issue space. Each committee member is assigned a personal ideal point in this space and the individual payoffs are a (declining) function of the distance between this ideal point and the point chosen by the committee. A commonly known decision-making institution determines how a committee decision is determined. These institutions describe, e.g., that the agenda setting, the communication and the majority rule used. Examples of this type of experiment include Berl et al. (1976), Fiorina and Plott (1978), Hoffman and Plott (1983), Holt and Anderson (1999), and Eavey and Miller (1984), who argue that this model can be considered to be a test of the Niskanen (1971) model of bureaucracy. The conclusions include:

i) if decisions are made by the simple majority rule and a Condorcet winner exists, the committee decision is close to that outcome;
ii) if decisions are made by the simple majority rule and no Condorcet winner exists, stable outcomes are often observed, though as yet no theoretical predictions for these outcomes are known;
iii) communication does not have a large effect on the outcome;
iv) the Niskanen model does not find support, in the sense that an agenda setter (bureaucrat) does not manage to make his ideal point the committee decision.

The median voter experiments study the interaction between political candidates and voters. The latter are given an ideal point in a one- or two dimensional policy space. Once again, their payoffs are a declining function of the distance between the chosen point and the ideal point. Candidates choose a position in the policy space, hoping to attract voters. If elected, their position is chosen and determines the

---

[8] A limited number of other topics related to voting have been studied experimentally. These include vote trading (McKelvey and Ordeshook 1980), voting on jury decisions (McKelvey and Palfrey 1998), voting to prevent 'public bad (Sutter 1999), and the aggregation of information through elections (Wit 1997).

voters' payoffs. Candidates' payoffs are positively related to the number of votes they receive (e.g., the winner receives a fixed amount and the loser receives nothing). Examples of this type of experiments include Collier et al. (1987), Williams (1991), McKelvey and Ordeshook (1993), and Olson and Morton (1994). The conclusions include:

  i) with complete information on ideal points and payoffs, the median voter model finds support;
 ii) even with incomplete information, there is convergence to the median, when it exists;
iii) costly information on candidates' positions does not affect the rate of convergence to the median.

The bottom line of both types of voting experiments is that quite some support for theoretical predictions (Condorcet winner, median voter) is observed. Moreover, stable outcomes when no Condorcet winner exists and convergence to the median voter even in case of incomplete information indicate that the voting mechanism can lead to even more robust results than predicted by theory.

Concluding Discussion

As mentioned in the introduction, there is a large body of experimental literature that is not directly related to public choice topics, but still very relevant for the analysis of these topics. This is the literature on individual motivations and (bounded) rationality. Most of public choice theory is based on the *homo economicus*, who pursues his selfish preferences in a perfectly rational way.

Both of these elements have been questioned, based on experimental results (see Schram 2000, or Charness and Rabin 2002). Many authors argue that preferences are only selfish in certain circumstances. Instead, it is argued that other aspects regarding such preferences as altruism or fairness can be widely observed for many subjects. This observation has led to theoretical models incorporating these preferences (e.g., Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Charness and Rabin 2002). By now, there is enough material to try to incorporate other preferences in some of the traditional public choice theory.

The assumption of perfect rationality also needs to be adjusted. It is becoming increasingly clear that emotions (Bosman 2001) and limits to rationality (Camerer 1998) can have major impacts on behavior. However, in this case it is less clear how existing models could be adapted to accommodate these results. As yet, there is no model of bounded rationality that seems to be applicable. On the other hand, these results do create space for theories based on 'reasonable' behavioral assumptions other than perfect rationality.

The examples given in this essay show that a variety of typical public choice topics have been studied in laboratory environments. In some cases, this has given support to existing theories and ideas (e.g., the median voter model). In others (e.g., rents seeking), it raises doubts about the validity of the theory in its present form.

In yet other cases (e.g., the turnout paradox), the experimental results can give hints as to ways to develop the theory further. In this way, experimental studies have proven to be a welcome addition to the public choice literature.

References

Andreoni, J. (1995a). "Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments." *Quarterly Journal of Economics*, 110, 1–22.

Andreoni, J. (1995b). "Cooperation in public-goods experiments: kindness or confusion"? *American Economic Review*, 892, 891–904.

Berl, J., McKelvey, R., Ordeshook, P., and Winer, M. (1976). "An experimental test of the core in a simple n-person cooperative non-side payment game." *Journal of Conflict Resolution*, 20, 453–479.

Blais, A. and Young, R. (1999). "Why do people vote? An experiment in rationality." *Public Choice*, 99, 39–55.

Bolton, G. and Ockenfels, A. (2000). "ERC: a theory of equity, reciprocity and competition." *American Economic Review*, 90, 166–193.

Bornstein, G. (1992). "The free-rider problem in intergroup conflicts over step-level and continuous public goods." *Journal of Personality and Social Psychology*, 62, 597–606.

Bosman, R. (2001). Emotions and economic behavior. PhD-thesis, University of Amsterdam.

Bosman, R. and van Winden, F. (2001). "Emotional hazard in a power-to-take experiment." *Economic Journal*, 112, 147–169.

Brandts, J., Saijo, T., and Schram, A.J.H.C. (2004). "A four country comparison of spite and cooperation in public goods games." *Public Choice*, 119, 381–424.

Brandts, J. and Schram, A. (2001). "Cooperation and noise in public goods experiments: applying the contribution function approach." *Journal of Public Economics*, 79, 399–427.

Camerer, C. (1998). "Bounded rationality in individual decision making." *Experimental Economics*, 1, 163–183.

Cason, T. and Mui, V. (2004). "Uncertainty and resistance to reform in laboratory participation games." *European Journal of Political Economy*, 21, 708–737.

ason, T., Saijo, T., and Yamato, T. (2002). "Voluntary participation and spite in public good provision experiments: an international comparison." *Experimental Economics*, (forthcoming).

Charness, G. and Rabin, M. (2002). "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117, 817–869.

Collier, K., McKelvey, R., Ordeshook, P., and Williams, K. (1987). "Retrospective voting: an experimental study." *Public Choice*, 53, 101–130.

Croson, R.T.A. and Marks, M.B. (2000). "Step returns in threshold public goods: a meta- and experimental analysis." *Experimental Economics*, 1, 239–259.

Davis, D.D. and Holt, C.A. (1993). *Experimental Economics*. Princeton: Princeton University Press.

Davis, D.D. and Reilly, R.J. (1998). "Do too many cooks always spoil the stew? An experimental analysis of rent-seeking and the role of a strategic buyer." *Public Choice*, 95, 89–115.

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

Eavey, C. and Miller, G. (1984). "Fairness in majority rule games with a core." *American Journal of Political Science*, 28(3), 570–586.

Fehr, E. and Fischbacher, U. (2003). "The Nature of Human Altruism." *Nature*, 425, 785–791.

Fehr, E. and Gächter, S. (2000a). "Fairness and retaliation." *Journal of Economic Perspectives*, 14, 159–181.

Fehr, E. and Gächter, S. (2000b). "Cooperation and punishment in public goods experiments." *American Economic Review*, 90, 980–994.

Fehr, E. and Schmidt, K.M. (1999). "A theory of fairness, competition and cooperation." *Quarterly Journal of Economics*, 114, 817–868.

Fiorina, M. and Plott, C.H. (1978). "Committee decisions under majority rule: an experimental study." *American Political Science Review*, 72, 575–598.

Goeree, J. and Holt, Ch. (2000). "An explanation of anomalous behavior in binary-choice games: entry, voting, public goods, and the volunteers' dilemma," mimeo, University of Virginia.

Großer, J., Elberfeld, W., and Giertz, T. (2002). "An experimental study of the polity game," mimeo, University of Amsterdam.

Großer, J. and Schram, A. (2006). "Neighborhood information exchange and voter participation— an experimental study." *American Political Science Review*, 100, 235–247.

Großer, J. and Schram, A. (2007). "Public opinion polls, voter turnout and welfare: an experimental study," mimeo, University of Amsterdam.

Güth, W. and Weck-Hannemann, H. (1997). "Do people care about democracy? An experiment exploring the value of voting rights." *Public Choice*, 91, 27–47.

Hoffman, E. and Plott, Ch. (1983). "Pre-meeting discussions and the possibility of coalition-breaking procedures in majority rule committees." *Public Choice*, 40, 421–439.

Holt, Ch. (2007). *Markets, Games and Strategic Behavior*. Boston: Pearson Education.

Holt, Ch. and Anderson, L. (1999). "Agenda and Strategic Voting." *Southern Economic Journal*, 65, 622–629.

Isaac, R.M., Walker, J., and Thomas, S. (1984). "Divergent evidence on free riding: an experimental examination of possible explanations." *Public Choice*, 43(1), 113–149.

Isaac, R.M., Walker, J., and Williams, A. (1994). "Group size and the voluntary provision of public goods: experimental evidence utilizing large groups." *Journal of Public Economics*, 54, 1–36.

Kagel, J.H. and Roth, A.E. (eds.) (1995). *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Kirchsteiger, G. and Prat, A. (2001). "Inefficient equilibria in lobbying." *Journal of Public Economics*, 82(3), 349–375.

Ledyard, J. (1984). "The pure theory of large two-candidate elections." *Public Choice*, 44, 7–41.

Ledyard, J. (1995). "Public goods: a survey of experimental research," in J. Kagel and A. Roth (eds.), *The Handbook of Experimental Economics*. Princeton: Princeton University Press, 111–194.

Levine, D. and Palfrey, Th. (2007): "The Paradox of Voter Participation? A Laboratory Study." *American Political Science Review*, 101, 143–158.

McKelvey, R. and Ordeshook, P. (1980). "Vote trading: an experimental study." *Public Choice*, 35(2), 151–184.

McKelvey, R. and Ordeshook, P. (1990). "A decade of experimental research on spatial models of elections and committees," in J.M. Enlow and M.J. Hinich (eds.) *Readings in the Spatial Theory of Voting*. Cambridge, England: Cambridge University Press, 99–144.

McKelvey, R.D. and Ordeshook, P.C. (1993). "Information and elections: retrospective voting and rational expectations," in D. Kinder and T. Palfrey (eds.) *Experimental Foundations of Political Science*. Michigan: The University of Michigan Press, 333–362.

McKelvey, R. and Palfrey, T. (1995). "Quantal response equilibria in normal form games." *Games and Economics Behavior*, 7, 6–38.

McKelvey, R. and Palfrey, T. (1998). "An experimental study of jury decisions," Discussion Paper, California Institute of Technology.

Millner, E.L. and Pratt, M.D. (1989). "An experimental investigation of efficient rent-seeking." *Public Choice*, 62, 139–151.

Niskanen, W. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton.

Offerman, T., Sonnemans, J., and Schram, A. (1996). "Value orientations, expectations, and voluntary contributions in public goods." *Economic Journal*, 106, 817–845.

Olson, M. and Morton, R. (1994). "Entry in spatial competition and strategic uncertainty," Discussion Paper, Tinbergen Institute, (TI 94–123), University of Amsterdam.

Palfrey, T.R. and Prisbrey, J.E. (1997). "Anomalous behavior in linear public goods experiments: how much and why?" *American Economic Review*, 87, 829–846.

Palfrey, T.R. and Rosenthal, H. (1983). "A strategic calculus of voting." *Public Choice*, 41, 7–53.

Plott, C.R. (1982). "Industrial organization theory and experimental economics." *Journal of Economic Literature*, 20, 1485–1587.

Potters, J., de Vries, C.G., and van Winden, F. (1998). "An experimental examination of rational rent-seeking." *European Journal of Political Economy*, 14, 783–800.

Potters, J. and van Winden, F. (1992). "Lobbying and asymmetric information." *Public Choice*, 74, 269–292.

Riedl, A. and van Winden, F.A.A.M. (2001). "Does the wage tax system cause budget deficits?" *Public Choice*, 109, 371–394.

Schram, A. (1991). *Voter Behavior in Economic Perspective*. Berlin Heidelberg New York: Springer.

Schram, A. (2000). "Sorting out the seeking: rents and individual motivation." *Public Choice*, 103, 231–258.

Schram, A. (2005). "Artificiality: the tension between internal and external validity in economic experiments." *Journal of Economic Methodology*, 12, 225–237.

Schram, A. and Sonnemans, J. (1996a). "Voter turnout as a participation game: an experimental investigation." *International Journal of Game Theory*, 25, 385–406.

Schram, A. and Sonnemans, J. (1996b). "Why people vote: experimental evidence." *Journal of Economic Psychology*, 17, 417–442.

Smith, V.L. (1994). "Economics in the laboratory." *Journal of Economic Perspectives*, 8, 113–131.

Sonnemans, J., Schram, A., and Offerman, T. (1998). "Public good provision and public bad prevention: the effect of framing." *Journal of Economic Behavior and Organization*, 34, 143–161.

Sutter, M. (1999). "Public bad prevention by majority voting on redistribution — experimental evidence," mimeo, University of Innsbruck.

van Winden, F. (1999). "On the economic theory of interest groups: towards a group frame of reference in political economics." *Public Choice*, 100, 1–29.

van Winden, F. and Bosman, R. (1996). "Experimental research in public economics," in *Experiments in Economics — Experimente in der Ökonomie*. Ökonomie und Gesellschaft, Jahrbuch 13, Campus Verlag, Frankfurt.

Weimann, J., Yang, C.-L., and Vogt, C. (2000). "An experiment on sequential rent-seeking." *Journal of Economic Behavior and Organization*, 41, 405–426.

Williams, K. (1991). "Candidate convergence and information costs in spatial elections: an experimental analysis," in T. Palfrey (ed.) *Laboratory Research in Political Economy*. Michigan: The University of Michigan Press, 113–135.

Wit, J. (1997). Dynamics and information aggregation in elections and markets. PhD-thesis, University of Amsterdam.

Chapter 33
The Political Economy of Swarming in Honeybees: Voting-with-the-Wings, Decision-Making Costs, and the Unanimity Rule[1]

Janet T. Landa

Introduction

Sociobiology has attracted a great deal of attention in the last decade or so. However, according to biologist Michael Ghiselin, the sociobiological approach has its limitations as it tries to explain all behavior including altruism, in terms of genetics. As Ghiselin puts it: 'Genes, of course, occur in all organisms, but it is the economic forces that really explain what organisms do.'[2] A 'bioeconomic' approach, incorporating cost- benefit calculations would better explain biological forms of organization.[3] The bioeconomic approach has been used by Becker (1981), Ghiselin (1978), Hirshleifer (1978, 1982), Tullock (1978), and Wilson (1978) in their work, and will also be used in this paper. The paper will explain the various aspects of the political economy of swarming in honeybees, focusing especially on the bees' collective choice of a new permanent nest site by the unanimous voting rule. The economic analysis draws on the work of Arrow (1974) on organizations, Buchanan and Tullock (1962) on the choice of Pareto-optimal voting rule and Schelling (1978) on critical mass phenomena. The paper will describe aspects of the political economy of honeybee swarming in Section 1. Section 2 uses economic analysis

---

[1] This paper is a revised and updated version of a paper published in *Public Choice*, 1986, 26–38.

[2] M. Ghiselin (personal communication, June, 1985).

[3] The term 'bioeconomic approach' is used by Ghiselin: 'Bioeconomics arose when a number of biologists and economists discovered that they were working on the same basic kind of problem, and that, furthermore, they had a great deal to learn from each other' (personal communication, June, 1985). According to Ghiselin (1982: 132–133) the economic approach to biology goes back to Darwin's work and even further back in time. According to Samuelson, writing on modes of thought in Economics and Biology: 'There is much territory between economics and biology that is still virgin ground. It will be tilled increasingly in the future' (1985: 172). And according to E.O. Wilson: 'I believe that insights from economists can add a great deal in helping even the purest of entomologists to look for new phenomena and theoretical constructions (personal communication, August 3, 1985).

to explain various aspects of swarming in honeybees. Section 3 provides a conclusion and suggests an extension of the line of research in this paper.

The Political Economy of Swarming in Honeybees

The classic study of swarming in honeybees is by Lindauer (1961). Swarming is the process whereby a bee colony divides itself into two groups: one group staying behind with a new queen, while the other group (the swarm) flies off with the old queen to establish a new colony (Lindauer, 1961; von Frisch, 1967; Wilson, 1971; Michener, 1974; Free, 1977; Seeley, 1982).[4] Among the various causes of swarming such as the presence of disease, poor temperature regulation, etc., overcrowding/congestion appears to be a critical factor. Swarming occurs in late spring to early summer in temperate climates when food supplies are plentiful. The abundance of food supplies facilitates the swarm to establish a new colony before the cold weather sets in.

The warming process is divided into two stages: one occurring within the nest and the other occurring outside the nest. The first stage is associated with the rearing of additional queens, so that at least one new queen remains in the parent colony. The workers construct queen cells in order to produce new queens. The nurse bees load these cells with large amounts of nutrient gel ('royal jelly'). Before the virgin queens emerge as adults, foraging bees are observed to be shirking from work in response to house bees' increasing unwillingness to accept food from foragers. Some of the foragers, unable to continue foraging begin to hunt for new nest sites. The former foraging bees have become scout bees, thus providing an example of 'sociogenesis, the procedures by which individuals undergo changes in caste, behavior, and physical location incident to colonial development' (Wilson, 1985: p. 1490).

The first virgin queen to emerge usually stings her potential rivals to death while they are still in their cells. Should two or more queens emerge more or less at the same time, they fight to the death until one queen is victorious. The victorious queen will leave on a mating flight and will return inseminated for life. She will either take over the parent nest or will participate in an afterswarm herself. Meanwhile, the old queen ceases to be an egg-laying machine as worker bees feed her less and less to ensure that she will be able to fly some considerable distance to the new nest site.[5] Before departing as a swarm from the parent nest, bees fill their stomachs with honey so that they can postpone foraging for food for a few days while they wait for scout bees to find them a new home.

---

[4] For a more detailed description of the socio-economic organization of honeybees, in-cluding a brief description of the swarming process, see J. Landa and A. Wallis (1985).

[5] Ghiselin has provided an explanation as to why the *old* queen, and not the young queen, leaves the nest: 'The older bee, the mother in the hive, leaves with a swarm of her daughters. Why is this? Well, I think the reason... is that she is leaving a group of relatives, so she gets something from providing her daughters with a very good place to stay. Also, she is probably an old bee and going to die pretty soon, so she should take the risk. If she were sending out a lot of propagules, as termites do, maybe she would be betteroff remaining in the hive (1982: 143).

The actual departure of the swarm occurs about a day or two before the emergence of the first young queen. In preparation for the actual swarm's flight from the parent nest, workers bees perform buzzing runs (*Schwirrlauf*). As they dance, bees intentionally touch other bees who, in turn, also begin to do buzzing runs and this spreads rapidly among the bee population until the swarm emerges from the nest and flies off with the old queen. The swarm is approximately half the original bee population of roughly 30,000 or so workers. It is not known how the population sorts itself into two sub-groups since swarm bees are workers of all ages. During the flight from the parent nest, and later-during the flight to the new nest site, the queen plays a crucial role, via pheromone odor, in maintaining order and cohesion of the swarm as an organized unit. If the queen is lost, the swarm becomes disorderly and will return to the old nest. The swarm, after emerging from the old nest, will fly a short distance and cluster at a temporary exposed site such as a tree; and there they wait for the scout bees to find them a new home.

The second stage of the swarming process is associated with the scout bees searching for potential nest sites and arriving at a unanimous agreement regarding the choice of a particular nest site. It is to be emphasized that only a very small sub-group of the swarm is involved in the searching and voting process -only scout bees constituting about 5% of the swarm. When a scout bee finds a suitable site, it performs a dance on the outside of the surface of the swarm. The tail wagging dance transmits information to other scout bees regarding the distance and direction of the nest site. Many potential nest sites therefore can be announced by scout bees. The scout bees convey information regarding the quality of the potential nest site by the vigour and duration of their dance; the more lively the dance, the better the quality of the potential nest site. A 'democratic' voting process is employed by scout bees in arriving at their choice:

> . . . the agreement appears to take place in a seemingly simple manner; the better the qualities a nesting place exhibits, the livelier and longer will be the messengers' dance after the inspection. In this way new messengers are recruited in the cluster for this place, which then likewise seek out and inspect this nesting place, and then they too solicit by means of the same lively dances. If those scouting bees which at first had only inferior or average dwellings to announce are persuaded by the livelier dances of their colleagues to inspect the other nesting place, then nothing more stands in the way of an agreement. They can now make a comparison between their own and the new nesting place, and they will solicit in the cluster for the better of the two (Lindauer, 1961: p. 50).

The process of soliciting or 'lobbying' for the best site continues until all conflicts and disagreements are democratically resolved and the scout bees converge to a unanimous choice. Lindauer has described in detail the actual process whereby a particular swarm arrives at its unanimous decision: 21 nesting places were discovered by scouting bees of this swarm, and it took five days before one particular nesting place became the unanimous choice. See Fig. 33.1.

Lindauer found that the most important reason delaying reaching consensus among the scout bees is that the bees wanted to ensure that the best site, from all possible potential nest sites that have been discovered, will be chosen. Being experienced foragers, scout bees can judge the quality of nesting sites, and determine the
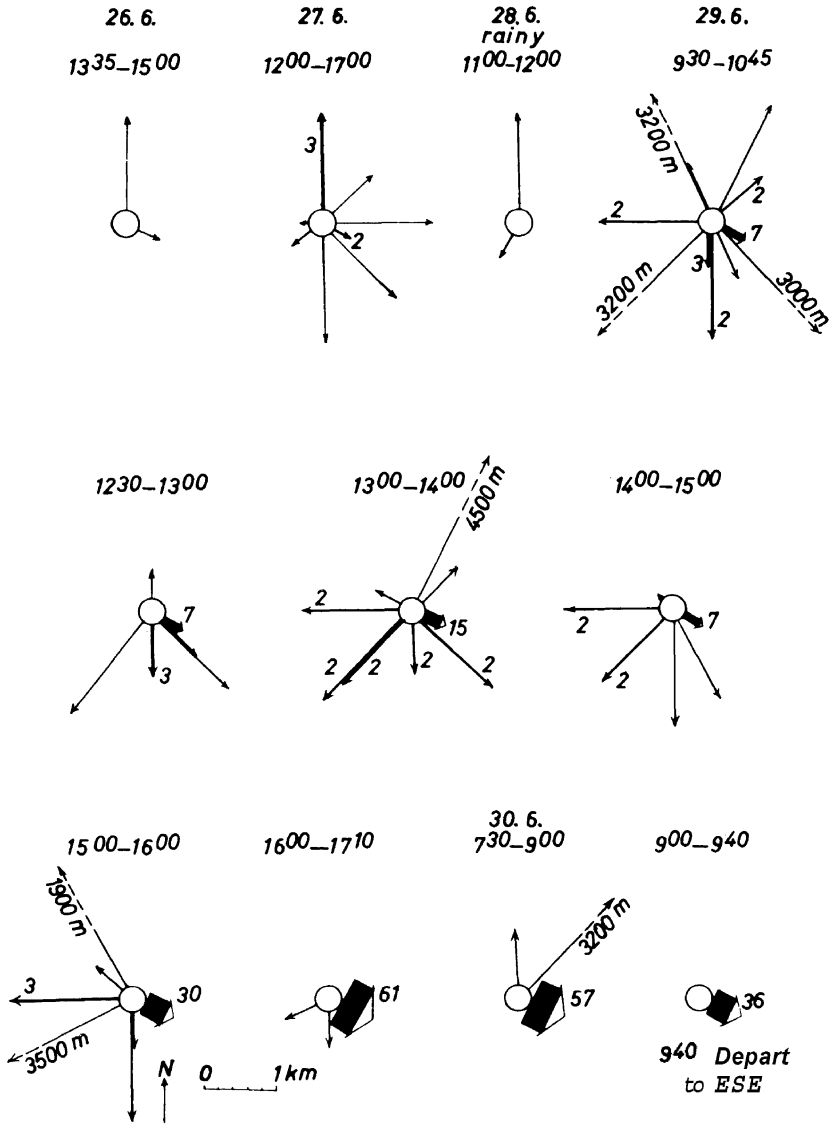
Fig. 33.1 The process of reaching unanimous agreement among Lindauer's scout bees (Source: Lindauer, 1961: 42)

properties which make a particular nesting site the 'best' site. Seeley explains why bees must choose their nesting sites very carefully:

> ... if a colony does survive the critical first winter, it will endure on the average for another five years. In short, a colony has the potential to survive for a long time but faces great risks in moving from an old nest to a new one. Therefore a swarm cannot rely on trial-and-error methods in finding a suitable site. Each colony must make a single, careful decision with which it can live for many years (Seeley, 1982: p. 168).

Because finding a suitable site is critical for colony's survival, it is rare for scout bees not to come to a unanimous agreement. In fact, out of 19 swarms observed by Lindauer, only 2 swarms failed to reach agreement. In the first case:

> ... two groups of messengers had got into competition: one group announced a nesting place to the northwest, the other one to the northeast. Neither of the two wished to yield. The swarm then finally flew off and I could scarcely believe my eyes -it sought to divide itself. The one half wanted to fly to the northwest the other to the northeast. Apparently each group of scouting bees wanted to abduct the swarm to the nesting place of its own choice. But that was naturally not possible, for one group was always without the queen, and there resulted a remarkable tug of war in the air, once 1Immediately both groups began again with their soliciting dances, and it was not until the next day that the northeast group finally yielded; they ended their dance and thus an agreement was reached on the nesting place in the northwest (Lindauer, 1961: p. 45).

Example of the recorded announcements of the scouting bees from the moment of swarming at 13:35 on 26 June until the moving into the new dwelling on 30 June. Each arrow represents a newly marked scouting bee that announced a nesting place by means of a dance in the cluster. Direction and length of the arrow give to scale the location of the reported nesting place. First of all nesting places in different directions and distances are announced but gradually an agreement is reached for the nesting place 350 meters to the southeast.

It is possible that in this first case, the swarm did not survive the winter. The second case ended in a totally unexpected way:

> ... for 14 days no agreement had been reached, and then when a period of rain set in, the scouting bees gave up their search for a dwelling and occupied themselves instead with the collection of nectar and pollen. The traveling stores of the swarm bees were apparently used up and it was high time for a replacement of provisions. Thus the activity of the hunters of quarters was completely suppressed, and the swarm made its abode at its first landing place, built honey combs in the bushes, and set up a normal nest for its brood (Lindauer, 1961: p. 45).

Regarding the first case, what caused the two subgroups of scout bees to disagree on the choice of a particular nest site? Could it be that the failure to agree is due to the fact that the two subgroups were genetically different so that both groups had difficulties in communicating with one another? This may have been the case. Lindauer himself, reporting on research pointed out that mixed colonies consisting of different species of bees have communication problems:

> If, for example, an Austrian bee receives information from an Italian bee about a food place 100 meters from the nest, she will fly 120 meters, because she interprets the 'Italian dialect' in her Austrian way. And conversely, the Italian bee will fly only 80 meters when given information for 100 meters by an Austrian bee. Similar misunderstandings also exist between other races. It is now clear that each geographic race has its own dialect (Lindauer, 1961: p. 61)

The idea of a swarm failing to agree on a particular site because of genetic heterogeneity of the swarm has crossed Dawkins' mind:

> There is no suggestion here that the two subgroups of bees were genetically different, *though they may have been* [italics mine]. What matters to the point I am making is that each individual follows local behavioural rules, the combined effect of which normally gives rise to coordinated swarm behaviour. These evidently include rules for resolving

'disputes' in favour of the majority. Disagreements over the preferred location for the outer wall of a termite mound might be just as serious for colony survival as disagreements over nesting sites among Lindauer's bees (colony survival matters, because of its effects on the survival of the genes causing individuals to resolve disputes). As a working hypothesis, we might expect that disputes resulting from genetic heterogeneity in termites would be resolved by similar rules (Dawkins, 1982: p. 206).

Once unanimous decision regarding choice of a new home is reached, scout bees prepare the swarm to lift-off. They perform zigzag runs are accompanied by wing buzzing to break up the cluster. The whole swarm then flies off. Approaching the nest site, scout bees will signal the swarm to stop. Soon the bees will be streaming inside the new home and:

> Within 30 minutes of lift-off nearly all the bees are safely inside their new home. Within a few hours they are cleaning out debris, constructing combs and flying off to forage for nectar and pollen. A new colony has been established (Seeley, 1982: p. 158).

Voting-with-the-Wings, Decision-Making Costs and the Unanimity Rule

There are two different voting procedures discussed in public choice theory:[6] traditional voting by ballot, and 'voting-with-the-feet' (Tiebout, 1956). The latter procedure sorts people into homogeneous groups of like tastes by allowing people to leave a community and enter another which has a mix of public goods which satisfy the preferences of the incoming population.

Both kinds of voting procedures are found in the bee society during the two-stage swarming process. The first stage of the swarming process may be viewed as 'voting-with-the-wings' procedure which sorts the original bee population into two homogeneous subgroups: those who stay and those who exit (the swarm) to establish a new colony (viewed as a public good).[7] The second stage of the swarming process involves scout bees reaching unanimous agreement regarding the choice of a new nest site. The voting procedure is the traditional voting by ballot (or 'voting by ballet/dancing in the case of bees). We will provide answers to the following questions:

1. What are the principles involved in the sorting process by which the original population is divided into two subgroups?
2. Why is voting for a new nest site not conducted within the parent nest, but outside the nest by the swarm?
3. Why is searching for the nest sites and voting for a particular nest site restricted solely to a small subgroup of scout bees?
4. Why do scout bees use the unanimity rule in arriving at their collective choice of a new nest site?

---

[6] Public choice theory is the 'Economics of politics', the application of economics to the study of political For a survey of public choice literature, see Mueller (1979).

[7] For a theory of the homogeneous occupational group the Chinese middleman group – as a low cost organization for contract enforcement, see J. Landa, see also J. Carr and J. Landa (1983).

5. What explains the high degree of cooperation and altruism exhibited by scout bees in their search for a new home; why do scout bees not free- ride on the house hunting efforts of other scout bees?

As noted earlier, the original bee population sorts itself into two subgroups. But it is not known what principles are involved in the sorting process. Perhaps the dynamics of population division leading to the emergence of the swarm, falls into the class of 'critical mass' phenomena (Schelling, 1978) involving some activity that is self-sustaining. Once that activity passes a certain minimum level, the original bee population sorts itself into two subgroups. The swarming process is analogous to the phenomenon of 'tipping-in' and 'tipping-out' (Schelling, 1978) applied to neighborhood migration, a special case of the critical mass phenomenon:

> Not only was the departure of a white population induced by the appearance of minorities, but minorities themselves would be more attracted the larger the minority colony and the faster its growth, with some minimum size required to get a self-sustaining influx started (Schelling, 1978: p. 101).

As noted earlier, voting for the new site is undertaken only after the swarm has emerged and left the nest. The reason is that if voting for the new nest site is conducted within the parent nest, information must be transmitted to all the bees within the nest and information is costly. It is clearly optimal to reduce the information costs by economizing on internal communication channels (Arrow, 1974).Thus, transmitting information only to members of the swarm, which constitute about half of the original bee population, greatly economizes on information costs. The 'voting-with-the-wings' procedure is both a voting and sorting procedure which greatly economizes on information costs because the swarm is the smallest unit consistent with the collective provision of a public good (the new colony).

Since searching for potential nest sites and voting on the choice of a particular nest site are costly activities, it is clearly efficient if activities are restricted to a small sub-group of specialists who possess the requisite skill and information in assessing the quality of the potential nest site. Scout bees are specialists since they are former experienced foragers. Their foraging skills turn on their ability to find new sources of food and the skill can be extended to searching for new nest sites. Furthermore, as Adam Smith once suggested, members of the same trade find it easy to communicate with each other because of their shared experiences. Thus it may well be that scout bees, being a highly homogeneous occupational group, can more easily arrive at a consensus regarding the choice of a particular nest - site. This may also partly explain why scout bees use the unanimity rule in arriving at their collective choice of a new nest site.[8]

To fully understand why scout bees use the unanimity rule in arriving at their choice, we need to understand the nature of 'external costs' and 'decision-making costs'. According to Buchanan and Tullock (1962) the least costly or the Pareto-optimal

---

[8] B. Grofman has suggested that it may be that bees cannot count, therefore it is easier to use the unanimity rule than the simple majority rule ('how do they know when they have 51% of the vote?'). Also, the fact that bees are swarming on a branch, *exposed* to the elements, gives them an added incentive to converge to an unanimous vote. It is analogous to the sequestration of juries to hasten reaching of unanimous jury decision(personal communication, June 8, 1985).

decision-making rule is the rule that minimizes both external costs and decision-making costs. From an individual's point of view, external costs are the costs imposed upon the person when an issue is passed by others without the person's consent. The external cost function *(C)* is downward sloping: when one person imposes his choice upon the group, the expected external costs are very high; external costs decrease as the number of individuals in the group who agree with each other increases and when all members of the group agree on an issue, external costs are reduced to zero. Decision-making costs are the time and effort and direct expenditures involved in persuading others in the group to agree. The decision costs function (D) is upward sloping: when one person is present, decision-making costs are zero, when two persons have to agree, decision costs become positive and these costs increase as the size of the group increases. The Pareto optimal rule is the rule that minimizes the sum of these two costs. There is no a priori reason why the simple majority rule is the least costly rule. The optimal rule depends on how high decision-making costs are relative to external costs. Where decision-making costs are extremely high, the one-person (dictatorship) rule is efficient; where external costs are very high relative to decision-making costs, the least costly rule is the unanimity rule. Figure 33.2 shows that the Pareto optimal rule is a simple majority rule.

Not only does the size of the group determine the magnitude of decision- making costs, but also the degree of homogeneity of the group. The more homogeneous the group, the more the individual favors a more inclusive majority rule because both external costs and decision-making costs are lower in a homogeneous group with similar tastes. Clearly, reducing the size of the voting group and increasing the degree of homogeneity of the voting group economizes on decision-making costs.
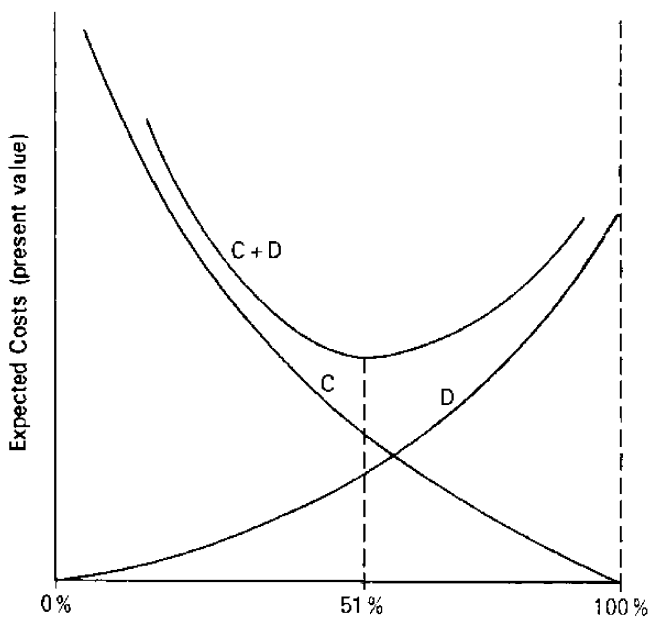
Fig. 33.2 The Buchanan -Tullock model: Simple majority rule as the efficient making rule

To reduce the size of the voting group, the larger group can delegate its voting rights on a particular issue to a small committee of specialists which represents the larger group's interests. Thus a shift from direct democracy to representative democracy economizes on decision-making costs.[9]

In the context of voting behavior in bees, restricting voting rights to a small subgroup of scout bees is efficient because: a) the size of the voting group is greatly reduced, since it is about 5% of the swarm; and b) of the high degree of occupational homogeneity of the scout bee subgroup. Thus the reduction in the size of the voting group together with the high degree of homogeneity of group membership means that decision-making costs are low. On the other hand, external costs are extremely high if scout bees use the 'one-bee' (dictatorship) rule, in which any one bee can impose its choice of a nesting site on the rest of the group. In the context of bees, the term 'external costs' requires some explanation. Since any bee can gain fitness *only* through the queen, there is really no question of external costs here. What corresponds to 'external costs' in this context boils down to a difference of opinion about what is best for the queen. The more inclusive the rule, the more certain the bee can be that its own choice of a nest site will not be over-ridden.' External costs are still high even when a majority of scout bees agree on a particular nest site since it may well be that the best site is discovered by a scout bee who is not yet in the majority coalition. Because of the high external costs relative to decision-making costs, the use of the unanimity rule by scout bees is efficient.

A key element in the process of arriving at a unanimous agreement is that individual bees are 'programmed' to *change their minds*, when making a binary choice of their own nest sites compared with other nest sites, always voting for the better quality nest.[10] In this way, a process of proto-coalition formation takes place in which more and more scout bees switch their allegiance and join the majority, voting for the best site until a grand coalition of all scout bees, voting for the best site emerges. This, together with the fact that each scout bee has only one preferred option, makes it possible for the bee society to avoid the well-known paradox of voting/cyclical majorities problem (Arrow, 1951) connected with any form of comparison majority rule.[11]

---

[9] I am indebted to the late Jack Hirshleifer for this personal communication, June 19, 1985.

[10] For a view that does not credit bees with rationality, see J. Bennett (1964).

[11] For an analysis of a proto-coalition formation process among a network of traders in the 'Kula Ring' gift-exchange system in the Trobriand Islands of Papua New Guinea, see B. Grofman and J. Landa (1983).

[12] For a survey of the social choice literature dealing with the voting paradox and related problems of collective choice in democratic societies, see D. Mueller (1979). For an applica-tion of social choice theory to voting in corporate law, see F.H. Easterbrook and Daniel R. Fischel (1983). The authors explain why only one class of participants in the corporation (the shareholders) hold dispositive voting rights at one time: 'It is well known, however, when voters hold dissimilar preferences it is not possible to aggregate their preferences into a consistent system of choices. ... The preferences of one class of participants are likely to be similar if not identical. This is true of shareholders, especially for people who buy and sell in the market so that the shareholders of a given firm at a given time are a reasonably homogeneous group with respect to their desires for the firm. So firms with single classes of voters are likely to be firms with single objectives, and single-objective firms are likely to prosper relative to others (p. 405).'

In their search for a new home, scout bees exhibit a high degree of cooperation among themselves. But because a new colony is a public good in which no bee can be excluded once it is established, a prior question arises: Why do we not observe scout bees exhibiting Prisoner's Dilemma behavior in choosing to free-ride on the efforts of other scout bees, thus shirking behavior among scout bees? Scout bees have no incentive to free-ride on other scout bees' efforts because the search for a suitable nest site is critical to the survival of the whole swarm, including the scout bees. If every scout bee free- rides, then the whole swarm cannot survive in the exposed temporary site However, in the process of searching for potential nest sites, scout bees may encounter dangers that bring death. Why then do we see altruism in scout bees who, despite occupational hazards, continue to hunt for suitable nest sites?

Here the work of socio-biologists can shed light on the question. Socio- biologists have emphasized the importance of genetic kinship (Hamilton, 1964; Dawkins, 1976, 1982) and reciprocity (Trivers, 1971) as the basis of altruism and cooperation in biological systems.[13] Altruism in scout bees - and even more so in guard bees (Landa and Wallis, 1985) -may be explained as the gene's eye view of natural selection (Dawkins, 1976, 1982). In the words of Axelrod and Hamilton:

> A gene, in effect, looks beyond its mortal bearer to the potentially immortal set of its replicas existing in other related individuals. If the players are sufficiently closely related, altruism can benefit reproduction of the set, despite losses to the individual altruist. In accord with this theory's predictions, almost all clear cases of altruism, and most observed cooperation -apart from their appearance in the human species -occur in contexts of high relatedness, usually between immediate family members. The evolution of the suicidal barbed sting of the honeybee worker could be taken as the paradigm for this line of theory (Axelrod, 1984: 89–90).

A stronger argument than that provided by Hamiltonian altruism is provided by Tullock: 'The worker bee simply cannot reproduce and hence has no prospect at all of reproducing her own genes directly. It is only by saving the queen that any of her genes may be made to survive.[14] Wilson (1985) citing the work of West-Eberhard, has argued that as death approach for insects, the 'optimum strategy for contributing genes to the next generation is to enhance colony welfare through more dangerous occupations such as defense and foraging, thus producing more brothers and sisters as opposed to personal offspring (p. 1494).' By this criterion, scout bees are engaged in a dangerous occupation which contributes genes to the next generation

---

[13] However, according to Ghiselin: 'In the *Origin of Species,* Darwin (who knew essentially nothing about genetics) explained the so-called altruism of bees on the basis of each society being selected as a single unit (an extended family). The colonies with the most economically efficient workers out-produce the less efficient colonies, and this occurs in spite of the fact that the workers are sterile. We need not consider such mattersfrom the point of "selfish genes". (Personal communication, June, 1985).

[14] G. Tullock (personal July 15, 1985). For this line of argument, see also G.F. Oster and E.O. Wilson (1978): 'Caste members are programmed to behave altruistically. They surrender most or all of their personal reproductive capacity in favor of the mother queen, undertake risky foraging trips, and sometimes literally throw their lives away in frenzied defense of the nest. What matters is not their personal survivaland reproduction, but rather that of the queen (p. 161).'

by enhancing colony survival. And Dawkins has introduced the concept of the 'extended phenotype' defined as 'all effects of a gene upon the world'; more specifically 'the effects [which] influence the survival chances of the gene, positively or negatively (Dawkins, 1982: p. 286). Dawkins gave an example of the beaver building a dam across the stream creating a lake; the lake may be regarded as 'a huge extended ph'enotype, extending the foraging range of the beaver in a way which is somewhat analogous to the web of the spider (Dawkins, 1982: p. 200).' Viewed in this way, the newly established bee nest site may be regarded as an extended phenotype, allowing the swarm to continue to forage in a new territory, away from the parent nest, thus lessening colonies competing against one another as independent units. The new site, especially when it is a suitable one, increases the survival chances of the new colony.

Conclusions

The political economy of bee swarming offers a fascinating study of collective action in biological systems. Various aspects of the organization of swarming in honeybees have been explained in this paper as devices bees have evolved to economize on information and decision-making costs associated with the economics and politics of establishing a new nest site that would ensure the swarm's survival, which in turn, would bring forth a new generation of bees.

There are opportunities for extending this line of research. One example is to explain schooling behavior of fishes.[15] Biologist E. Shaw (Extending the line of research in this paper to schooling fishes is suggested by Ghiselin (personal communication, June, 1985). Biologist E. Shaw (1978) has pointed out that schooling fishes remain an exception among vertebrates in that their organization is non-hierarchical, consisting of a 'truly egalitarian state in which all members of the social group are alike in influence and importance. In the school, no fish dominates another and no peck order or hierarchy exists. The fate of a few could well be the collective fate of all because schooling fish act in synchrony behaving as one toward a source of food or away from the engulfing sweep of a predatory trawl' (E. Shaw, 1978: p. 166). Using the concepts of 'external costs' and decision-making costs, it may be that the egalitarian form of organization is the least costly organizational form in large schools of fish where the cost of establishing leadership is prohibitive.

Afterword, December 2007

My paper was written over twenty years ago, and published in *Public Choice* 51, 25–38, 1986). It was my first bio-economics paper, using economic theory (specifically, public choice theory), to examine aspects of the swarming of honeybees.

---

[15] Extending the line of research in this paper to schooling fishes is suggested by Ghiselin (personal communication, June, 1985).

I used the Buchanan-Tullock cost approach to the choice of Pareto-optimal rules for group decision-making to analyze Martin Lindauer's (1961) pioneering experimental findings that scout bees use the unanimous voting rule to arrive at their choice of a new nesting site.

New experimental findings by entomologist Thomas Seeley and his co-workers since the 1990s (Seeley et al. 2006*) found that the essence of group decision-making by scout bees is their use of a kind of *quorum-sensing* ("sufficient number of scouts" ranging from 10–20 bees) voting rule, rather than the consensus rule, in arriving at their collective choice of a sufficiently "good", often the "best" but not necessarily always the "best" nest site.

The new experimental findings by Seeley and his co-workers necessitate a revision of my theory that the unanimous voting rule is the Pareto-optimal rule for scout bees' collective choice of their best new nest site. A meeting with Thomas Seeley at a biology conference—the 33rd Biology Symposium, "The Biology and Evolution of Communication Systems" held at York University, March 11, 2006 led to a co-authored paper by Janet Landa & Thomas Seeley entitled, "Honeybees Swarms Making Decisions about Choice of a New Nest: New Empirical Evidence and a Refinement of 'Voting with the Wings' - A Bioeconomics and Public Choice approach." (in progress).

Our paper discusses Seeley's new empirical findings of aspects of the bees' swarming behavior, focusing on bees' individual as well as group decision-making in their search for a new nest site. Based on the new empirical findings, the paper provides a refinement of my theory of the Pareto-optimal rule for honeybee scouts' collective choice of a new nest site. Our paper has practical policy implications regarding efficient decision-making rules for all kinds of groups in non-human and human societies.

References

Arrow, K.J. (1951). *Social choice and individual values*. New York: Wiley.

Arrow, K.J. (1974). *The limits of organization*. New York: Norton.

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Becker, G. (1981). Families in nonhuman species. In G. Becker (Ed.), *A treatise on the family*. Cambridge: Harvard University.

Bennett, J. (1964). *Rationality: An essay toward an analysis*. London: Routledge and Paul.

Buchanan, J.M. and Tullock, G. (1962). *The calculus of consent*. Ann Arbor: The University of Michigan Press.

Carr, J. and Landa, J.T. (1983). The economics of symbols, clan names and religion. *Journal of Legal Studies* 12: 135–156.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Dawkins, R. (1982). *The extended phenotype*. Oxford: Oxford University.

Easterbrook, F.H. and Fischel, D.R. (1983). Voting in corporate law. *Journal of Law Economics* 26: 395–427.

Free, J.B. (1977). *The social organization of honeybees*. London: Edward Arnold.

Ghiselin, M.T. (1978). The economy of the body. *American Economic Review* 68: 233–237.

Ghiselin, M.T. (1982). General discussion: Evolution and human behavior. In P. Rubin (Ed.), *Research in law and economics* 4: 143.

Grofman, B. and Landa, J.T. (1983). The development of trading networks among spatially separated traders as a process of proto-coalition formation: The Kula trade. *Social Networks 5:* 347–365.

Hamilton, W.D. (1964). The genetical evolution of social behaviour I. *Journal of Theoretical Biology* 7: 1–16.

Hirshleifer, J. (1978). Natural economy versus political economy. *Journal of Social and Biological Structure* 1: 319–337.

Hirshleifer, J. (1982). Evolutionary models in economics and law: Cooperation versus conflict strategies. In P. Rubin (Ed.), *Research in Law and Economics* 4: 1–60.

Landa, J.T. (1981). A theory of the ethnically homogeneous middleman group: An institutional alternative to contract law. *Journal of Legal Studies* 10: 349–362.

Landa, J.T., and Wallis, A. (1985). *The socio-economic organization of honeybees: The new law and economics approach to sociobiology*. Paper presented at the Public Choice Society Annual Meeting, New Orleans, Feb. 21–23.

Lindauer, M. (1961). *Communication among social bees*. Cambridge: Harvard Press.

Michener, C.D. (1974). *The social behaviour of the bees*. Cambridge: Harvard University Press.

Mueller, D. (1979). *Public choice*. Cambridge: Cambridge University Press.

Oster, G.F. and Wilson, E.O. (1978). *Caste and ecology in the social insects*. Princeton: Princeton Press.

Samuelson, P.A. (1985). Modes of thought in economics and biology. *American Economic Review* 75: 166–172 (Papers and Proceedings of the Ninety-Seventh Annual Meeting of the American Economic Association).

Schelling, T.C. (1978). *Micromotives and macrobehavior*. New York: Norton.

Seeley, T.D. (1982). How honeybees find a home. *Scientific American* 247: 158–168.

Seeley, T.D., Visscher, P.K., and Passino, K.M. (2006). Group decision-making in honey bee swarms. *American Scientist* 94: 220–229

Shaw, E. (1978). Schooling fishes. *American Scientist* 66: 166–175.

Tiebout, C.M. (1956). A pure theory of local expenditures. *Journal of Political Economy* 64: 416–424.

Trivers, R.L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology* 46: 35–57.

Tullock, G. (1978). Altruism, malice, and public goods. *Journal of Social and Biological Structure* 1: 3–10.

von Frisch, K. (1967). *The dance language and orientation of bees*. London: Oxford University Press.

Wilson, E.O. (1971). *The insect societies*. Cambridge: Harvard University Press.

Wilson, E.O. (1978). The ergonomics of caste in the social insects. *The American Economic Review* 68: 25–35.

Wilson, E.O. (1985). The sociogenesis of insect colonies. *Science* 228: 1489–1495.

Index