THE ELGAR COMPANION TO LAW AND ECONOMICS

The Elgar Companion to Law and Economics

Second Edition

Edited by
Jürgen G. Backhaus

Krupp Chair in Public Finance and Fiscal Sociology, Erfurt University, Germany

Edward Elgar
Cheltenham, UK • Northampton, MA, USA

© Jürgen G. Backhaus 2005

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise without the prior permission of the publisher.

Published by
Edward Elgar Publishing Limited
Glensanda House
Montpellier Parade
Cheltenham
Glos GL50 1UA
UK

Edward Elgar Publishing, Inc.
136 West Street
Suite 202
Northampton
Massachusetts 01060
USA

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

ISBN 1 84542 032 2 (cased)

Typeset by Manton Typesetters, Louth, Lincolnshire, UK
Printed and bound in Great Britain by MPG Books Ltd, Bodmin, Cornwall

Contents

List of figures ix
List of tables x
List of contributors xi

Introduction 1
Jürgen G. Backhaus

PART I BASICS OF THE LAW AND ECONOMICS APPROACH

1 Coase theorem and transaction cost economics in the law 7
Francesco Parisi
2 Property rights and their partitioning 40
Christian Müller and Manfred Tietzel
3 Legal change in economic analysis 53
John N. Drobak and Douglass C. North
4 Positive, normative and functional schools in law and economics 58
Francesco Parisi
5 Commons and anticommons 74
Francesco Parisi and Ben Depoorter

PART II PRIVATE LAW AND ECONOMICS

6 The economics of tort law 87
Giuseppe Dari Mattiacci and Francesco Parisi
7 Family 103
Margaret Brinig
8 Inheritance 119
Richard E. Wagner
9 Intellectual property and the markets of ideas 127
Giovanni B. Ramello
10 Incomplete contracts and institutions 145
Antonio Nicita and Ugo Pagano

PART III PUBLIC LAW AND ECONOMICS

11 Central bank 165
Željko Šević

12 Constitutional economics I 184
Francesco Farina
13 Constitutional economics II 223
Ludwig Van den Hauwe
14 Administrative law and economics 239
Jean-Michel Josselin and Alain Marciano
15 Property 246
Thomas J. Miceli
16 The European Union's institutional design 261
Elisabetta Croci Angelini
17 Subsidiarity 280
Jürgen G. Backhaus

PART IV LABOUR LAW AND ECONOMICS

18 Labour contracts 289
Don Bellante
19 Company board representation 297
Jürgen G. Backhaus
20 Employment security through dismissal protection: market versus policy failures 311
Christoph F. Buechtemann and Ulrich Walwei

PART V REGULATION, TAXATION AND PUBLIC ENTERPRISE

21 Structures of public enterprise 329
Jürgen G. Backhaus
22 Environmental law and economics 345
A. Allan Schmid
23 Environmental policies choice as an issue of informational efficiency 350
Donatella Porrini
24 Tradable emission rights 364
Edwin Woerdman
25 Regulatory taxation 381
Jürgen G. Backhaus

PART VI DISPUTE RESOLUTION

26 Dispute resolution 393
Thomas J. Miceli

PART VII DIFFERENT SOURCES OF THE LAW

27 Judicial independence 407
Sophie Harnay
28 General norms and customs 424
Jean-Michel Josselin and Alain Marciano
29 Science as a source of law 433
Peter R. Senn
30 Social science as a source of the law 442
Peter R. Senn
31 Cognitive science 453
John N. Drobak
32 Connections with law and society research 459
Jürgen G. Backhaus

PART VIII TOWARDS AN IDEAL ECONOMIC ANALYSIS OF A LEGAL PROBLEM

33 Towards an ideal economic analysis of a legal problem 465
Jürgen G. Backhaus

PART IX CLASSICAL AUTHORS IN LAW AND ECONOMICS

34 Cesare Beccaria (1738–94) 475
Francesco Parisi and Giampaolo Frezza
35 Franz Böhm (1895–1977) 489
Heinz Grossekettler
36 John R. Commons (1862–1945) 499
Warren J. Samuels
37 Walter Eucken (1891–1950) 508
Leland B. Yeager
38 Otto von Gierke (1841–1921) 519
Jürgen G. Backhaus
39 Augusto Graziani (1865–1938) 522
Giampaolo Frezza and Francesco Parisi
40 Robert Lee Hale (1884–1969) – legal economist 531
Nicholas Mercuro, Steven G. Medema and Warren J. Samuels
41 Friedrich August von Hayek (1899–1992) 545
Ludwig Van den Hauwe
42 Theodor Herzl (1860–1904) 559
Peter R. Senn

43 Rudolf von Jhering (1818–92) and the economics of justice 568
J.L.M. Elders
44 Franz Klein (1854–1926) 576
Peter Lewisch
45 Etienne Laspeyres (1834–1913) 585
Wolfgang Drechsler
46 Friedrich List (1789–1846) 590
Arno Mong Daastöl
47 Achille Loria (1857–1943) 607
Giampaolo Frezza and Francesco Parisi
48 Karl Marx (1818–83) and Friedrich Engels (1820–95) 618
Heath Pearson
49 Carl Menger (1840–1921) 627
Richard E. Wagner
50 Plato (c. 427–349 BC) 635
Wolfgang Drechsler
51 Wilhelm Roscher (1817–94) 642
Erich Streissler
52 Emil Sax (1845–1927) 652
Manfred Prisching
53 Gustav von Schmoller (1838–1917) 662
Helge Peukert
54 Adam Smith (1723–90) 672
Helge Peukert
55 Werner Sombart (1863–1941) 683
Günther Chaloupek
56 Lorenz von Stein (1815–90) 689
Heinz Grossekettler
57 George Joseph Stigler (1911–92) 700
Peter R. Senn
58 Pietro Trimarchi (1934–) 709
Giampaolo Frezza and Francesco Parisi
59 Thorstein Veblen (1857–1929) 727
Heath Pearson
60 Max Weber (1864–1920) 733
Helge Peukert
61 Christian Wolff (1679–1754) 745
Wolfgang Drechsler

Index 751

Figures

5.1	Game matrix	79
10.1	The unilateral hold-up problem	149
10.2	The bilateral hold-up problem	149
12.1	Role of institutions in economic integration	214
13.1	Classic prisoner's dilemma	229
13.2	Modified prisoner's dilemma	233
16.1	Pay-off matrix of the EU	274
16.2	Games: pay-off matrices	276
16.3	Games: tied hands matrices	278
23.1	Scheme of different environmental policy instruments	351
32.1	Sombart's modern capitalism	460

Tables

7.1	A systematic representation of family law	104
15.1	Optimal remedies for externalities	251
16.1	Features of EU institutional bodies	264
19.1	Codetermination as an interlocking system	297
21.1	The Treuhand legacy	329
33.1	Different approaches to law and economics	468

Contributors

Jürgen G. Backhaus, Krupp Chair in Public Finance and Fiscal Sociology, University of Erfurt, Erfurt, Germany.

Don Bellante, Professor and Chair, Department of Economics, University of South Florida, Tampa, FL, USA.

Margaret Brinig, William G. Hammond Professor of Law, College of Law, University of Iowa, Iowa City, IA, USA.

Christoph F. Buechtemann, Community Resources and Information Service (CRIS), Santa Barbara, CA, USA. Mr Buechtemann died during the preparation of this volume.

Günther Chaloupek, Vienna Chamber of Labour, Vienna, Austria.

Elisabetta Croci Angelini, Professor of Economics, Department of Economics, University of Macerata, Macerata, Italy.

Arno Mong Daastöl, Kolbotn, Norway.

Ben Depoorter, Professor of Law, Center for Advanced Studies in Law and Economics, Ghent University, Belgium, and Fellow, Center for Law, Economics and Public Policy, Yale Law School, Yale University, New Haven, CT, USA.

Wolfgang Drechsler, Professor and Chair of Public Administration and Government, University of Tartu, Tartu, Estonia.

John N. Drobak, John Alexander Madill Professor of Law and Economics, Washington University, St. Louis, MO, USA.

J.L.M. Elders, Professor Emeritus, Amersfoort, The Netherlands.

Francesco Farina, Professor of Economics, Department of Economics, University of Siena, Siena, Italy.

Giampaolo Frezza, LUMSA University, Rome, Italy.

Heinz Grossekettler, Professor of Economics, Department of Public Finance, Westfälische Wilhelms University, Münster, Germany.

Sophie Harnay, Department of Economics, University of Reims Champagne Ardenne, Reims, France.

Ludwig Van den Hauwe, Brussels, Belgium.

Jean-Michel Josselin, Professor of Economics, Department of Economics, University of Rennes, Rennes, France.

Peter Lewisch, Professor of Law, Imadec University, Vienna, and Professor, Department of Criminal Law, Imadec University, Vienna, Austria.

Alain Marciano, Professor of Economics, University of Reims Champagne Ardenne, Reims, France.

Giuseppe Dari Mattiacci, Associate Professor, Amsterdam Centre for Law and Economics, University of Amsterdam, School of Economics, Utrecht University, Utrecht, The Netherlands and Visiting Professor, George Mason University, School of Law, Arlington, VA, USA.

Steven G. Medema, Professor of Economics, Department of Economics, University of Colorado at Denver, Denver, CO, USA.

Nicholas Mercuro, Professor in Residence, Michigan State University, College of Law, East Lansing, MI, USA.

Thomas J. Miceli, Professor of Economics, Department of Economics, University of Connecticut, Storrs, CI, USA.

Christian Müller, Lecturer (Akademischer Rat) Department of Business and Economics, Duisburg-Essen University, Duisburg, Germany.

Antonio Nicita, Assistant Professor, Department of Political Economy, University of Siena, Siena, Italy.

Douglass C. North, Spencer T. Olin Professor in Arts and Sciences, Department of Economics, Washington University, St. Louis, MO, USA.

Ugo Pagano, Professor of Economic Policy, Department of Political Economy, University of Siena, Siena, Italy.

Francesco Parisi, Professor of Law, School of Law, George Mason University, Fairfax, VA, USA.

Heath Pearson, Visiting Assistant Professor, Department of History, University of California, Berkeley, CA, USA.

Helge Peukert, Lecturer, Krupp Chair in Public Finance and Fiscal Sociology, University of Erfurt, Erfurt, Germany.

Donatella Porrini, Associate Professor, Faculty of Economics, 'Antonio de Viti de Marco', University of Lecce, Lecce, Italy.

Manfred Prisching, Professor, Institute of Sociology, University of Graz, Graz, Austria.

Giovanni B. Ramello, Department of Economics, University of Carlo Cattaneo, Castellanza, Italy.

Warren J. Samuels, Professor Emeritus of Economics, Department of Economics, Michigan State University, East Lansing, MI, USA.

A. Allan Schmid, Distinguished Professor Emeritus, Department of Agricultural Economics, Michigan State University, East Lansing, MI, USA.

Peter R. Senn, Professor Emeritus of Economics, Evanston, IL, USA.

Željko Šević, Professor of Economics, The Business School, University of Greenwich, London, UK.

Erich Streissler, Professor Emeritus, Institute of Economics, University of Vienna, Vienna, Austria.

Manfred Tietzel, Professor of Public Finance and Economic Methodology, Department of Business and Economics, Duisburg-Essen University, Duisburg, Germany.

Richard E. Wagner, Harris Professor in Economics, Department of Economics and Center for Study of Public Choice, George Mason University, Fairfax, VA, USA.

Ulrich Walwei, Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, Nuremberg, Germany.

Edwin Woerdman, Associate Professor of Law and Economics, Department of Law and Economics, University of Groningen, The Netherlands.

Leland B. Yeager is former Ludwig von Mises Distinguished Professor Emeritus of Economics at Auburn University. He is an expert on monetary policy and international trade.

Introduction

Jürgen G. Backhaus

> This book, I am told, needs an introduction. I therefore hasten to supply a guide to the reader who may find the volume difficult to read, and who has to be prepared for a journey through the most varied and partly inhospitable terrain, in which the ultimate goal and purpose of every single step can hardly be clear to him at every moment (Steindl, 1952, p. v).

The purpose of the *Companion* is to provide a reference work for the active researcher in law and economics. In so doing, care has been taken to avoid a possible overlap with other works in the field. In particular, the *Companion* does not intend to duplicate the ambitious *New Palgrave*, which aims to balance its pointedly formal focus by emphasizing institutional economics (Newman, 1998). The comprehensive set of chapters in the *Companion*, mainly in the Chicago tradition of law and economics (Posner and Parisi, 1997), allows us to focus on other mainly European aspects of law and economics and the historical sources of law and economics research, which explains its structure (Bouckaert and de Geest, 1999).

The *Companion* has not only been updated and revised for its second edition, but has also been substantially amended. Parts I–VIII cover the main areas of law and economics, including basic issues as well as different sources of the law, while Part IX offers 26 scholarly biographies of the key figures involved. These biographies have been written with a view to encouraging further research into neglected areas in the field which have been taken up at some point but are not part of the current scholarly discussion in law and economics.

Roots

Law and economics has its roots in those natural law philosophies, such as Christian Wolff's (1740), from which they developed as separate disciplines. For Wolff, for instance, applying an economic analytical argument to a legal question was still a standard approach. Only after the disciplines had gone their separate ways would it seem natural for an economic problem to be met with an economic analytical tool, and a legal problem with the proper legal analytical tools. The possibility that a legal problem might be tackled using an economic approach is novel and obviously requires the separation of the two disciplines. However, although genuinely innovative, the practice has been a long-standing one, for example, see authors such as Henri Storch, Wilhelm Roscher, Adolph

Wagner and Gustav von Schmoller, or Rudolf von Jhering, who, with his emphasis on the *purpose* of the law, clearly adopts an economically inspired approach to organizing an entire dogmatic civil legal system.

Still on the European continent, the extensive codifications which took place mainly during the nineteenth century were partly fuelled by economic analytical arguments. Oddly enough, the Englishman Jeremy Bentham was most influential outside his home country, as his explicit legal economic analysis leading towards not only codifications but also specific problem solutions to well-defined policy puzzles form early masterpieces of successful legal economic analysis. One can generalize by saying that continental economics had a strong law and economics undercurrent until the early 1930s (Backhaus, 1987). For instance, the name of the leading journal in economics in the German language area was *Annals of Legislation, Administration and Political Economy*. Hence, legislation and administration were clearly seen as the economic policy areas most likely to be used.

This particular European tradition was transported to America, and here the early institutionalist scholars continued a brand of economics which merged seamlessly with what we now understand as the old law and economics, when attempts to regulate market forces required economic analysis as inputs into administrative and judicial decisions.

Chicago, Yale, Virginia et al.

A totally different picture emerged after the Second World War, partly evolving from these continental roots, but facing a different challenge altogether. Economics had now developed into a science focusing on human decisions under constraints, and it was these constraints that required specific attention, since many of them arguably could be defined as being part of the law. The University of Chicago, with its many emigrant scholars, started to pioneer a new law and economics approach leading to the seminal work of Aaron Director, Ronald Coase, George Stigler, Richard Posner and Frank Easterbrook, to name but a few, which can be characterized as the distinct insertion of an economic analytical skeleton into legal dogmatics, just as the earlier writers had done on the European continent, witness Jhering or Otto von Gierke. However, these writers had to deal with a mass of amorphous case law, not codified law, and this made the task of seeking an organizing theoretical analytical framework a much more urgent one. In this these scholars excelled and, most notably, Richard Posner rendered the entire body of private law, and later all the other relevant bodies of law, including constitutional, administrative, and penal law, into one well-organized system, whose dogmatic structure is clearly borrowed from price theory.

But other schools did not remain on the sidelines. At Yale, a different approach was taken, with a more activist agenda being adopted. Here we

think of Guido Calabresi's classic, *The Cost of Accidents* (1970), which analyses the problem of how a legal system has to formulate policies that minimize the (necessary) cost of accidents in a modern society, when it is well understood that modern technologies will be adopted and cannot be rejected. In the same vein, Calabresi continued with his *Tragic Choices* (Guido and Bobbitt, 1978), while Susan Rose Ackerman (1992) explicitly started to reconsider the progressive agenda from a law and economics point of view.

Very different from this political bent is the Virginia School in Law and Economics in modern America, with the important contributions by Gordon Tullock on basic issues of the law from the law and economics point of view (including public choice considerations) (see, for example, Tullock 1971 and 1980), James Buchanan's constitutional approach to public choice, and the numerous studies that the public choice camp has produced on the impact of the regulatory state on economic activity, including the substantial costs of this regulatory activity; witness the theory of rent seeking pioneered by Tullock.

These different new approaches to the new economic analysis of law have found their publishing outlets in five leading journals in the field. The University of Chicago publishes the *Journal of Law and Economics* and the *Journal of Legal Studies*. Closer to the Yale approach is the *Journal of Law, Economics and Organization*. A more formal approach is taken by the *International Review of Law and Economics* and applied issues, particularly in a European context, are the focus of the *European Journal of Law and Economics*.

References

Ackerman, Susan Rose (1992), *Re-Thinking the Progressive Agenda*, New York: Macmillan.

Backhaus, Jürgen (1987), *Mitbestimmung im Unternehmen: Eine rechtsökonomische Analyse des Verfassungsgerichts Urteils vom 1. März 1979, als Beitrag zur Theorie der wirtschaftlichen Rechtspolitik* (Co-determination in the corporation: a legal–economic analysis of the decision of the German Constitutional Court of 1 March 1979 as an introduction to the theory of legal economic policy), Göttingen: Vandenhoeck & Ruprecht.

Bouckaert, Boudewijn and Gerrit de Geest (eds) (1999), *Encyclopedia of Law and Economics*, I–V, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Calabresi, Guido (1970), *The Cost of Accidents: A Legal and Economic Analysis*, New Haven, CT: Yale University Press.

Calabresi, Guido and Philip Bobbitt (1978), *Tragic Choices: The Conflicts Society Confronts in the Allocation of Tragically Scarce Resources*, New York: Norton.

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

Commons, John R. (1924), *Legal Foundations of Capitalism*, New York: Macmillan.

Easterbrook, Frank H. and Daniel R. Fischel (1991), *The Economic Structure of Corporate Law*, Cambridge, MA: Harvard University Press.

Leoni, Bruno (1961), *Freedom and the Law*, Los Angeles: Nash.

Newman, Peter (1998) (ed.), *The New Palgrave Dictionary of Economics and the Law*, London: Macmillan/New York: Stockton.

Posner, Richard A. (1986), *Economic Analysis of Law* (3rd edn), Boston, MA: Little Brown.

Posner, Richard E. and Francesco Parisi (1997), *Law and Economics*, I–III, Cheltenham, UK and Lyme, USA: Edward Elgar.
Roscher, Wilhelm (1965), *Die Geschichte der Nationalökonomie in Deutschland*, [The History of Political Economy in Germany], New York : Johnson Reprint.
Steindl, Joseph (1952), *Maturity and Stagnation in American Capitalism*, Oxford: Basil Blackwell.
Storch, Henri (1823–1824), *Cours d'économie politique*, [Course in Political Economy] I–V, Paris: Aillaud.
Tullock, Gordon (1971), *The Logic of the Law*, New York: Basic Books.
Tullock, Gordon (1980), *Trials on Trial: The Pure Theory of Legal Procedure*, New York: Columbia University Press.
Von Jhering, Rudolf (1883), *Der Zweck im Recht*, [Law as a Means to an End], Leipzig: Breitkopf und Härtel.
Wolff, Christian (1740), 'Von einer Erwegung der Staatsgeschäfte' [Contemplations About Public Affairs], *Gesammlete kleine philosophische Schrifften*, Halle: Renger.

PART I

BASICS OF THE LAW AND ECONOMICS APPROACH

1 Coase theorem and transaction cost economics in the law

Francesco Parisi[1]

This chapter discusses the pervasive methodological implications of Ronald H. Coase's contribution to economics and the law. Coase's reconceptualization of the firm as an institutional device to minimize transaction costs has triggered an entire field of research. The traditional view of production, where labour and capital are the primary inputs, is refocused and replaced by the important role of governance structures in firms. Similarly, Coase's assertion that an initial assignment of property rights is often irrelevant to overall welfare has occasioned one of the most intense and fascinating debates in the history of legal and economic thought. In the following pages, I shall examine the state of legal and economic scholarship in the wake of Coase's well-known methodological breakthroughs.

Transaction costs and Coase's theory of institutions

In his classic 1937 paper, 'The nature of the firm', Coase developed an economic theory of the firm which laid the foundation for understanding a wide range of institutional and organizational structures.[2] Coase's pathbreaking insight was that the comparative costs of organizing transactions within firms, rather than through markets, are the main factors that explain the existence and evolution of firms.[3] Likewise, the size and scope of firms is determined by the relative costs of accessing the market versus governing an organization at the various levels of production.[4]

A wide range of empirical and theoretical issues have arisen as a result of Coase's contribution. The most significant extension of his 1937 work has been the application of the transaction cost hypothesis to other forms of institutional structures. These extensions have become a central part of the transaction cost economic tradition. Indeed, several scholars have exploited the explanatory power of the transaction cost hypothesis in order to enhance the understanding of economic organization generally.

The puzzle of the firm: prices versus organizations

Coase developed his theory of the firm contrary to the prevailing economic theory. Economists had demonstrated the informational and functional superiority of the price mechanism over alternative allocative systems based on

centralized planning. Coase observed that such a hypothesis did not fit at all within the firm. His view was that the allocation of scarce resources among competing uses within a firm rests not on a price mechanism, but rather on the planning of the entrepreneur who makes allocative decisions without the aid of prices.[5] Market transactions are replaced with the controlled choices of the firm's manager. Coase generalized from this point that the distinguishing feature of firms is, indeed, the suppression of the price mechanism.

Coase's theory of the firm thus unveils an important puzzle, that is, why a firm emerges in a specialized exchange economy. Coase considered several possible explanations for the emergence of firms: first, the preference of workers to be subject to a command structure; second, the desire of entrepreneurs to have exclusive control over the planning of production; third, and finally, the cost of using a price mechanism.[6]

Coase's analysis established the importance of the third explanation: the price mechanism is costly to use.[7] Coase provided examples of the implicit costs, such as the difficulty of determining the relevant values of joint inputs and outputs, and the preferability of long-term contracts over spot-market prices for risk-averse individuals.[8] Additionally, market transactions are often treated differently from the internal decisions of the firm for both tax and legal purposes. The legal system may, in fact, create additional costs for the use of the price mechanism in the marketplace. Thus, in the internal setting, the firm becomes an island of exemption from those external costs.[9]

The subsequent transaction cost literature has explored the relative advantages of alternative institutional solutions under various real-world settings.

Transaction costs and the economics of institutions

Transaction cost economics views the firm and the market as alternative means of contracting. Building upon Coase's analysis, Williamson (1985) identified the limitations of the neoclassical analysis of models of perfect competition.[10] He reached beyond the assumptions of the neoclassical analysis to consider the roles played by other crucial variables. Williamson and other exponents of the new institutional economics explained the emergence and functioning of economic and legal institutions, not only as a production function, but as an intricate mode of contracting, and as a governance framework alternative to the market.

The allocation of economic activity as between firms and markets is taken as a datum under the neoclassical approach; firms are characterized as production functions;[11] markets serve as signalling devices; contracting is accomplished through an auctioneer; and disputes are disregarded because of the presumed efficacy of court adjudication.[12] Williamson criticized the neoclassical economic approach to the market and the firm for relying on such simplistic assumptions that too often limit the explanatory power of their models.

In the classical model of economics, the market is a frictionless institution characterized by perfect competition, ease of entry and exit, product homogeneity, unbounded rationality and perfect information. Self-interest and opportunism are not ignored in the classical model, but are only accounted for in the bargaining stage of contract, not in the execution stage.

The new institutional economics makes three additional assumptions regarding contracts. First, contracting is characterized by actors with bounded rationality, second, that those contracting also act opportunistically in the execution stage, and third, that the dimension of asset specificity must be added to the model assumptions. When all three of these elements are present, the contracting outcome calls for a governance solution. Thus, the new institutional economics attempts to explain how institutions with a governance structure emerge as transaction cost-minimizing devices in a world characterized by *ex post* opportunism and *ex ante* cognitive imperfections. More specifically, the new institutionalists criticize the alternative perspectives for their unconditional reliance on unrealistic assumptions: planning assumes perfect cognitive competence;[13] contract as promise assumes absence of *ex post* opportunism in the execution stage of the contract;[14] the perfect competition model ignores the crucial role played by asset specificity in the execution stage of the contract.[15] Williamson points out that when all three of these conditions – bounded rationality, opportunism and asset specificity – are present, the three classical contracting processes fail. In response to these shortcomings, the new institutional economics governance approach is interested in the governance structure and non-standard forms of contracting that emerge in the presence of bounded rationality,[16] opportunism[17] and asset specificity.[18]

Coase's legacy in the new institutional economics

As indicated above, the literature of the new institutional economics looks at the firm not only as a production function, but also as a governance function. This school of thought recognizes the intellectual legacy of Ronald Coase, tracing the roots of transaction cost or institutional economics to the writings of John Commons and Coase's 1937 article, 'The nature of the firm'. Coase's idea of the firm as an institutional device to minimize transaction costs is applied to other institutional settings where exchange market transactions are eliminated. The primary role of economic institutions is to decrease transaction costs associated with coordinating market activity. Scholars of the new institutional school generally credit Commons with recognizing that the transaction should be regarded as the basic unit of analysis. Commons recognized that economic organization is not merely a response to technological features, economies of scale or economies of scope, but often has the purpose of harmonizing relationships. The new institutionalists take this analysis one

step further, and posit that the imperative of profit maximization should be replaced with the organizational imperative to organize transaction costs so as to economize on bounded rationality while simultaneously limiting the hazards of opportunism. Transaction cost economies are realized by assigning transactions to governance structures and comparing institutional alternatives. With its extensions into the efficiency of institutional alternatives, this trend of research can thus be characterized as pursuing two general themes: the study of incentives generated by alternative legal and economic institutions, and the transaction cost optimization as a main determinant of the institutional choices. The incentive approach is predominantly *ex ante*, hence its utility in property rights and agency theories. The basic idea is that if rules are formulated so as to properly align incentives, fewer market distortions will result. Without these distortions, outcomes will more closely approximate the ideal outcome of global optimization. The approach followed by the transaction cost economists and by several new institutionalists places great emphasis on the *ex post* perspective, as contrasted with the traditional perspective of neoclassical economics. The basic unit of analysis is the transaction and the basic idea is to determine which governance structure is best suited to which type of transaction.

This approach is key to understanding the intellectual emphasis of the new institutionalists and their distinctive view of the firm and other governing structures. In this respect, Coase's legacy is well served by the widespread recognition that the neoclassical view of labour and capital as being the primary components of production had to give way to the central role of governance structures within the firm.

The genesis of the problem of social cost

The study of property rights and institutions has vividly engaged the attention of economists, philosophers and lawyers alike. Private property is often explained as the unavoidable byproduct of scarcity in a world where common-pool losses outweigh the sum of contracting costs and enforcement of exclusive property rights. At the turn of the twentieth century, the underlying assumption in the economic literature was that private property emerged out of a spontaneous evolutionary process because of the desirable features of private property regimes in the creation of incentives for constrained optimization.

This understanding of the relationship between scarcity and the emergence of legal entitlements characterized mainstream property rights theory when Coase entered the academic world as an undergraduate student in economics.[19]

Property rights theory at the London School of Economics

In the early 1930s, while Coase was conducting his undergraduate studies in commerce at the London School of Economics (LSE), one of Coase's teach-

ers, Sir Arnold Plant, was re-examining the theme of property rights from a novel perspective.[20] In Plant's view, the traditional justification for private property, scarcity, was incapable of serving as the sole intellectual foundation for this institution. In those years, Plant completed two papers on issues of intellectual property and copyright laws, showing that issues of incentives, rather than scarcity, were at the core of the property rights problem.[21]

There is, indeed, a striking correspondence of methodology and thematic between the later works of Coase and those of his undergraduate teacher. Both the focus on the incentive structure of legal rules, and the analysis of the effect of alternative laws on the final allocation of human and physical resources, reveal a remarkable affinity of technique, and the explicit use of legal rules as an object of economic research makes the comparison even more telling.[22]

It is indeed this fortunate combination of methodology and subject matter that would prove so valuable in Coase's research. Coase acknowledges the importance of his encounter with Plant at LSE, and pinpoints that 'great stroke of luck' as the origin of his interest in property rights theory.[23] For Coase, the encounter with Plant was a true revelation. Plant's repeated teaching that '[t]he normal economic system works itself',[24] and his belief that prices in a competitive market lead resources to their highest valuing uses, provided Coase with a powerful insight on the dynamic of the economic system:

> I was then 21 years of age, and the sun never ceased to shine. I could never have imagined that these ideas would become some 60 years later a major justification for the award of a Nobel Prize. And it is a strange experience to be praised in my eighties for work I did in my twenties.[25]

The experience of the following years in conjunction with the London School of Economics laid the methodological foundations of what would later become Coase's theorem on the problem of social costs.

The University of Virginia years: the birth of an ingenious idea

All the ingredients of Coase's revolutionary analysis on the debated theme of social cost had been profiled during his LSE years.[26] But it is not until the late 1950s that Coase verbalizes such a simple – and yet ingenious – idea. He had first expounded the core of his later theorem in an article published in 1959 – a fact not always remembered in the bibliographic citations.[27] In those pages, one grasps what would later become the underlying theme of Coase's celebrated argument:

> Whether a newly discovered cave belongs to the man who discovered it, the man on whose land the entrance to the cave is located, or the man who owns the

> surface under which the cave is situated is no doubt dependent on the law of property. But the law merely determines the person with whom it is necessary to make a contract to obtain the use of the cave. Whether the cave is used for storing bank records, as a natural gas reservoir, or for growing mushrooms depends, not on the law of property, but on whether the bank, the natural gas corporation, or the mushroom concern will pay the most in order to be able to use the cave.[28]

The discussion of the rationale of property rights under Coase's highest-bidder framework obviously contained an attack on the Pigouvian approach to the problem.[29] The point was rather self-evident to Coase, but not so for some of the Chicago economists. George Stigler was among Coase's early critics:

> Ronald Coase criticized Pigou's theory rather casually, in the course of a masterly analysis of the regulatory philosophy underlying the Federal Communication Commission's work. Chicago economists could not understand how so fine an economist as Coase would make so obvious a mistake. Since he persisted, we invited Coase (he was then at the University of Virginia) to come and give a talk on it. Some twenty economists from Chicago and Ronald Coase assembled one evening at the home of Aaron Director. … In the course of two hours of argument the vote went from twenty against and one for Coase to twenty-one for Coase. What an exhilarating event![30]

According to Coase, the objections that were raised to his Federal Communication Commission (FCC) paper were the basis of his later 1960 article on the problem of social costs.[31] In the course of his meeting with the Chicago economists, Coase had occasion to refine some of the arguments that he had outlined in his earlier work, arguments that he was later asked to put together in the form of an article for the *Journal of Law and Economics*.[32] He entitled this paper 'The problem of social cost'.

The Coase theorem

Coase's 1960 article was soon to be recognized as a milestone in legal and economic literature – a milestone later characterized as the Coase theorem. In the course of his austere discussion, Coase does not reveal any sign of anticipated realization of the revolutionary power of his insight. Indeed, he insists that he never intended to convey his thoughts in the precise and analytical form of a theorem.[33]

A few years after the publication of 'The problem of social cost', a sizeable number of commentaries and theoretical elaborations were developed on Coase's newly presented theme.[34] The unpretentious style of Coase's article had thus been crowned by a notoriety rarely attained by legal writings of any sort.[35] Part of the uproar is explained by the fact that the article challenged an established principle of public finance.[36] Before 'The problem of social cost', very little attention had been given to the possibility that the problem of

externalities could be resolved through free market exchanges.[37] In this way, the thesis advanced by Coase resulted in a rather revolutionary statement, one at the core of a central theme of economic science.[38]

Coase boldly attacks the conclusions reached by the Pigouvian tradition:

> It is strange that a doctrine as faulty as that developed by Pigou should have been so influential, although part of its success has probably been due to the lack of clarity in the exposition. Not being clear, it was never clearly wrong. … I propose to show the inadequacy of this Pigouvian tradition by demonstrating that both the analysis and the policy conclusions which it supports are incorrect.[39]

Coase contrasts the Pigouvian approach by demonstrating that, in the absence of transaction costs, generators of externalities and victims will negotiate to an efficient allocation of resources, independent of the initial assignment of rights among them.[40] In confuting the conclusions of the Pigouvian tradition, Coase gives life to a model with much broader potential, a revolutionary new perspective for the evaluation of an unlimited number of legal and social issues.

Stigler, in 1966, was the first scholar to restate Coase's model in the form of a theorem: 'under perfect competition private and social costs will be equal'.[41] In 1967, Alchian Demsetz defined the theorem in the following terms:

> There are two striking implications of this process that are true in a world of zero transaction costs. The output mix that results when the exchange of property rights is allowed is efficient and the mix is independent of who is assigned ownership (except that different wealth distributions may result in different demands).[42]

Soon thereafter, Guido Calabresi stated the same principle more descriptively: 'Thus, if one assumes rationality, no transaction costs, and no legal impediments to bargaining, all misallocations of resources would be fully cured in the market by bargains'.[43]

The implicit premise of Coase's analysis draws upon a fundamental postulate of microeconomic theory: the free exchange of goods in the market moves goods towards their optimal allocation, such that, when every possibility of beneficial exchange is satisfied, resources will reach their optimal allocation according to the criterion of Pareto efficiency.[44]

The law creates many subjective juridical positions that are also susceptible to exchange and transfer. Coase, applying by analogy[45] the proposition of the free exchange of goods in the market, maintains that the transferability of rights in a free economy leads towards their best use and to a Pareto-efficient final allocation.[46] The voluntary transfer of individual rights in the marketplace, thus, will cure a non-optimal allocation of legal entitlements.[47]

The Coasean methodological revolution

Coase's article, discussing widely cherished themes in the legal and economic traditions, constitutes, according to many commentators, the first example of an economic analysis of law in North American literature. The novelty of his approach inspired an entire generation of scholars – pioneers in this new branch of applied economics. In 1991, 30 years after the publication of 'The problem of social cost', Coase received the Alfred Nobel Memorial Prize in Economic Science. Through the prestige of this award, recognition came at last to honour the tradition of economic analysis of law that Coase so authoritatively represents. Only a few months prior to the award, Coase was recognized, together with Calabresi, Henry G. Manne and Richard A. Posner, as a founding father of law and economics.[48] This late recognition follows many years of challenging debate. Many of the writings that developed around 'The problem of social cost' tested the premises of Coase's model, seeking to undermine its operative conditions. The corollary literature, which was almost unanimous in acknowledging the theoretical soundness of Coase's approach, often stressed the lack of practical reach of his analysis.

Acknowledging the risk of an inaccurate first impression, it is possible to observe that the various criticisms pertained to three fundamental points, relating to the operative possibilities and practical effects of Coase's model. One group of critics observed that the Coase theorem disregarded the inter-industrial long-term effects of the system.[49] These critics argued that Coase utilized tools of static analysis, disregarding the possible disequilibria which may occur subsequent to the negotiation, and that the conclusions reached by Coase needed to be tested in light of the dynamic changes in the initial equilibrium. A second group of critics concentrated, instead, on the distributional effects of the model.[50] According to their criticism, to affirm that, in the absence of transaction costs, the final allocation of resources will be efficient in no way implied – much less guaranteed – the absence of transfers of wealth induced by the changed legal rule. Further, these critics observe that, even disregarding the distributional effects of the rule, a different assignment of the right could in some cases create the conditions for strategic behaviour in negotiation capable of disturbing the efficiency of the final allocation.[51] A third group of authors focused on the scarce realism of the no-transaction cost assumption.[52] According to this criticism, the true Achilles' heel of Coase's analysis was in the unrealistic assumption of absence of costs in the process of negotiation and transfer of the right. Many commentators observed that in order to obtain the efficient operation of Coase's model, it was not enough to eliminate legal impediments to the free transferability of individual rights; it was necessary as well to operate in an imaginary world with no costs involved in the negotiation or transfer of the right. These authors observed that the idea of a transaction without cost is a logical fiction rather

than a real possibility, and that by unveiling such a fiction, the theorem remains a mere tautology.

The following sections will discuss in greater detail the significance of these criticisms, and the impact that the emerging debate has had on the traditional approach to legal interpretation.

The self-curing failures of the Coasean bargaining

In a situation in which there is a non-optimal allocation of rights between two individuals, the Coase theorem predicts that the interested parties will contract with each other and that they will reallocate their respective rights so as to maximize their combined welfare. Coase postulates that the efficiency of the result is independent of the initial allocation of rights. According to some critics of Coase's model, however, a change in the allocation of rights is the potential origin of disequilibria in the system. The thrust of this criticism follows.

The dynamic effects of alternative liability rules

Calabresi[53] and Stanislaw Wellisz[54] are notable among the scholars who criticized Coase's model for disregarding the inter-industrial long-term effects of the Coasean bargaining. According to these authors, Coase's scheme does not take into account the dynamic effects of alternative liability rules among the various parties, and consequently, it ignores the long-term effects of the rules on different industries. In Coase's scenario, if the right has been assigned to the ranchers, the farmer will have to pay local ranchers until they all relinquish their right of pasture. The entire cost will, thus, burden the farming industry. Farmers will either have to bear the burden of the injury caused by the livestock or agree to pay the price demanded by the ranchers, whichever is less, assuming costless negotiation. Under this liability rule, the cost of ranching will not reflect the cost imposed on the farmers. The transfer of rights and liability from one group to another will, therefore, result in a shift in the relative wealth and costs associated with the two industries.

The criticism claims that, in the long run, every shift of wealth will lead to an inter-industrial disequilibrium. Even in the absence of transaction costs, a different assignment of rights can alter the equilibrium between different industries, with consequential effects on the cost and quantity of their relative products. In our example, if the farmers must suffer the losses caused by the herd during grazing – or pay the ranchers to avoid the damage – the unit cost of the farming product will inevitably be higher than would have resulted from a different allocation of liability. The entire farming industry will have higher costs of production and will, therefore, suffer a decrease in income. Consequently, some of the resources invested in that industry are likely to be

channelled towards more lucrative investments, with potential for resulting disequilibria.

Formulated in this way, the criticism appears to be on the mark. The efficiency of the Coase theorem is demonstrated only through a static analysis. If dynamic adjustments are taken into consideration, the structure of the model reveals its incapacity to consider the long-run inter-industrial effects of different initial allocations. However, in 1968, Calabresi, one of the initial proponents of this criticism, reconsidered his analysis regarding the long-term effects of the Coase theorem.[55] While elaborating on the conclusions reached in two previous works,[56] he noted that, in the presence of determined conditions, the conclusions of Coase remain as true in the long run as in the short term:

> Various writers – including me – accepted that conclusion for the short run, but had doubts about its validity in the long run situation. The argument was that even if transactions brought about the same short run allocation, liability rules would affect the relative wealth of the two joint cost causing activities, and in the long run this would affect the relative number of firms and hence the relative output of the activities. Further thought has convinced me that if one assumes no transaction costs … and if one assumes, as one must, rationality and no legal impediments to bargaining, Coase's analysis must hold for the long run as well as the short run.[57]

In this way, Calabresi carried on the logic of his earlier argument to reach opposite conclusions. The dynamic adjustments of the equilibrium that he had identified as the cause of the inter-industrial misallocations of resources, were, in reality, self-curing. The same dynamic strength of the market was capable of resolving the inter-industrial disequilibria denounced by Calabresi in his 1965 article.

Calabresi's later analysis re-established the authority of the Coase theorem, at least on this point. It became clear that Coase had not ignored the long-term effects of his model. Perhaps not explicitly, but he had considered them to their logical extreme. Calabresi proceeds: 'The reason is simply that (on the given assumptions) the same type of transactions which cured the short run misallocation would also occur to cure the long run ones. … This process would continue until no bargain could improve the allocation of resources'.[58]

Harold Demsetz on the long-term effects

In 1972, Harold Demsetz entered into this debate, demonstrating with a more systematic analysis that the conclusions reached by Coase are not corroded by the long-term effects of a change in the assignment of property rights.[59] Demsetz's reasoning finds its basis in the principle according to which the process of allocation of scarce resources among alternative uses is analogous

to the process of constrained optimization of the single owner of two conflicting activities.

In order to better understand Demsetz's reasoning, imagine a situation in which the two activities, farming and ranching, 'belong' to the same individual. This person has every interest in making the optimal allocational choice in the use of his/her limited resources between the two activities, and will tend to maximize the sum of his/her benefits at the net of the costs. Such a choice would lead to the optimal use of his/her resources in both the short and long terms, regardless of the equilibrium reached in the two industries or activities.

Approached in this manner, the true problem seems to remain that of pervasive scarcity of resources, not of assignment of rights.[60] In our example, the single owner of the two activities will not be interested in establishing whether the herd is creating a nuisance to the crops or whether the crops are becoming an obstacle to the ranching activity. The identification of the internal boundaries between different rights is entirely irrelevant for the integrated owner of multiple activities, whose only interest is that of attaining an optimal choice in the employment of limited resources. The problem for the single agent, as for society as a whole, is one of constrained optimization in a world characterized by pervasive scarcity. The theoretical concern for possible disequilibria between various activities, foreign to the preoccupations of the single owner, must also remain foreign to the debate on the Coase theorem.[61] The competitive allocation of limited resources between different activities is in no way different from the internal dilemma of a single individual who must make an optimal choice between alternative uses of his/her assets.

These conclusions, however, do not appear to be fully shared by Donald Regan, who observes that the self-curing dynamic of Coase's theorem is destined to remain a phenomenon foreign to the reality of the market.[62] Everything is in theory corrected through the internal mechanism of a market with no transaction costs, even the inefficiency generated by the monopolist.[63] Individual consumers will be willing to pay the monopolist to increase the production of goods to the desired level. In the absence of transaction costs, the negotiations will proceed until the optimal equilibrium of a perfectly competitive market is achieved. But this solution is not without its shortcomings. According to Regan's view – a view not shared by Coase – the free exchange of rights in the market produces irreversible transfers of wealth between the parties. According to this perspective, the market solution outlined by Coase and Stigler – and to some extent endorsed by Calabresi – while resolving a problem on one side, immediately creates a problem on the other. Regan and other commentators direct their attention to this point of collateral effect.[64] An account of their reasoning and a tentative assessment of their findings follow in the next section.

From property rights to individual incentives

A number of issues related to the distributional effects induced by a Coasean negotiation have been raised in the economic and legal literature. Economists tend to appraise the issue in terms of final distribution of resources derived from initially different allocations. The question is whether the optimality of the final allocation predicted by Coase guarantees, or implies as such, identical final allocations.

The debate on the uniformity of the final allocations of resources serves as a logical premise to the issue of distributional effects. Microeconomic theory teaches us that different points along a contract (or conflict) curve correspond to different distributions of goods among the various players. In other words, notwithstanding the Pareto optimality of every agreement that falls along a contract curve, any change in the initial endowments necessarily generates a different final distribution of resources. Following this logic, Regan and Nutter elucidate the strict tie between uniformity of allocations and distributional effects.[65] To affirm that the efficiency of the final allocation does not imply identical allocative outcomes at the equilibrium point, means to admit that the change of legal rules, although corrected by the Coasean negotiation on the level of efficiency, will always cause shifts of wealth between the various parties.

The observations that follow endeavour to account for the debate on the allocational uniformity and the distributional effects of the Coase theorem.

Property rights and social costs

The debate on this point was also initiated by Calabresi[66] and Wellisz,[67] who considered the issue of allocational uniformity as intrinsically related to that of distributional effects.[68] In order to evaluate the significance of their analyses, let us consider again Coase's scenario with a farm and a ranch coexisting in the same environment. In such a setting, a change in the allocation of rights and liabilities affects the relative values of the two activities. If we assign the right to the ranchers (that is, exclude their liability for the loss suffered by the farmers), we force the various farmers to bribe the ranchers into reducing the number of animals in their herds. Conversely, by assigning the right to the farmers, we force the ranchers to compensate the farmers for the damage to their crops. Because of these side payments, the different assignment of rights makes its mark on the profitability of the two activities, and on the value of the resources irreversibly invested in those enterprises. According to this argument, the transfer of primary and residual liability from one subject to another occasions a transfer of wealth.[69]

In his 1988 notes on the problem of social cost, Coase argues against the soundness of this logic: 'I consider this argument to be wrong, since a change in the liability rule will not lead to any alteration in the distribution of

wealth'.[70] Coase's argument is that, if the right is assigned to the farmer, then the cost of the rancher's liability will be discounted from the price necessary to acquire or rent the ranching business. The land destined for the activity of the rancher, subsequent to the shift of liability will be less valuable. Analogously, the farming land, protected by the liability rule, will yield a greater revenue and will consequently demand a higher price on the market. The change in the relative costs of the two businesses will, thus, offset the patrimonial effects of the modified legal rule.

With these observations, Coase responds to the various criticisms on the distributional effects of his model by affirming that, as soon as the assignment of rights between the two industries is known, it will be reflected in the relative prices of their products. The wealth of prospective farmers, ranchers and land-owners will remain unaltered since the changes in the prices of their entitlements promptly will balance the momentary disequilibrium caused by the changed system of rights.[71]

Coase's analysis, however, seems to presuppose a static system of legal rules in which, regardless of what may be the initial allocation of rights, a final equilibrium will be reached on a system of prices that fully offsets the distributional effects of the legal rule. In his view, once the legal rule is known, the adjustments in the prices of the affected factors of production will prevent any alteration in the respective supply and demand curves. But the previous analysis is questionable if, eliminating the assumption of staticity, one takes into consideration the possibility of sudden and recurrent changes in the assignment of property rights.[72] The system of prices will not be capable of offsetting the losses suffered by property rights owners as a consequence of an unexpected change in the legal rule. The preserved optimality in the set of legal incentives is obtained in total disregard of vested rights and property interests.

Coase does not overlook the possibility of a similar objection, and tries to reconcile the conflict through contractual devices. He believes that the distributional effects can be avoided even in the case of dynamic changes in the legal system through a different mechanism, which remains faithful to the nature of his model. According to Coase, in fact, the parties can agree to tie the price paid for the acquisition of any given property right to possible changes in the law.[73] By means of such contractual provisions, the parties would be able to obtain an effective shield against involuntary transfers of wealth due to exogenous changes in the assignment of rights and liability.[74]

Allocational effects and the problem of extortion

According to the Coase theorem, in the absence of transaction costs, the voluntary exchange of property rights would lead to an efficient allocation of resources between alternative uses. According to Calabresi[75] and Wellisz,[76]

however, the use of strategic behaviour in the process of contract formation risks altering such a result.[77] Elaborating on this variation on the general theme of distributional effects, Callabresi and Wellisz observe that the change in the rule of law creates the conditions for possible extortion on the part of the rights holders against the other individuals who are bound by the rule. The argument is that individuals are likely to threaten the use of their own rights in a measure which exceeds the optimal level, in order to maximize the gain from the release of their own legal entitlements.[78] In our example, if the right is assigned to the ranchers, they will be induced to threaten to increase the size of the cattle herds in order to strengthen their bargaining position towards the farmers.[79]

In order to clarify the point, consider a situation in which the optimal size for the rancher's herd is 1000 head. In such a scenario, imagine that, in order to reduce the damage to his/her own crops, the farmer would be willing to compensate the rancher for a reduction of his/her herd to 800 head. According to Coase, this agreement generates an optimal allocation of resources between the two activities. The criticism claims that, by introducing the possibility of strategic behaviour in the negotiation, the result may differ from such an ideal equilibrium. If the rancher threatens – for strategic reasons – to increase the size of his/her herd to 1500 head, the final agreement is likely to diverge substantially from the efficient allocation of resources boasted by Coase. The rancher will, in fact, seek to maximize the profit from the conceded reduction on the first 500 head (which, however, would have constituted an inefficient oversize for his/her firm), and the agreement will likely be reached on different terms from those predicted by the theorem. As a consequence of such strategic bargaining, there would still be too many cattle and too much damage to the crops.

In his 1988 notes on the problem of social cost, Coase did not elaborate on the theme of extortion. His silence on this point perhaps implies a tacit reference to the work of Demsetz, which in 1972 had supplied a convincing answer to this criticism.[80] According to Demsetz, the possibility of strategic behaviour in the negotiations does not alter the efficiency in the final allocation of resources between the two activities.[81] Despite possible uses of strategic bargaining, the number of cattle in our example will always be reduced to the point at which the sum of the values of the two activities is maximized. The optimal allocation will obtain regardless of the internal distribution of the contractual surplus between the parties. If the extortion is not capable of altering the efficiency of the final allocation of the rights reached through Coasean bargaining, the problem is, thus, confined within terms of relative advantage in the apportioning of surplus between the two activities.[82]

Strategic stalls in Coasean bargaining

The residual problem of the allocative effects, often used to cast doubt on Coase's model, merits one further word of clarification.[83] The credibility of the threat made in the course of strategic bargaining finds its limits in the market structure in which the Coasean negotiation takes place. A rancher who threatens to raise the number of his/her cattle beyond the maximum capacity of his/her industrial structure, or to a size that exceeds the absorption of the beef market, for example, would make use of a non-credible threat, one incapable of playing any role in the negotiations.[84] In general, the competitive structure of the market eliminates much of the advantage that can be obtained through strategic behaviour in the negotiation process. Inasmuch as the market of resources is competitive (in our example, as long as there are alternative locations for the farming or ranching activity), strategic bargaining is not capable of bringing about any abnormal return.[85] If the farmer demands a level of compensation that exceeds the market price for that right (in addition to the cost necessary to move the herd to another locality), the rancher will opt for more economical alternatives, relocating elsewhere.

Thus the non-competitive structure of the market and the credibility of the threat become the only situations which seem to justify the concerns for the use of contractual strategies in Coase's model. Beyond these marginal hypotheses, the existence of a competitive market will exclude the possibility of any contractual mark-up that goes beyond the normal returns of a profit-maximizing firm. The criticism, however, appears to be justified when it argues that, in some marginal situations, the curing role of the free exchange may still be impeded. For example, consider reversing the assignment of property rights between the rancher and the farmer. In such a situation, the farmer is likely not to have an equally large number of alternatives. The transfer of a farm from one place to another is costly, and farming unavoidably requires the undertaking of location-specific investments. Since some capital investment is irreversibly locked into that specific location, the farmer has less opportunity to relocate than the rancher.[86] The rancher, consequently, finds him-/herself in a position of local monopoly in the sale of his/her property right. Demsetz considers the monopoly that affects this feature of the Coasean exchange as identical to the standard monopoly of microeconomic analysis:

> The appropriate economic label for this problem is nothing more nor less than monopoly. It takes on the cast of such legal classifications as extortion only because the context seems to be one where the monopoly return is received by threatening to produce something that is not wanted – excessively large herds. The conventional monopoly problem involves a reduction or a threat to reduce the output of a desired good. In the unconventional monopoly problem presented here there is a threat to increase herd size beyond desirable levels. But this difference is

> superficial. The conventional monopoly problem can be viewed as one in which the monopolist produces more scarcity than is desired, and the unconventional monopoly problem discussed here can be considered one in which the monopolist threatens to produce too small a reduction in crop damage. Any additional sum that the rancher succeeds in transferring to himself from the farmer is correctly identified as a monopoly return.[87]

According to Demsetz, the concerns for possible monopolistic structures in the market of rights considered by Coase must not, however, be used, to raise again the already resolved problem of the initial allocation of rights:

> The temptation to resolve this monopoly problem merely by reversing the rule of liability must be resisted. Should the liability rule be reversed and the owner of ranchland now be held liable for damage done by his cattle to surrounding crops, the specific monopoly problem that we have been discussing would be resolved. But if the farmer enjoys a local monopoly such that the rancher has nowhere else to locate, the shoe will now be on the other foot. The farmer can threaten to increase the number of bushels of corn planted, and hence the damage for which the rancher will be liable, unless the rancher pays the farmer a sum greater than would be required under competitive conditions. The potential for monopoly and the wealth redistribution implied by monopoly is present in principle whether or not the owner of ranchland is held liable for damages. Both the symmetry of the problem and its disappearance under competitive conditions refute the allegation that Coase's analysis implicitly endorses the use of resources in undesirable activities.[88]

Having freed the discussion from concerns on the potential use of contractual strategies in a Coasean bargaining situation, we can now move to the examination of the controversial assumption of no transaction cost contracting, by many identified as the true weakness of Coase's model.

Transaction costs and market failures

The very basis of Coase's 1937 article on the theory of the firm – that is, transaction costs in the functioning of the price mechanism – becomes central to Coase's 1960 seminal contribution on the problem of social cost. Ironically, what was necessary to support Coase's 1937 hypothesis is later assumed away in his 1960 paper. In a world without transaction costs, firms would not exist. In a world with transaction costs, the Coase theorem would not hold. Much of the debate surrounding Coase's work dealt with the all-inclusiveness of the category of transaction costs, which risked transforming Coase's 1960 assertion into an empty tautology.[89]

The notion of transaction costs has had a peculiar development in the history of economic thought. Almost every term adopted by economic science has in time assumed a precise, mathematically definable, content. The notion of transaction costs has never been defined in an equally rigorous fashion.[90]

Much of the secondary literature on the Coase theorem has directly or indirectly dealt with the content of this concept. In its narrow sense, the term 'transaction costs' contemplates material expenses and the opportunity cost of the time and energies necessary to reach an agreement on the transfer of a right.[91] To these factors, one should add the various costs necessary for the preparation, strategic implementation and execution of the agreement, including information costs, and all the costs necessary for an effective monitoring of the other party's performance.[92]

The normative Coase theorem

If the sum of the various transaction costs exceeds the net benefit of the contract, no exchange will take place in the market. For a right to be exchanged it is necessary that transaction costs be less than the difference between the demand and supply prices. If this condition is not met, then Coasean bargaining will not be carried out, and rights will remain in a non-optimal allocation.

In the face of a similar clarification, one must question the relevance of Coase's analysis when the assumption of no transaction costs is relaxed.[93] According to Coase's prediction, without transaction costs, the final allocation of scarce resources would coincide with the use that an individual who is the single owner of different activities would make of his/her endowments. Moving into a more realistic environment with positive transaction costs, however, an exchange will be pursued only to the point at which its marginal benefit equals the marginal cost of the transaction.

In this phase of the analysis, the positive transaction costs of Coase's model play a role analogous to transportation costs in international trade or more generally, to the contracting costs in the economics of exchange.[94] This conclusion is rather obvious and consonant with criteria of economic rationality, but, as Demsetz notes, the question cannot be reduced merely to this observation.[95] It is necessary, in fact, to keep in mind that the positive Coase theorem indicates the market as a general cure for inefficient allocations of property rights. To recognize that the reallocation may not take place in the presence of positive transaction costs means to concede that the market solution postulated by Coase may fall short of rectifying the inefficiency in the case at hand. This would yield to other remedies of a public nature, addressing the problem through legislative, judicial or governmental intervention, models of taxation, or other structural corrections of the system.

The effect of positive transaction costs on the Coase theorem has been extensively examined by the secondary literature. The following is a classic illustration.[96] The smoke of a factory soils laundry which is line drying on five neighbouring properties. The losses amount to $150 for each neighbour, for a total of $750. The damage could be eliminated through the installation

of a purifying filter on the industrial smoke stack or through the acquisition of electric dryers on the part of each one of the neighbouring owners. The cost of the filter would amount to $300, while the dryers would impose a cost of $100 per household, for a total of $500. The first solution is obviously more efficient, since the acquisition of five dryers would require an expenditure superior to that of the single filter. The Coase theorem predicts that in the absence of transaction costs, the efficient solution will be chosen independently of the initial assignment of property rights. Even assuming an initial allocation of polluting right to the industry (that is, fully legalizing industrial emissions), the landowners would jointly offer to buy the industrial filter at their expense. Sharing the cost of the filter in equal parts, each owner would face a cost of only $60, with a relative saving of $40 compared to the otherwise necessary acquisition of a personal dryer.

Relaxing the initial assumption of no transaction costs, the initial allocation of property rights is no longer immaterial.[97] Imagine that each owner has to face a cost of $120 in order to negotiate the contract with his/her neighbours and with the owner of the industrial plant. If the right is assigned to the industry, each landowner will have to choose whether to bear the loss of his/her soiled laundry for $150, to acquire the electric dryer for $100, or, finally, to undertake the negotiation process for a total pro-quota cost of $180. Considering these alternatives, each rational landowner will choose to acquire his/her own dryer, generating a socially non-optimal outcome. By relaxing the no transaction cost assumption, thus, the choice of legal regimes appears capable of affecting the final equilibrium. In this particular case, the assignment of property rights to the neighbouring residents, rather than to the polluting industry, would minimize the effect of positive transaction costs, since the industry will have incentives to install the filter, without any need for Coasean bargaining with the neighbours.[98] The original formulation of Coase's proposition, thus, can be restated as a normative theorem, by maintaining that, in the presence of positive transaction costs, the efficiency of the final allocation is not independent from the choice of the legal rule, and that the preferable initial assignment of rights is that which minimizes the effects of such transaction costs.[99]

Coase on the issue of transaction costs

Coase's positive theorem shows that, in a world with no transaction costs, the parties will reallocate rights among themselves, to maximize their aggregate welfare. Whatever might be the more efficient device to maximize the combined welfare of farmers and ranchers – the use of a cow hand to watch over the herd, the construction of a fence to protect all the crops, or even the use of a tiger to keep the cows far from the farmland[100] – it will eventually be chosen by the parties through their negotiations. The critics have often argued

that Coase's proposition risks becoming a mere tautology when applied to real-life situations with positive transaction costs. Coase firmly refutes such allegations.

In his retrospective analysis, Coase explains that 'The problem of social cost' was developed as an economic essay aimed at economists.[101] Coase intended to carry the standard economic assumption of no transaction cost to its logical extreme, demonstrating the inconsistency of the generally accepted idea that government intervention was necessary to improve the working of the economic system.[102] But, according to Coase, this argument was only 'a preliminary to the development of an analytical system capable of tackling the problems posed by the real world of positive transaction costs'.[103] With a rather telling consistency, Coase has attempted to correct what he perceives to be a general error in the understanding of his theorem, refuting the generalized identification of his own model with an imaginary universe of transactions without costs:

> The world of zero transaction costs has often been described as a Coasian world. Nothing could be further from the truth. It is the world of modern economic theory, one which I was hoping to persuade economists to leave. What I did in *The Problem of Social Cost* was simply to bring to light some of its properties.[104]

Already in his early writings, Coase revealed a mature understanding of the crucial role played by positive transaction costs in the economic system. As early as 1937, Coase had shown that, in the absence of transaction costs, there would be no economic basis for the existence of a firm.[105] Following the same logic, his work on the problem of social cost showed that, in the absence of transaction costs, it does not matter what the law is, since individuals will contract with each other to an optimal allocation of legal entitlements.[106]

None of his works, however, merely stop at the investigation of the properties of an abstract world without transaction costs, and Coase clearly restates that it is necessary to introduce positive transaction costs explicitly into the analysis, in order to understand the functioning of the real world: 'Without the concept of transaction costs, which is largely absent from current economic theory, it is my contention that it is impossible to understand the working of the economic system, to analyze many of its problems in a useful way, or to have a basis for determining policy'.[107] According to Coase, this important part of his argument has systematically been overlooked by the numerous commentaries to his theorem. Coase laments that his emphasis on positive transaction has practically been ignored in the secondary literature: 'This has not been the effect of my article. The extensive discussion in the journals has concentrated almost entirely on the "Coase Theorem", a proposition about the world of zero transaction costs'.[108]

According to Coase, the insistence upon the no transaction cost assumption risks undermining the normative significance of the theorem for real-world problems. The normative Coase theorem addresses the problem of positive transaction costs as the origin of a failure in the spontaneous contracting of the parties. Coase remembers: 'Law came into the article because, in a regime of positive transaction costs, the character of the law becomes one of the main factors determining the performance of the economy'.[109] It is, indeed, by relaxing the assumption of zero transaction costs, that Coase's analysis offers the most valuable insight on the effective potential of contractual arrangements in the correction of inefficient allocations of property rights.[110] The Coase theorem considers the effect of positive transaction costs. In its normative version, the theorem indicates that legal rules that minimize the effects of such costs are to be preferred for being relatively more efficient.[111] In its more complex formulation, the Coase theorem provides, indeed, a guide for such a choice.

Coase is wary, however, of simplistic generalizations, noting that no single universal formula exists for the creation of an optimal system of incentives:

> The result brought about by different legal rules is not intuitively obvious and depends on the facts of each particular case. It may be for example, as was shown earlier in this section, that the value of production will be greater if those generating harmful effects are not liable to compensate those who suffer the harm they cause.[112]

Coase theorem and other market failures

Two further situations, both related to the general notion of market failure, have been indicated as potential obstacles to the working of Coase's model. The first situation of alleged insufficiency of Coasean bargaining is occasioned by the non-excludability of the rights that are the object of Coasean negotiation. In order to shed light on the significance of this problem, one should observe that in Coase's scenario, the property right which was exchanged between the farmers and the ranchers was characterized by its excludability (that is, by the fact that individuals other than the right-holder could be excluded from its enjoyment).[113]

Economists describe this category of commodities as private goods. Difficulties arise, however, when the object of the Coasean bargaining is an entitlement which has the nature of a public good (that is, a situation in which third parties cannot be excluded from the enjoyment of that right, with no feasible way to require them to share in the costs of that resource).[114] The market may fail to cure a non-optimal allocation of rights that falls within this category.[115] In order to understand this point, consider a scenario in which the object of the Coasean negotiation consists of a non-excludable right, such as the right to enjoy pollution-free air in a residential environ-

ment. Imagine the legal limit of air emissions in that area to be fixed at twice the optimal level. In this situation, it will be the legal rule, and not the outcome of an ideal negotiation between the interested parties, that will determine the actual amount of emissions. Clean air is, in fact, a public good in the sense described above. The benefit derived from the reduction of industrial emissions is, on the one hand, non-consumable – or, more properly, not subject to rivalry in use – and, on the other hand, non-excludable.[116] It is non-consumable because the normal enjoyment of a unit of clean air by one resident does not reduce the possibility of enjoyment from others; it is non-excludable because the reduction of pollution provides a benefit to all the residents, with no feasible way to exclude those who did not agree to pay for the reduction of the industrial emissions.

When non-excludability and absence of rivalry in use prevail in the characterization of the right, the possibility of costless exchange of individual entitlements on the market is unlikely to cure a non-optimal initial assignment of property rights. Individuals will not reveal their own preferences through the price system, placing public goods among those cases that are most recidivistic to the Coasean antidote.

A second obstacle results from the absence of any barrier to the entry of new operators in the market (in economic jargon, 'ease of entry'). It is necessary to keep in mind that ease of entry is one of the characterizing features of a competitive market. The discussion that follows, therefore, far from being a mere theoretical speculation, directly regards the market structure that serves as the ideal scenario for the Coasean negotiation.

In order to more fully understand the effect of ease of entry on the Coasean model, consider the hypothesis in which the exchange of rights takes place, according to Coase's prediction, so that the ranchers agree, upon compensation, to reduce the dimension of the herd until a point of Pareto efficiency is reached. If such an agreement is reproduced on a large scale, the total quantity of meat produced in the market will fall and, given the usual negative slope of the demand curve, the price of the meat will rise. If this occurs, the agreement between the ranchers and the farmers will be short-lived. The high price of meat will attract new enterprises interested in exploiting the new potential for profit in the industry. These entrepreneurs will immediately jeopardize the stability of the initial agreement by disturbing the momentary equilibrium reached through Coasean negotiation. The farmers, in order to reduce the damage to their crops, will be forced to pay the newly entered ranchers to limit the size of their cattle herds as well. This phenomenon would recur in cycles, rendering any further agreement useless. A static analysis of the equilibrium appears incapable of weighing the applicative significance of Coase's theorem, in which the dynamic adjustments of the initial equilibrium risk corroding the holding structure of his model.

Described in these terms, however, the problem risks overstatement. One needs, in fact, to observe that the absence of barriers to the entry of new enterprises into the market does not represent, by itself, an obstacle to the functioning of Coase's model. The problem arises only when the same transfer of rights occurs on a large scale, thus influencing the prices of the market.[117] The individual rancher who reduces the dimension of his/her own herd is unlikely to influence the price of meat on the market. No dynamic adjustments will, therefore, take place in the equilibrium reached through the initial agreement, nor will there be new entries in the market. In spite of the absence of barriers, Coase's analysis would thus remain a sound prediction of individual behaviour.

Coase's legacy in law and economics

In writing 'The problem of social cost', Coase intended to correct a consolidated error of the Pigouvian interpretative tradition. From the vantage point of over 30 years, Coase does not appear to show any sign of repentance. In the conclusion of his 1988 annotations on the problem of social costs, Coase reiterates his belief that there are few reasons to give credibility to the Pigouvian approach: 'My point was simply that such tax proposals are the stuff that dreams are made of. In my youth it was said that what was too silly to be said may be sung. In modern economics it may be put into mathematics'.[118] According to Coase, the usual shortage of the data necessary to establish the level of Pigouvian taxes or subsidies renders the Pigouvian solution replete with imprecise estimations.[119] To admit that in the presence of perfect information, the system of Pigouvian taxes is impeccable,[120] does not imply that it may be the same in a world full of unknowns.

Coase formulates a rigorous argument in support of his severe conclusions. His point is that, in addition to the relative difficulty of gathering the necessary information for an effective use of the Pigouvian taxes,[121] this type of approach lacks symmetry in addressing the problem of externalities. According to Coase, economists in the Pigouvian tradition fail to consider the possible reciprocity of the effects of individual choices.[122] By labelling one agent as injurer and the other as victim, the Pigouvian tradition presumes an initial allocation of rights.[123] In such a manner, this approach falls into a serious methodological error.[124] By taxing the generator of the externality in a measure corresponding to the difference between the private cost and the social cost of his own activity, Pigou's followers fail to consider the effects of potential victims' behaviour. If the social cost of the industrial emissions is calculated by aggregating the economic disadvantages of the residents that are negatively affected by the smoke, the figure will vary with the number of individuals who fix their residence in that area. If the Pigouvian tax is imposed on the industrial activity only, there will be less incentive for each

resident to consider moving into a different neighbourhood. Coase illustrates his argument with the following scenario:

> The example I used to illustrate my argument was that of a factory whose smoke would cause damage of $100 per annum but in which a smoke-prevention device could be installed for $90. Since emitting smoke would involve the owner of the factory in paying taxes of $100, he would install the smoke-prevention device, thereby saving $10 per annum. Nevertheless the situation may not be optimal. Assume that those who would suffer the damage could avoid it by taking steps which would cost $40 per annum. In this case, if there were no tax and the factory emitted the smoke, the value of production would be greater by $50 per annum ($90 minus $40).[125]

Coase renders his reasoning even more convincing by taking into consideration the choices of new individuals who locate their residence in that area, without considering the potential increase in the costs imposed on the industrial activity. Through these arguments, Coase's analysis demonstrates the incapacity of the Pigouvian approach to consider the interdependence of the harmful effects generated by individual choices.

These are, in truth, complex questions which have engaged a whole generation of economists and policy makers. The two approaches – sound in their respective analyses – must be evaluated in light of the specific circumstances. Lawyers and policy makers will have to be particularly attentive to the respective assumptions of each tradition, weighing the relative strengths of each remedy in the treatment of the complex situations they encounter. In considering the Pigouvian and Coasean traditions, one can no longer think of two directly opposed theories, in which the first identifies the solution with the choice of the optimal parameters for the fiscal imposition, while the other maintains that, in the long run, all externalities will be cured in the market.

The issue for the jurist is of a broader scope. It deals with the unstable relationship between private and public remedies in the pursuance of social goals. The jurist may have limited opportunity to enter actively into this debate, but he/she should always treasure Coase's intuition in reappraising familiar legal issues.

Conclusions

Coase's analysis occasioned a paradigmatic shift in legal and economic analysis. His theorem, short of providing a simplistic formula for the social cost problem, suggests an alternative approach, one based on the evaluation of the relative costs of alternative assignments of rights. In 1960, Coase entrusted legal and economic scholars with the challenging task of deriving the implications of his theorem in their areas of research:

> Satisfactory views on policy can only come from a patient study of how, in practice, the market, firms and governments handle the problem of harmful effects. … It is my belief that economists, and policy makers generally, have tended to overestimate the advantages which come from government regulation. But this belief, even if justified, does not do more than suggest that government regulation should be curtailed. It does not tell us where the boundary line should be drawn. This … has to come from a detailed investigation of the actual results of handling the problem in different ways.[126]

Coase's invitation was, indeed, taken up by a number of economists and lawyers who experimented with the unparalleled analytical potential of Coase's theorem in their research. The new methodological approach suggested by Coase provided a key for the solution of difficult normative choices. The scientific and academic credibility of legal analysis no longer relies on the self-contained methodology of the Langdellian tradition, a system of instruction and analysis which had relied almost exclusively on the self-contained framework of case analysis and classification, viewing law as little more than a filing system. Coase has disclosed new grounds for a more coherent appraisal of legal and policy issues. In retrospect, Coase is aware of the far-reaching achievement of his analysis:

> Legal scholarship … moves forward in a new spirit. Ernest Rutherford said that science is either physics or stamp collecting, by which he meant, I take it, that it is either engaged in analysis or in operating a filing system. Much, and perhaps most, legal scholarship has been stamp collecting. Law and economics, however, is likely to change all that and, in fact, has begun to do so.

The revolutionary perspective of Coase's intuition unavoidably will continue to encounter the ideological and academic resistance that Thomas Kuhn predicts whenever radical shifts of analytical paradigms are involved.[127] But, in spite of any resistance, the present generation of legal scholars has witnessed an irreversible process of transformation in contemporary legal science, and as Henry Manne concluded, 'it is hard to imagine law ever again being free of the influence of the techniques and findings of objective economic analysis'.[128]

Notes

1. The author would like to express his appreciation for the helpful comments received from Harold M. Demsetz, David M. Levy, Henry G. Manne and Walter E. Williams.
2. R.H. Coase, 'The nature of the firm', 4 *Economica* 386 (1937).
3. Ibid., p. 392.
4. Ibid., pp. 393–4
5. Ibid., p. 388.
6. Ibid., p. 396.
7. Ibid., p. 391.
8. Ibid., pp. 391–2.

9. Ibid., p. 393.
10. Oliver E. Williamson, *The Economic Institutions of Capitalism* (New York: Free Press, 1985).
11. In the neoclassical conception of the firm as a production function, labour and capital are symmetrically juxtaposed. Non-standard or complex forms of organization are viewed through the monopoly approach.
12. Williamson, *supra* not 10, at 7.
13. In an ideal world characterized by unlimited rationality (in the sense of *ex ante* solution of all possible contingencies) *ex ante* solution of possible problems in the execution stage can be afforded. In coordinating individual behaviour, if one assumes perfect cognitive competence, as the neoclassical economists do, there is no need to worry about *ex post* opportunism on the part of the contracting parties. The idea is that the parties can plan for every contingency or possible problem. Williamson refers to this implied contracting process as 'planning'.
14. Whereas neoclassical economics glossed over behaviour by assuming self-interest to be both omnipresent (the classic incentive of capitalist economies) and negligible (thus the lack of any adequate analysis of strategic behaviour and opportunism), new institutional economics places a great emphasis on human behaviour. Indeed, the new institutional economics model suggests that strategic behaviour is infused throughout the contracting process, not only as *ex ante* self-interest in profit-maximization, but also as *ex post* opportunism in the execution of an already formed contract. If one assumes opportunism does not exist, there will only be self-interest at the negotiations stage (*ex ante*). The parties will maximize profits before executing a contract. Once the contract is signed, however, it will be executed in good faith. Even if you have limited or bounded rationality, you can still have asset specificity without exploitation because the promise binds the parties. Once the contract is signed, all problems and contingencies are dealt with on a win–win basis. Williamson refers to this implied contracting process as 'promise'.
15. The new institutional economics looks to *ex post* activities and investment strategy of the firm. In order to be successful the established firm must often invest in durable, transaction-specific assets. Williamson (*supra* note 10, at 376) argues that if an established firm has capital assets already in place (sunk costs) in a highly particularized area, potential entrants will be deterred from attempting entrance since the likelihood of success, especially in the short term, is low. Potential entrants will seek more secure investments in non-transaction-specific assets. This is the problem identified by Williamson in the standard entry barrier model where supposedly potential entrants have access to the same long-run average total cost curve as the established firms (ibid., p. 375). This model fails to take account of the different costs associated with specific ($k = 0$) and non-specific ($k > 0$) assets (ibid., p. 33) The governance contracting model assumes that assets are not fungible and that firms will make efforts to safeguard their investments.
16. Bounded rationality describes behaviour that is 'intendedly rational but only limitedly so'.
17. Williamson describes opportunism as 'self-interest seeking with guile'.
18. The market can be an adequate basis for contracting in a world where every component of production is redeployable to its next best alternative. In a perfectly competitive market, such redeployment of non-specific assets would result in no loss or gain. However, in a world that includes assets specific to a particular production scheme, redeployment can be either impossible or can result in substantial loss. In such a world, governance is a critical factor and thus the firm becomes a critical alternative means of contracting.
19. Coase began his undergraduate studies at the London School of Economics in 1929, as a candidate for a Bachelor degree in commerce.
20. Arnold Plant was appointed Sir Ernest Cassel Professor of Commerce at the University of London in 1930.
21. A. Plant, 'The economic theory concerning patents for inventions', 1 *Economica* 30–51 (1934) (presented before the British Association at Leicester, September 1933); A. Plant, 'The economic aspects of copyrights in books', 1 *Economica* 167–95 (1934) (presented

as the Presidential Address to the London Economic Club, 13 March 1934). Both papers are now collected in A. Plant, *Selected Economic Essays and Addresses*, (London: Routledge and Kegan Paul, 1974).

22. On 21–23 March 1981, a conference jointly sponsored by the Law & Economics Center (at the time located at Emory University and now at George Mason University) and the University of California at Los Angeles intended to bring together the distinguished group of scholars that directly contributed to the birth of the law and economics movement. Present on that occasion were William Adams; Armen A. Alchian; Gary S. Becker; Walter Blow; Robert H. Bork; Ward S. Bowman; Steven N.S. Cheung; Ronald H. Coase; Harold Demsetz; Aaron Director; Milton Friedman; Edward W. Kitch; Benjamin Klein; William M. Landes; Wesley J. Liebeler; Henry G. Manne; Jesse W. Markham; John S. McGee; John H. Moore; Thomas B. Morgan; Sam Peltzman; Richard A. Posner; George L. Priest; Kenneth E. Scott; Bernard H. Siegan; George L. Stigler; Michael J. Trebilcock; W. Allen Wallis; and Richard O. Zerbe. The transcription of the historical discussion was published as E.W. Kitch (ed.), 'The fire of truth: a remembrance of law and economics at Chicago, 1932–1970', 26 *Journal of Law and Economics* 163 (1983). During that discussion, Coase casually suggested the affinity of methodology and thematic with his teachers and fellow students at the London School of Economics: 'I might add something about Britain. The antecedents to events when I went to LSE have been left out. One has to realize that British economists have never been hostile to the study of law, in fact, they have always encouraged people. They thought that economists ought to study law, or at least they used to. Edwin Cannan, who was the professor that Plant studied with, and also Lionel Robbins … was very anxious that economists should study law and supported the development of the law faculty at the London School of Economics. … There was a dispute between him [Alfred Marshall] and Maitland about how it should be taught. Essentially, the older generation of British economists were not hostile but welcomed the introduction of law into their studies. However, that influence has really disappeared' (at 215–16).
23. R.H. Coase, 'The institutional structure of production', 82 *American Economic Review* 713, 715 (1992). The paper represents the lecture delivered by the author upon receiving the Alfred Nobel Memorial Prize in Economics, on 9 December 1991.
24. Plant's quotation was taken from Sir Arthur Salter (J.A. Salter, *Allied Shipping Control* 16–17 (New York: Garland Publishing, 1921). It is with this exact quotation that Coase starts his analysis on the nature of the firm. See Coase, *supra* note 2, at 388.
25. Coase, *supra* note 23, at 716.
26. Coase actually presented some of the elements of his later work in a lecture that he gave in Dundee at the beginning of October 1932. But the anticipation was mainly regarding his theory of the firm. A full account of these events can be found in O.W. Williamson and S.G. Winter (eds), *The Nature of The Firm: Origins, Evolution, and Development* 34–5 (Oxford: Oxford University Press, 1991).
27. R.H. Coase, 'The Federal Communications Commission', 2 *Journal of Law and Economics* 1 (1959).
28. Ibid., p. 25.
29. The reference is obviously to A.C. Pigou, *The Economics of Welfare* (Brooklyn, NY: AMS Press, 1920), particularly to the analysis contained in the second part of the fourth edition of 1932, on the theme of externalities.
30. G.J. Stigler, *Memoirs of an Unregulated Economist* 75–6 (Chicago: University of Chicago Press, 1988).
31. Coase recalls that he was urged to omit that section of his FCC article, something that he refused to do. R.H. Coase, 'Law and economics at Chicago', 36 *Journal of Law and Economics* 239, 249–50 (1993): '[The] discussion of the rationale of property rights, with its attacks on A.C. Pigou, was, however, thought to be erroneous by the economists at Chicago. I was even urged to omit this section from the FCC article. But I refused, arguing that, even if my argument was an error, it was a very interesting error. … Had it not been that these Chicago economists thought that I had made a mistake in the article on "The Federal Communication Commission", it is probable that "The Problem of Social Cost" would never had been written'.

32. The meeting was held at Aaron Director's home. Among the participants, Coase remembers the presence of Milton Friedman, George Stigler, Arnold Harberger, John McGee and Reuben Kessel. (ibid.). Further recollections of that event can be found in Stigler, *supra* note 30, at 73–90; and, more extensively on the early years of law and economics at Chicago, Kitch, *supra* note 22. Coase's arguments were memorialized in R.H. Coase, 'The problem of social cost', 3 *Journal of Law and Economics* 1 (1960).
33. R.H. Coase, 'Notes on the problem of social cost', in *The Firm, the Market, and the Law* 157 (Chicago: University of Chicago Press, 1988). Coase recognizes the merit of G.J. Stigler, *The Theory of Price* 113 (3rd edn, London: Macmillan, 1966) [hereinafter Stigler, *Theory of Price*]. The argument was subsequently restated in, G.J. Stigler, 'The law and economics of public policy: a plea to the scholars', 1 *Journal of Legal Studies* 1, 12 (1972) [hereinafter Stigler, 'Law and economics of public policy'].
34. It is sufficient here to cite H. Demsetz, 'The exchange and enforcement of property rights', 7 *Journal of Law and Economics* 11 (1964) [hereinafter Demsetz, 'Enforcement of property rights']; H. Demsetz, 'Toward a theory of property rights', 57 *American Economic Review* 347 (1967) [hereinafter Demsetz, 'Theory of property rights']; and G. Calabresi, 'Transaction costs, resource allocation and liability rules. A comment', 11 *Journal of Law and Economics* 67, 68 (1968).
35. See F.R. Shapiro, 'The most cited *Law Review* articles', 73 *California Law Review* 1540 (1985).
36. H.G. Manne, *The Economics of Legal Relationships* 123–6 (1975).
37. In order to verify the prior monopoly of the Pigouvian approach in the treatment of this issue, see F.H. Knight, 'Some fallacies in the interpretation of social cost', 38 *Quarterly Journal of Economics* 582 (1924), reprinted in *Readings in Price Theory* 160 (G.J. Stigler and K. Boulding eds, Chicago: R.D. Irwin Publishing, 1952); W. Fellner, 'External economics and diseconomies', 33 *American Economic Review* 493 (1943), reprinted in *Readings in Price Theory* 242; and T. Scitovsky, 'Two concepts of external economics', 62 *Journal of Political Economy* 143 (1954). Further reference citations are J.E. Meade, 'External economies and diseconomies in a comparative situation', 62 *Economic Journal* 54 (1952); O.A. Davis and A. Whinston, 'Externalities, welfare and the theory of games', 70 *Journal of Political Economy* 241 (1962); J.M. Buchanan and W.C. Stubblebine, 'Externality', 29 *Economica* 371 (1962); R. Turvey, 'On divergences between social cost and private cost', 30 *Economica*, 309 (1963). More recently, see also W.J. Baumol and A.S. Blinder, *Economics: Principles and Policy* (3rd edn, New York: Harcourt, Brace Jovanovich, 1985), particularly, chs 27 and 31. For a historical retrospective of the Marshall–Pigouvian tradition, see, among others, M. Blaug, *Economic Theory in Retrospect* 370–424 (4th edn, Homewood, IL: Richard D. Irwin, 1980).
38. For an influential study of the theme of externalities, almost contemporaneous with the contribution of Coase, and exemplifying the diverse approach utilized at the time, see Buchanan and Stubblebine, *supra* note 37.
39. Coase, *supra* note 2, at 39. The reference is, again, to Pigou, *supra* note 29.
40. Both Knight and Fellner had critically examined the theme of externalities, and their analyses were certainly conducive to a free market solution to external economies. Fellner, *supra* note 37, at 510, acutely observed: 'Where there are genuine diseconomies ignored by the competitive producer – smoke nuisance, wasteful exploitation of resources, etc. – these results follow not from the atomistic character of production, but from technical or institutional circumstances as a sequence of which scarce goods are treated as though they were free; and the divorce of scarcity from effective ownership may be equally complete for atomistic, oligopolistic, and monopolistic private enterprise'. Coase acknowledges the role played by Knight's paper 'Some fallacies in the interpretation of social cost' (*supra* note 37), in the formulation of his 1960 article: 'I would say that the title of my paper came from Frank Knight. … If there are traces of what Knight says in my work, it wouldn't surprise me' (see Kitch *supra* note 22, at 215).
41. Stigler, *Theory of Price*, *supra* note 33, at 113. Such a formulation, in its brevity, contains every element of the Coasean analysis. On one side, the presence of perfect competition provides the existential condition of Coase's model; on the other side, the

coincidence of private costs and social costs indicates the complete internalization of externalities. I hope to develop both these themes in the discussion that follows.

42. Demsetz, 'Theory of property rights', *supra* note 34, at 349.
43. Calabresi, *supra* note 34, at 68.
44. In his 'Notes on the problem of social cost', *supra* note 33, at 160, Coase once again uses the citation of an influential economist of the last century whose writings provided – in his formative years – the principle source of inspiration for the successful formulation of his theorem. This source is Francis Y. Edgeworth, *Mathematical Psychics* (London: C. Kegan Paul, 1881), in which the author argues that the exchange of goods between two individuals will always take place along the so-called 'contract curve' (that is, the aggregate of the best possible combinations of exchange). If it were not like this – the argument continues – the possibility of different wealth-improving agreements would remain open, with the reciprocal advantage for each of the contracting parties.
45. For yet another analysis on the validity and the limits of this analogy, see generally, Demsetz, 'Enforcement of property rights', *supra* note 34.
46. In line with Edgeworth's reasoning, an important proposition of the Coase theorem has been to affirm that while the legal system establishes the initial allocation of rights and liabilities, in the end it is the dynamic of the market to determine their final allocation. Whenever the initial allocation is not optimal, the owners of the rights will have an incentive to transfer them to other individuals who value them more. Such an exchange will continue until there is no further room for reciprocal profit. As we learn from microeconomic theory, the potential for reciprocal benefit in the exchange will not be exhausted until each right is in the hands of the highest valuing individual. The Coase theorem, thus, predicts that in a competitive market environment without transactional impediments, the final allocation of transferable rights cannot be improved upon.
47. According to this understanding of the Coase theorem, in order to guarantee the efficiency of the system, it is enough to remove every legal and material impediment to the free transfer of individual rights. As R. Cooter observes in 'Coase theorem', in 1 *The New Palgrave. A Dictionary of Economics* 457 (J. Eatwell, M. Milgate, P. Newman eds, London: Macmillan, 1987), one of these impediments – perhaps that of the greatest interest to the jurist – can be identified in the existence of uncertainties regarding the content of the rights. An imprecise definition of such content renders its valuation and exchange problematic. Similarly, situations of legal instability (inducing uncertainty with respect of future laws) may occasion informational asymmetries delaying the process of reallocation of legal entitlements. The transferability of some subjective legal positions finds different constraints in different legal systems. On the issue of freedom of object in contractual agreements, see M.J. Trebilcock, *The Limits of Freedom of Contract* (Cambridge, MA: Harvard University Press, 1993). Generally, in order for the conditions for the free exchange to be verified in Coase's model, it is necessary to define the content of the rights without ambiguity, and to provide legal means to enforce their transfer. For further analysis, see also Demsetz, 'Enforcement of property rights', *supra* note 34; and Demsetz, 'Theory of property rights', *supra* note 34.
48. The occasion was the First Annual Meeting of the American Law and Economics Association, held in Champain-Urbana (Illinois) on 24–25 May 1991. Those who had the good fortune to be at that meeting will never forget the affable presence of Professor Coase, who agreed to be the first of the four authoritative speakers only after ascertaining that the chosen criterion, age, called upon him to open that historical event.
49. Among the principle sources of this criticism see S. Wellisz, 'On external diseconomies and the government assisted invisible hand', 31 *Economica* 345 (1964); and G. Calabresi, 'The decision for accidents: an approach to non-fault allocation of costs', 78 *Harvard Law Review* 713 (1965).
50. Among these, see D.H. Regan, 'The problem of social cost revisited', 15 *Journal of Law and Economics* 427, 431–3 (1972); and G.W. Nutter, 'The Coase theorem on social cost: a footnote', 16 *Journal of Law and Economics* 503 (1968).
51. For this variation on the general theme of distributional effects, see again, Wellisz, *supra* note 49; and Calabresi, *supra* note 49.

52. The citations on this point would be overwhelming. In holding most references for the discussion that follows, at this point it is sufficient to cite Cooter, *supra* note 47, at 457, who analogizes a transaction without costs to an unattainable theoretical model, such as a plane with no air resistance in physics. The assumption of no transaction costs, according to Cooter, calls for the adoption of a legal rule capable of minimizing the costs necessary for the transfer of the right. According to this reading of the Coase theorem, legislators will be freed from allocative concerns, but they would remain burdened with the important task of promoting the elimination of legal and fiscal impediments to the free exchange of rights in the market. In a word, voluntary transfers of entitlements should be encouraged, and every effort should be made to reduce the risk and costs associated with subsequent litigation.
53. Calabresi, *supra* note 49. For a substantial change in his views on this point, see Calabresi, *supra* note 34.
54. Wellisz, *supra* note 49, at 345.
55. Calabresi, *supra* note 34.
56. Calabresi, *supra* note 49, at 730–31; and G. Calabresi, 'Fault, accidents and the wonderful world of Blum and Kalven', 75 *Yale Law Journal* 216, 231–2 (1965).
57. Calabresi, *supra* note 34, at 67.
58. Ibid., pp. 67–8. The reasoning used here by Calabresi is very similar to that which will later be elaborated by Stigler, 'Law and economics of public policy', *supra* note 33, at 12. Stigler argues that, in the absence of transaction costs, even a monopoly situation will be cured through the voluntary bargaining of producers and consumers, so that total output of natural monopolists will be identical to that of a competitive producer. This chapter will briefly examine the significance of his analysis.
59. H.M. Demsetz, 'When does the rule of liability matter?', 1 *Journal of Legal Studies* 13 (1972).
60. Ibid.
61. Ibid., p. 19.
62. Regan, *supra* note 50, at 432–3.
63. Stigler, 'Law and economics of public policy', *supra* note 33, at 12.
64. Regan, *supra* note 50, at 431–2; and Nutter, *supra* note 50.
65. Ibid.
66. Calabresi, *supra* note 49.
67. Wellisz, *supra* note 49.
68. For additional arguments on the existence of distributional effects, see P. Burrows, 'On external costs and the visible arm of the law', 22 *Oxford Economic Papers* 39 (1970); and E.J. Mishan, 'The economics of disamenity', 14 *Natural Resources Journal* 55, 62–4 (1974). With partially different conclusions, see Regan, supra note 50, at 433, according to whom there are wealth transfers only in the presence of individual gains not reflected in the price of the legal entitlement.
69. For a discussion of the problem of social cost that preceded Coase's analysis, see Knight, *supra* note 37.
70. Coase, *supra* note 33, at 171.
71. Ibid.
72. The key here might be that the assumption of zero transaction costs ideally implies that the affected parties have perfect knowledge of expected fluctuations in the system of property rights. This means that the affected parties are never surprised by a legal change, and can fully protect themselves from any wealth redistribution. See Coase, *supra* note 33, at 171–2.
73. This analysis by Coase seems to rely on the earlier study of S.N.S. Cheung, 'Transaction costs, risk aversion, and choice of contractual arrangements', 12 *Journal of Law and Economics* 23 (1969).
74. Coase, *supra* note 33, at 172–3. On this point, the logic of Coase's analysis has been criticized for lack of practicality. The inevitable increase of transaction costs tied to a similar contractual mechanism, has been pointed out by O.E. Williamson, 'Contract analysis: the transaction cost approach', in *The Economic Approach to Law* 39 (P.

Burrows and C.G. Veljanovski eds, Burlington, MA: Butterworth-Heinemann, 1981). Coase appears to be aware of the practical difficulties of such a contractual dynamic. Even though stressing the limited significance of the criticism, Coase, *supra* note 33, at 174, concedes: 'It cannot be denied that it is conceivable that a change in the criteria for assigning ownership to previously unrecognized rights may lead to changes in demand which in turn lead to a difference in the allocation of resources, but, apart from such cataclysmic events as the abolition of slavery, these effects will normally be so insignificant that they can safely be neglected. This is also true of those changes in the distribution of wealth which accompany a change in the law when there are positive transaction costs and it is too costly for the contracts to cover all contingencies'. Alternative schemes of compensation for losses occasioned by changes in the legal rule have been examined by G. Tullock, 'Achieving deregulation – a public choice perspective', *Regulation* 50 (November/December 1978); J. Quinn and M.J. Trebilcock, 'Compensation, transition costs and regulatory change', 32 *University of Toronto Law Journal* 117 (1982); and, more extensively, by J.L. Knetsch, *Property Rights and Compensation: Compulsory Acquisition and Other Losses* (Dublin: Butterworths, 1983).

75. Calabresi, *supra* note 49.
76. Wellisz, *supra* note 49.
77. The transaction cost literature seldom differentiates between *no* transaction cost, which is a theoretical construct, and *low* transaction cost, which may often be a real-world phenomenon. Many commentators seem to discuss Coase's assumptions without ever defining which situation they are addressing. In the absence of transaction cost, there are no strategic or factual impediments to bargaining, property rights are perfectly defined, and information is costless. The debate on the point will greatly benefit from a more clear distinction between information and negotiation costs, on the one hand, and the costs created by strategic bargaining on the other.
78. For an interesting perspective on the role of contractual strategies in the area of externalities, see also Davis and Whinston, *supra* note 37, at 241–62.
79. For a subjectivist approach to this problem in a low transaction cost setting, see J.M. Buchanan, 'Rights, efficiency, and exchange: the irrelevance of transactions cost', in *Economics: Between Predictive Science and Moral Philosophy* 153 (College Station, TX: Texas A&M University Press, 1987).
80. Demsetz, *supra* note 59, at 21.
81. As observed by Demsetz, the negative implications usually carried by the idea of strategic behaviour should not affect the understanding of this point. Whether the strategy is part of the faculties of the owner of the right, or whether instead it should be considered as a form of extortionistic behaviour in abuse of the right itself, is an ethical question which falls outside the interest of the economist (ibid.).
82. The entire analysis presupposes that the so-called 'income effect' can be ignored. In general, a different allocation of property rights implies a different distribution of wealth between the individuals involved. Different initial endowments generate different final allocations, notwithstanding an equal level of efficiency. In order for the final allocations to be identical, it is necessary that the utility functions of the individuals involved are almost linear. The absence of the income effects implies, in this sense, that the demand functions for the good are independent of the income level.
83. On this point, see as well the discussion of Calabresi, *supra* note 34, at 68.
84. In game theory, such a threat in the strategy would be classified as a non-credible threat. The availability of information as to the costs and the benefits of the other contracting party is necessary in order to ascertain the credibility of the threat of others. Nevertheless, in situations of incomplete information, it is possible to have some indicia to evaluate the credibility of the adverse strategy. The game-theoretic literature on the subject is plentiful.
85. On the limits of contractual strategies found in the competitive structure of the market, see M. Shubik, *Strategy and Market Structure* (New York: Wiley, 1959).
86. Economists classify these situations of economic advantage as situations of *rent*. See Wellisz, *supra* note 49, at 345–62, who conducts a first investigation on the theme of

rents and their connection with the model of Coase. See also Coase, *supra* note 33, at 163–70. The content of Coase's analysis on the point will emerge in the course of this chapter.

87. Demsetz, *supra* note 59, at 24.
88. Ibid., pp. 24–5.
89. Most recently, Mancur Olson, 'The Coase theorem is false?', presented at the May 1997 American Law and Economics Association Annual Meeting held in Toronto, Canada.
90. For a retrospective assessment of the role played by his transaction cost paradigm, see the introductory chapter of R.H. Coase, *The Firm, the Market and the Law* 6 (Chicago: University of Chicago Press, 1988). On this point, see R. Cooter and S. Marks, 'Bargaining in the shadow of law: a testable model of strategic behavior', 11 *Journal of Legal Studies* 225, 242 (1982).
91. Here it is enough to think of the transaction costs that are in all likelihood necessary in order to reach an agreement between an elevated number of distant parties.
92. Cooter, *supra* note 47, at 457–58. According to the author, the problem of information costs should be addressed within the broader framework of perfectly competitive markets. See also R. Cooter, 'The cost of Coase', 11 *Journal of Legal Studies* 1 (1982). As evidenced by O. Schultze, *The Public Use of Private Interest* (Washington, DC: The Brookings Institution Press, 1977), market structures different from that of perfect competition can be an obstacle to the market dynamic contemplated by Coase. According to Schultze, therefore, the legal system will have to foster the conditions for a healthy competitive market for the transfer of rights. On the point, see more generally C.J. Dahlman, 'The problem of externality', 22 *Journal of Law and Economics* 148 (1979).
93. For an influential and, yet, controversial examination of the problem, see Williamson, *supra* note 74.
94. Demsetz, *supra* note 59, at 20.
95. Ibid.
96. This example is borrowed from A.M. Polinsky, *An Introduction to Law and Economics*, 11–14 (2nd edn, Boston, MA: Little, Brown, 1989), who examined the significance of Coase's analysis in the presence of positive transaction costs.
97. We shall observe later that clean air – in economic terms – is a *public good*, lacking the features of *excludability* and *rivalry* in their use, which instead characterizes private goods. In situations of this kind, there is little viability for a market solution to externalities. For a systematic investigation on the theme, see R. Cornes and T. Sandler, *The Theory of Externalities, Public Goods, and Club Goods* (Cambridge: Cambridge University Press, 1986).
98. The industry, having to choose between paying damages to his/her neighbours ($750), the acquisition of five dryers ($500) or the installation of a filter ($300), will rationally opt for the last possibility, being the most economical. Thus, the efficiency of the result will not be altered in the presence of positive transaction costs. Further proof can be found in Polinsky, *supra* note 96.
99. Among the various attempts to reformulate the Coase theorem in the presence of positive transaction costs, see again Polinsky, *supra* note 96. Calabresi, *supra* note 34, at 72–3, observes that many authors read the Coase theorem as if it suggests that the absence of public intervention is in reality the major cure for externalities. Calabresi does not share this common interpretation, observing that Coase's analysis offers invaluable instruments for the choice of liability rules and for the identification of the areas in which public intervention becomes desirable.
100. The unusual example is not the license of this author but is in fact provided by Coase himself, *supra* note 33, at 175.
101. Coase, *supra* note 31, at 251–3.
102. Coase, *supra* note 2, at 18. Coase criticizes some economists and policy makers for overestimating the advantages of government regulation.
103. Ibid., pp. 24, 15.
104. Coase, *supra* note 33, at 174.
105. Coase, *supra* note 2.

106. Ibid., p. 14.
107. Ibid., p. 6.
108. Ibid., p. 15.
109. Coase, *supra* note 31, at 251.
110. In this direction, see again, Calabresi, *supra* note 34, at 72–3.
111. See Polinsky, *supra* note 96, at 14.
112. Coase, *supra* note 33, at 178–9. For a discussion on the point, see R.O. Zerbe, 'The problem of social cost: fifteen years later', in *Theory and Measurement of Economic Externalities* 29, 33 (S.A.Y. Lin (ed.) Academic Press: 1976).
113. For an excellent study on this point, see S. Cheung, 'The structure of a contract and the theory of a non-exclusive resource', 13 *Journal of Law and Economics* 49, 49–70 (1970).
114. Here reference is made to all those goods which lack *excludability* in their enjoyment. For further analysis, see again Davis and Whinston, *supra* note 37; Turvey, *supra* note 37; Demsetz, 'Enforcement of property rights', *supra* note 34. The nature of public goods creates problems for the functioning of the price system. In the typical example of a public good, the lighthouse that each sailor utilizes, but for which no one is willing to pay, the market is incapable of inducing consumers to reveal their preferences through prices. Consequently, the supply of public goods will not be induced by the consumers' demand, but will be determined by public choices. Each person will consequently be in a position to derive a benefit from these goods, even though they are not contributing to its cost. More attentively on the subject, O. Davis and A. Whinston, 'On the distinction between public and private goods', 57 *American Economic Review* 360 (1967); and P.A. Samuelson, 'The pure theory of public expenditure', 36 *Review of Economics and Statistics* 387 (1954). Coase himself discusses the problem in 'The lighthouse in economics', 17 *Journal of Law and Economics* 357 (1974).
115. For a recent argument against this common belief, see the excellent study by F. Foldvary, *Public Goods and Private Communities, the Market Provision of Social Services* (Cheltenham: Edward Elgar, 1994).
116. Economic literature prefers to speak of *nonrivalry*, describing this characteristic as the absence of an increase in the marginal costs of production as an effect of a marginal increase in the number of users. The example often used is that of a tract of interstate road with low traffic, which is fit to satisfy the need of an additional user at no additional cost. For a further analysis, see R.S. Pindyck and D.L. Rubinfeld, *Microeconomics*, 638–41 (Prentice Hall: 1989). For further ideas on the problem, see S.H. Gordon, 'The economic theory of common property resources: the fishery', 62 *Journal of Political Economy* 124 (1954).
117. In economic terms, the assumption is that the single economic agent is a *price taker*, incapable of influencing the prices in the market with his/her own decisions. In an ideal market of atomistic competition, each producer offers a small fraction of the total supply of the market, so that the choice of the quantity produced or demanded by the individual agent will not have any effect on the price of the product. The price will be, instead, determined by the aggregate supply and demand curves.
118. Coase, *supra* note 33, at 185.
119. Coase is certainly not the only author to denounce this shortcoming of the Pigouvian approach. Many authoritative economists have contributed to the understanding of the limits of this solution. The citations would be lengthy and hardly relevant for the analysis conducted here. For a representative sample, see J.M. Buchanan, 'Politics, policy and the Pigouvian margins', 29 *Economica* 17 (1962); C. Plott, 'Externalities and corrective taxes', 33 *Economica* 84 (1966); and E.J. Mishan, 'Reflections on recent developments in the concept of external effects', 31 *Canadian Journal of Economics and Political Science* 3, 33–4 (1965). For the applicative difficulties of the system of Pigouvian taxes in the hypothesis of goods with joint supply, see also, J.M. Buchanan, 'Joint supply, externality, and optimality', 33 *Economica* 404 (1966).
120. Coase here refers to a statement made by W.J. Baumol, 'On taxation and the control of externalities', 62 *American Economic Review* 307 (1972), in which the influential econo-

mist affirms that 'taken on its own grounds, the conclusions of the Pigouvian tradition, are, in fact, impeccable' (ibid., p. 307).

121. Coase does not examine the difficulties of a political nature connected with the adoption of a system of Pigouvian taxes.
122. Of this shortcoming Coase is critical of Baumol. For Baumol's criticism of Coase's approach, see Baumol, *supra* note 120.
123. Cornes and Sandler, *supra* note 97, at 59.
124. Coase's argument on the point seems to be refuted by various empirical psychological studies. While rules must always be balanced, the human psyche seems to value current entitlements more than their market price (that is, the actual consumers' surplus is not captured by the objective market price). On the point, see D. Kahneman, J.L. Knetsch and R.H. Thaler, 'Experimental tests of the endowment effect and the Coase theorem', 98 *Journal of Political Economy* 1325–48 (1990), where the authors explain how measures of willingness to accept greatly exceed measures of willingness to pay. They call this phenomenon the 'endowment effect'. They use this notion to explain the observed undertrading of entitlements in a Coasean setting. For an introduction to this literature, see E. Hoffman and M.L. Spitzer, 'Symposium on law and economics: experimental law and economics. An introduction', 85 *Columbia Law Review* 1037 (1985); R.E. Scott, 'Error and rationality in individual decision-making: an essay on the relationship between cognitive illusions and management of choices', 59 *Southern California Law Review* 329 (1986); R.C. Ellickson, 'Bringing culture and human frailty to actors: a critique of classical law and economics', 65 *Chicago-Kent Law Review* 23 (1989); R.L. Hausen, 'Comment: efficiency under informational asymmetry. The effects of framing on legal rules', 38 *University of California-Los Angeles Law Review* 391 (1990).
125. Coase, *supra* note 33, at 180.
126. Coase, *supra* note 2, at 18–19.
127. T.S. Kuhn, *The Structure of Scientific Revolutions* (2nd edn, Chicago: University of Chicago Press, 1970).
128. H.G. Manne, *The Intellectual History of George Mason University School of Law* 4 (Arlington, VA: Law and Economics Center Publication, 1993).

2 Property rights and their partitioning

Christian Müller and Manfred Tietzel

Property rights can be defined as socially recognized entitlements of individuals to use a good. Here, the term 'property' is used in a broad sense and is meant to encompass the relations of actors to all scarce goods yielding utility, including rights not only to material resources but also to immaterial, human rights such as the right to vote and that of free speech (Furubotn and Pejovich, 1974, p. 3). One commonly distinguishes between the right to use a resource (*usus*), the right to appropriate returns (*usus fructus*), the right to change the form and substance of assets (*abusus*) and the right to sell or lease some or all of these rights to another user (*alienation*). Neoclassical microeconomics implicitly assumes all these rights to be fully laid in the hands of one single user and focused only on 'the forces determining the price and the number of units of a good to which these rights attach' (Demsetz, 1967, p. 347). In sharp contrast to this, the theory of property rights (see the surveys in Furubotn and Pejovich, 1972; DeAlessi, 1980; Tietzel, 1981; Eggertsson, 1990; Richter and Furubotn, 1996) emphasizes the possibility of differences between entitlement structures. For reasons discussed below it is held that any given property rights structure functions as a system of incentives consisting of rewards and punishments; this extended approach sheds light on the institutional aspects of choice that are taken to be given exogenously in orthodox theory. The traditional focus is regarded as being too narrow, as it neglects the fact that, owing to different property rights structures, physically homogeneous goods can differ considerably from the users' points of view. The property rights approach therefore holds that it is not 'specific commodities', defined by their technical properties, that are exchanged in markets but 'effective commodities', that is, rights bundles shaped by a particular set of legal restrictions determining the socially recognized use of goods (Furubotn and Pejovich, 1974, pp. 4f.).

The ideal of private property

The more completely and privately the property rights to an asset are defined, the more will its holder be inclined to maximize the full value of the resource. This insight, which can be taken as the central starting point of property rights economics, is not novel (it is foreshadowed, for example, in the biblical 'parable of the good shepherd'; see John 10: 11–13). What is new in the theory of property rights is the systematic analysis of the behavioural consequences that are to be expected when the salutary role of private rights is

violated. This question has been neglected and left unanswered within orthodox economics, for the only type of property that neoclassical theory takes into account is fully defined private property. Each owner of a commodity is the unrestricted residual claimant of the asset who bears *all* benefits and costs (the *usus fructus*) of resource use. In such a highly specific property rights regime, any person benefiting another will be fully rewarded by the beneficiary and any person harming another has to fully compensate the victim. To put it into common economic terminology, Pareto-relevant externalities, positive or negative, are implicitly assumed to be nil.

As the celebrated Coase theorem (Coase, 1960) states, the existence of such a first-best world requires all transactions costs to be zero – including all monetary and time costs that accrue as a result of the specification, exchange, supervision and enforcement of property rights. If, and to the extent that, people can costlessly bargain for rights, the particular assignment of legal liabilities has no effect on the achievement of allocative efficiency. The Coasean approach contradicts the orthodox Pigouvian perspective of an externality relation according to which the person who harms another has either to fully compensate the harmed person or to give up his/her activity. In sharp contrast to this, Coase takes into account that externalities always consist in two-sided relationships. Consequently, the Coase theorem holds that, in a world of costless bargaining, also the externally affected party may compensate the acting party in order to achieve efficiency. Regardless of who initially owned the (*de facto*) right to affect another person's welfare, the emerging property rights allocation is efficient since that person ultimately turns out to be the owner of the right who values it most highly. Of course, there will be different distributional implications of both solutions.

Transactions costs and the partitioning of rights

Transactions costs will seldom be zero under real-world conditions. It may be costly to provide information on product qualities in order to identify profitable production opportunities or exchange partners. Furthermore, any bargaining requires time and other resources to be spent, as well as, at a later time, costs to monitor and enforce the keeping of the contract. With transactions costs being positive, property rights will never be fully defined since individuals will have to bear costs in order to establish exclusive rights and to exploit the full potential of a formally owned resource. As long as the definition of property rights is costly, the degree of rights delineation will depend on an individual cost–benefit calculus of the affected parties. As a well-known theorem of Demsetz (1967) states, exclusive property rights will be established when net gains from exclusivity are positive. Since perfect delineation of rights will be prohibitively costly, there will always be domains with property rights being absent and, hence, public.

Moreover, even when rights are formally assigned to private owners, entitlements, owing to a similar calculus of gains and costs of enforcement, may be attenuated: asset proprietors will find it worthwhile to claim their ownership rights as long as marginal benefits exceed marginal costs (Demsetz, 1964; Barzel, 1989, pp. 64ff.). As owners, therefore, deliberately abstain from exploiting the entire potential of 'their' resource, there will always be an optimal degree of externality leaving some valued attributes placed in the public domain. Shopping centres, for example, often allow for free parking and tolerate non-customers' use of their parking space (Demsetz, 1964). Shoppers, by making their purchases, produce positive externalities for all non-shopping parkers. The shopping centre may well fence in the parking space and pay a guardian to exclude non-payers by employing a price mechanism. This solution, however, may be disadvantageous even to the shoppers, in that the parking price would have to cover the exclusion costs and might well exceed their marginal valuation of using the parking lot. Thus even purchasers producing external economies to other users may prefer free parking access for all rather than costly exclusion of free-riders. Another illuminating example is that of restaurant owners who supply their patrons with free salt while overpricing other attributes to cover its costs. Equally, in their pricing schemes they fail to distinguish between fast and slow eaters, although the latter use the provided space and china for a longer period than the former (Barzel, 1989, pp. 66, 72). An unusual pricing device to economize on transactions costs was observed by Cheung (1980) in Hong Kong, where the owners of movie theatres underpriced the better seats in order to get them fully occupied. This allowed them to economize on the costs of price discrimination.

There are two main forms of attenuation of property rights (Alchian and Demsetz, 1973, p. 18), the first being *decision sharing*, with rights held by a group of people who are only able to exercise their entitlements collectively, as is often the case with public ownership. A second important form of rights attenuation consists in the *domain partitioning* of rights uses by several people. Consider, for instance, a unique piece of land:

> A may possess the right to grow wheat on it. B may possess the right to walk across it. C may possess the right to dump ashes and smoke on it. D may possess the right to fly an airplane over it. E may have the right to subject it to vibrations consequent to the use of some neighboring equipment. And each of these rights may be transferable. In sum, private property rights to various partitioned uses of the land are 'owned' by different persons. (Alchian, 1977, pp. 132f.)

The problem of common property

Inasmuch as, because of transactions costs, property rights are not exclusive, privately perceived benefits and costs will differ from total gains and costs

(Furubotn and Pejovich, 1972, p. 1144). As long as nominal owners and actual holders of rights to rival goods are not the same persons, the latter are able to use the nominal entitlements of the former as common property while imposing their use costs on the nominal rights holders. To the extent of the positive externality, demand for the resource exceeds the optimal level *because* others pay its price. The resulting problem of overexploitation of commonly owned resources may be viewed as the central problem of property rights economics. Using the terminology of standard public goods theory, overexploitation is to be expected to occur whenever the consumption of an asset is rival (subtractible) and non-paying users are not excluded from extracting benefits from it. The problem of overuse is not confined to resources with property being formally delineated as communal. Owing to positive transactions costs, common-property problems may also emerge in cases where there is no formal ownership, as well as in the cases of communal property or even private property, of a resource.

Consider first the anarchical case of *no property* where people *de facto* hold a 'right to all' (Hobbes, 1949, p. 27) that 'entitles' each and every person to appropriate whatever he/she wants and as much as he/she is able to get, including the 'right' to take by force goods produced by other individuals (for a formal model of anarchical behaviour, see Bush and Mayer, 1974; Grossman and Kim, 1995). As even in anarchy any 'right' has a 'duty' as a correlative, a 'right to all' cannot accrue to anyone in a multi-person society. Rather, individuals will engage in an appropriation race against each other that Hobbes (1914, p. 64) vividly described as a 'warre of every man against every man'. Commonplace examples of overuse problems of resources to which no property rights are devised are those of natural resources where formal rights are non-existent. Air, fishing grounds, oil pools, forests or groundwater basins are cases in point. All these examples of resources with exclusive rights being absent are usually connotated with the notion of a 'tragedy of the commons' since Garret Hardin (1968) in his celebrated article paradigmatically explored his example of a 'pasture open to all' with many villagers driving on their cattle. Each herdsman, as a rational non-altruist, will try to keep as many cattle on the commons as will meet his *individual* profit maximum. While the gains of his effort are strictly private, the associated costs are shared by all herdsmen, with himself bearing only a small fraction. Since a similar calculus holds for each individual, the villagers are locked into a prisoner's dilemma where *collective* welfare – which is maximized at a lower than the individually optimal level of effort – is unattainable owing to individually rational behaviour (a more formal analysis of common-property problems was presented earlier by Gordon, 1954; for an important exception, see Tietzel 2001).

Similar problems may arise when property, by its formal definition, is *communal*. Consider the example of highways that, at least in principle,

could be supplied as club goods where free-riders are excluded and market provision is possible (this example is taken from Alchian and Demsetz, 1973, p. 21). However, governments often choose to supply roads as freeways, with each driver being entitled to free access. During rush hours the actual use of the allegedly purely public good day by day exceeds its capacity. In order to avoid congestion, each driver may be willing to pay other persons to use alternative routes during these hours. However, even were such bargainings not prohibitively costly, the communal right system encourages the drivers to have the others pay, for any payer would individually bear the cost of providing a public good the benefits of which are dissipated among all users. Hence the government's decision to leave open the access to highways creates a free-rider problem. Furthermore, in the long run, the communal right to roads may induce additional traffic, since persons, other things being equal, have an incentive to substitute the apparently free use of carways for the costly use of public transport.

There are many other examples of common-property problems that are government induced and regularly the unintended (and often also unexpected) byproducts of an inefficient distributional intervention. In the case of the so-called 'merit goods' (Musgrave, 1987; see Tietzel and Müller, 2002 for a critique from a constitutional-economics perspective), paternalistic governments try to change the market allocation of private or club goods, the demand for which is regarded as being 'insufficiently low'. Here the state intervenes, for example, by (wholly or partly) subsidizing the market supply of opera performances, building of houses or school lunches or by directly regulating prices to a lower than equilibrium level. In the extent of the price reduction of these commodities, the state intervention amounts to a governmental definition of communal property rights to these goods, voluntarily leaving some or all of the goods' valued attributes in the public domain. From a government's egalitarian point of view it may be desirable equally to treat private and public goods as free goods. However, the distinguishing feature between communal property rights to free as opposed to scarce resources is that, because of their non-subtractibility, the former remain communal in the consumption process. Thus, in view of a good's abundance, the establishment of exclusive rights is neither possible nor necessary. In sharp contrast to this, the consumption of scarce resources, owing to their shortage, requires any one rationing procedure defining exclusive rights to them. In a limited world where goods are obtainable at a zero price, rationing will regularly occur in terms of time and other resources spent to appropriate these goods. Even when no private rights to these resources are formally defined, subtractibility establishes a *de facto* system of rights on a first come, first served basis. Consequently, the right to use the asset is communal only as long as it is not appropriated by someone; it turns into a *private* right to use the resource once

it is captured (Alchian and Demsetz, 1973, p. 22). In this situation, no one will have an incentive to form an orderly queue in which each waiting person grants the individuals in front of him (or her) the *de facto* right to obtain the good at a lower time price than he himself pays (Barzel, 1974, p. 86). Rather, all individuals will engage in an appropriation race, competing with each other in order to be the first to get the private right to the resource use. Overexploitation and rent dissipation are the natural consequences to be expected.

The common-property problem is even aggravated if one assumes a dynamic perspective. If a government decides to allocate non-public goods under a communal property regime along with taxation according to the ability-to-pay principle, the correlation between demand and supply of public services is interrupted. From an individual's point of view, under such a rights arrangement, it is worthwhile not only to demand the goods supplied at a higher level than the optimal, but also to press for an expansion of governmental supply of private goods in order to externalize the costs of satisfying private needs. This individually rational cost–benefit calculus, however, leads directly to a collective self-damage. Permanent increase of claims to the budget and, if satisfied, an explosion of government expenditure, will be the consequence (Bonus, 1978, pp. 75ff.). Since this, *ceteris paribus*, curbs economic efficiency and increases unemployment, benevolent (and vote-maximizing) politicians will hasten to increase the government supply of scarce resources, this time in order to mitigate the social damages induced by their own policy. Again, this will give rise to government growth, and so on. Thus it seems that, to some extent, the problems currently faced by welfare states can be explained in terms of an inefficient government definition of property rights.

It contributes to the explanatory power of the property rights paradigm that it even allows for an analysis of common-property problems when *private* entitlements are established. Suppose, for instance, an employment relationship between an owner–manager of an enterprise and a labour supplier. A twofold exchange of property rights is involved here: first, the entrepreneur acquires a private right to exploit temporarily the agent's effort; second, the principal delegates some of his (or her) exclusive rights to the use of firm equipment to the agent. Since information and measurement costs of labour efforts often prevent the employee's compensation being tied either to his (or her) effort or to his marginal product (which, for example, is impossible in the case of 'team production' in the sense of Alchian and Demsetz, 1972), fixed-wage contracts are usual substitutes for input- and output-dependent compensation. To the extent that the principal, as a result of prohibitive monitoring costs, is unable to exclude the agent from a misuse of his property rights, the employee has an incentive to appropriate what is offered to him free. Thus, as in the communal rights cases discussed above, the employee

may take advantage from pursuing his private interests and externalize the costs on the principal.

Consider for simplicity the case of the principal's complete inability to monitor the agent's effort. The 'tragedy of the commons' then has two facets: first, by 'shirking' (Alchian and Demsetz, 1972) an opportunistic employee may reappropriate the private right to his effort that he has sold to the principal. Since with a fixed-wage contract payment does not depend on his labour supply, he will only provide that level of effort which directly yields intrinsic utility to himself. Any supply of labour exceeding this level would produce external benefits to the principal and would, from the agent's point of view, amount to paying a 100 per cent tax on the increase in output induced by greater effort (Barzel, 1989, pp. 37f.). A second kind of common-property problem is posed by an agent's 'consumption on the job' (Jensen and Meckling, 1976, pp. 312ff.) by which he appropriates fringe benefits from the firm equipment the principal delegated to him. A managerial agent, for example, may, at the cost of the firm's owner, derive utility from a larger than optimal computer, from a luxurious company car satisfying his needs for status or from judging secretaries by their appearance rather than their productivity.

Solutions to the problem of common property

In principle at least, there are two conceivable solutions to the problem of common property that one might label 'regulation' and 'privatization' respectively or, according to the two mainstreams of welfare economics, refer to as the Pigouvian and the Coasean solutions to the commons problem. By means of *regulation*, a bundle of given rights to a commonly used resource remains unaltered but the harmed persons try to enforce their rights more fully by narrowing access to the resource to a clearly defined group of users and by legislating rules for internal governance of the commons. As many empirical studies indicate, such organizations of joint users of natural resources have turned out to be successful which succeeded in specifying time, place, technology and/or quantity of resource units to be appropriated (see Ostrom, 1990; Anderson and Simmons, 1993; Ostrom et al., 1994). Similarly, access to highways and other roads is frequently regulated by user tolls and traffic regulations that serve as rules for internal governance which, for instance, determine in which lane and how fast to drive. By the same token, employers may choose to mitigate the commons problem in agency relations by engaging in policing the employee's effort and output and in sanctioning violations of rules of governance. According to Alchian and Demsetz's hypothesis (Alchian and Demsetz, 1972), firms are founded mainly in order to allow for easier monitoring in order to cope with common-property problems involved in team production. All these solutions to overuse problems are 'Pigouvian'

in the sense that the imposition of tolls and sanctions is equivalent to Pigouvian taxes that are imposed on the harming persons. As in Pigouvian welfare theory, the externality relation is taken to be one-sided, in that it is always the harmed persons (or a benevolent government acting in their place) to whom the residual claim to the common property is assigned and who take the active part in mitigating the problem. Correspondingly, the role of the harming persons is entirely reactive.

The contrary is true when *privatization* of the resource's free attributes is the strategy used to reduce common-property problems. Privatization means that, in order to achieve allocative efficiency, the residual claim to common property should be assigned to that part of the externality relation that mainly affects the outcome. The strategy of privatization may be interpreted as 'Coasean' in that it is not only the harmed person to whom the right to the commonly used attributes may be granted. The harming person may also become the residual claimant of the unpriced attributes of a resource. Consider once again the case where legal property to a commonly used resource is absent, as, for instance, in the Hobbesian jungle or in Hardin's herdsmen example. These anarchical or anarchy-like situations most vividly demonstrate the welfare gains that all can achieve by accepting general behavioural constraints such as a system of socially endorsed and enforced property rights that presupposes everybody's abandonment of the 'right' to steal from others' endowments or of the 'right' to graze cattle at the expense of others. If the anarchical 'right to all' is divided into generally recognized parts and individually assigned to the users, all will gain, for two reasons. First, each owner is now granted a private right to use his/her part of the resource which accrues to him independently of its actual exercise. Thus no more productive resources have to be spent on participation in an appropriation race with others. Second, since each person is enabled to alienate parts of his property, voluntary exchange becomes possible, which will give each owner an additional incentive to maintain the market value of his asset. Both of these aspects illustrate the salutary role a state can play in order to reduce the commons problem in anarchy (see, for example, Buchanan, 1975).

In firms, privatization of free attributes in agency relationships means making the agent residual claimant of the employment contract to the extent of his (or her) information advantage. Barzel (1989, p. 61) states as an empirical hypothesis: 'The central principle underlying an organization is that the greater is the inclination of a transactor to affect the mean outcome, the greater is the claim on the residual the transactor will assume'. Accordingly, an efficient solution to the incentive problems described above would be to conclude a fixed-rent contract where the agent pays a predetermined amount of money to the principal while himself appropriating the residual claim to his effort. In this case, the agent, even in absence of monitoring, can

be expected to use the employer's rights to his effort and to firm equipment as if he himself were the principal. Alchian and Demsetz's monitor (Alchian and Demsetz, 1972) simply becomes superfluous. The validity of this result, however, presupposes equal inclinations to risk of both the principal and the agent. Otherwise, a conflict would arise between efficiency and Pareto-optimal risk allocation. This conflict is maximal if, as it seems reasonable to suppose, employees are more risk-averse than their employers. Optimal risk allocation, then, indicates a fixed-wage labour contract, with the agent's disincentive to work being maximal. Hence a simultaneous solution to the motivation and the risk allocation problem – which is the main object of normative principal–agent theory – is impossible (see, for instance, Shavell, 1979, pp. 59f.).

Evolution of property rights

The distinction between 'regulation' and 'privatization' is made mainly for analytical reasons. In many cases, owing to the physical nature of the common-pool resource, only one of the two policies will be available. In other cases, the choice between the alternatives will be gradual rather than binary, with an infinite set of intermediate solutions combining elements of each strategy. Which of the solutions, 'regulation' or 'privatization', will be chosen and to what degree?

The cost–benefit models mentioned above, according to which individuals engage in defining and enforcing exclusive rights if their net gains are positive, imply, other things being equal, that changes in the current definition of rights must be due to a change of exogenous factors such as the development of new technologies, the opening of new markets or a change in the costs of property rights definition. Demsetz (1967) employed this approach to explain the sudden emergence of private property rights to land among Indian hunters on the Labrador peninsula, triggered by an exogenous increase in the value of beaver furs that significantly raised the hunting intensity and consequently led to an overexploitation of the commonly used land. As Demsetz argues, exclusive rights allowed the hunters to avoid this common-property problem. Using a similar model, North and Thomas (1973) explained the development of private property in medieval western Europe as a consequence of population growth which substantially increased the demand for goods and, hence, for exclusive rights. Anderson and Hill (1975) argued in the same vein that, in the nineteenth century, ranchers in the American West began to devote more resources to defining and enforcing private entitlements to land as soon as settlement became denser and land values increased and when the introduction of barbed wire significantly decreased the costs of fencing farm land. Field (1989) has shown that historical development is not a one-way street from common to private property but may take the other direction as well.

According to his study, the ultimate direction will depend on a calculus of exclusion costs and costs of internal governance ('transactions costs') which in turn are determined by the number and size of equal plots (commons). When marginal governance costs exceed marginal exclusion costs it will pay the users of a common-pool resource to reduce the size of plots and to adopt the 'privatization' strategy; otherwise 'regulation' of the commonly used asset will be optimal.

The conclusions drawn from Field's model seem to be supported by many empirical studies on the common-pool problem. In many cases the costs of excluding potential free-riders from using a commonly used resource may be prohibitive owing to physical attributes of the asset. Using a variant of Field's model, Eggertsson (1993) explained the long-standing use of common-property arrangements for Icelandic mountain pastures (*affrettir*) as a result of relatively high exclusion costs for individual plots in comparison to the costs of internal governance. Similarly, dividing up inshore fishing grounds or even an entire ocean may be prohibitively costly. The fishing grounds can principally be divided into different areas of private property. This, however, does not resolve the problem, for no fish remains in one place for long and the productivity of fishing areas varies frequently. In these cases management costs may fall short of exclusion costs, and regulated common-property arrangements may prevail. The founding of exclusive 'harbour gangs' (Acheson, 1993) that control access to fishing territories and regulate internally the harvesting efforts of their members may be a much more effective answer to the common-pool problem than devising private property rights. A clever method of regulation approximating the advantages of private entitlements was introduced by a cooperative of fishermen at Alanya, Turkey, which assigns by lot the fishing locations available on a yearly basis (Berkes, 1986). The random allocation of fishing grounds not only temporarily assigns to each individual the residual claim to a certain territory, but also reduces total management costs by creating a 'veil of uncertainty' that induces rights recognition even by holders of poor spots, who in later periods may find themselves in the position of legal users of more productive fishing areas.

Cost–benefit models of property rights evolution can be expected to work well when individually defined property rights are tolerated by the government. Otherwise, a fuller explanation of rights evolution would be necessary which, according to Eggertsson (1990, ch. 8), would require considering the political process as the 'supply-side' determinant of rights emergence, including rent-seeking activities of private interest groups. This, however, is not necessarily an extension of the property rights approach if one takes into account that government behaviour and rent seeking of interest groups can in turn be explained in terms of common-property arrangements. On the one hand, governmental officials, by their power to tax, are able to use the

incomes and fortunes of the governed as if they were common property. By presenting themselves as benevolent donators of public means, politicians may maximize their electoral votes. This may, on the other hand, induce private interest organizations to enter into rent-seeking competition for parts of the public budget which is 'public' only as long as it is not appropriated by someone. Thus, in principle, the above 'demand-side' models of rights emergence should be applicable to the 'supply side' as well. In accordance with the economic theory of politics, these models predict that vote-maximizing governments will tend to close more fully budget access to rent seekers and hence stop budget growth if the net gains are positive from the median voter's point of view.

Cost–benefit explanations of rights evolution may nevertheless fail to be adequate because of their lack of institutional embeddedness. Prisoner's dilemma considerations or Olson's theory of group behaviour indicate that the existence of gains from cooperation is merely a necessary condition for the evolution of property rights, whereas the viability of institutional settings under which they can be appropriated is the sufficient condition. As recent research on common-pool allocation suggests (Libecap, 1989; Ostrom, 1990; Anderson and Simmons, 1993; Ostrom et al., 1994), the key variable to self-organized solutions to the commons problem seems to be a relatively small number of repeatedly interacting appropriators. In such a closed setting, individuals are often able to resolve problems that, from a theoretical point of view, seem to be insurmountable at first sight. Overcoming the common-property problem, then, does not necessarily require an external 'Leviathan' to monitor rule compliance and to sanction free-ridership. Common owners may economize on governance costs by peer monitoring which in small groups is only a byproduct of the appropriators' strong motivation to use the resource efficiently (Ostrom, 1990, p. 95). In sharp contrast to conventional wisdom, under certain conditions 'covenants, without the sword' (Hobbes, 1914, p. 87), are more than mere words. As shown by Ostrom et al. (1993, 1994), even in 'one-shot games', communication alone may lead to substantial improvements in joint outcomes of resources held in common property.

References

Acheson, J.M. (1993), 'Capturing the commons: legal and illegal strategies', in Anderson and Simmons (eds), pp. 69–83.

Alchian, A.A. (1977), 'Some economics of property rights', *Economic Forces at Work*, Indianapolis: Liberty, pp. 127–49.

Alchian, A.A. and H. Demsetz (1972), 'Production, information costs and economic organizations', *American Economic Review*, **62**, 777–95.

Alchian, A.A. and H. Demsetz (1973), 'The property rights paradigm', *Journal of Economic History*, **33**, 16–27.

Anderson, T.L. and P.J. Hill (1975), 'The evolution of property rights: a study of the American West', *Journal of Law and Economics*, **18**, 163–79.

Anderson, T.L. and R.T. Simmons (eds) (1993), *The Political Economy of Customs and Culture: Informal Solutions to the Commons Problem*, Lanham, MD: Rowman & Littlefield.
Barzel, Y. (1974), 'A theory of rationing by waiting', *Journal of Law and Economics*, **17**, 73–95.
Barzel, Y. (1989), *Economic Analysis of Property Rights*, Cambridge: Cambridge University Press.
Berkes, F. (1986), 'Marine inshore fishery management in Turkey', in *Proceedings of the Conference on Common Property Resource Management*, National Research Council, Washington, DC: National Academy Press, pp. 63–83.
Bonus, H. (1978), 'Verzauberte Dörfer, oder: Solidarität, Ungleichheit und Zwang' ['Enchanted villages, or: solidarity, inequality and coercion'], *ORDO*, **29**, 49–82.
Buchanan, J.M. (1975), *The Limits of Liberty – Between Anarchy and Leviathan*, Chicago: University of Chicago Press.
Bush, W.C. and L.S. Mayer (1974), 'Some implications of anarchy for the distribution of property', *Journal of Economic Theory*, **8**, 401–12.
Cheung, S.N.S. (1980), 'Rose Bowl vs. Hong Kong: the economics of seat pricing', *Research in Law and Economics*, Supplement 1, 27–49.
Coase, R.H. (1960), 'The problem of social cost', *Journal of Law and Economics*, **3**, 1–44.
DeAlessi, L. (1980), 'The economics of property rights: a review of the evidence', *Research in Law and Economics*, **2**, 1–47.
Demsetz, H. (1964), 'The exchange and enforcement of property rights', *Journal of Law and Economics*, **3**, 11–26.
Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57**, 347–59.
Eggertsson, T. (1990), *Economic Behavior and Institutions*, Cambridge, MA: Cambridge University Press.
Eggertsson, T. (1993), 'Analyzing institutional successes and failures: a millennium of common mountain pastures in Iceland', in Anderson and Simmons (eds), pp. 109–26.
Field, B.C. (1989), 'The evolution of property rights', *Kyklos*, **42**, 319–45.
Furubotn, E.G. and S. Pejovich (1972), 'Property rights and economic theory: a survey of recent literature', *Journal of Economic Literature*, **10**, 1137–62.
Furubotn, E.G. and S. Pejovich (1974), 'Introduction: the new property rights literature', in E.G. Furubotn and S. Pejovich (eds), *The Economics of Property Rights*, Cambridge, MA: Ballinger, pp. 1–9.
Gordon, H.S. (1954), 'An economic theory of a common property resource', *Journal of Political Economy*, **62**, 124–42.
Grossman, H.I. and M. Kim (1995), 'Swords or plowshares? a theory of the security of claims to property', *Journal of Political Economy*, **103**, 1275–88.
Hardin, G. (1968), 'The tragedy of the commons', *Science*, **162**, 1243–8.
Hobbes, T. (1914), *Leviathan*, edited with an introduction by A.D. Lindsay; reprint 1970, London: Everyman's Library.
Hobbes, T. (1949), *De Cive or The Citizen*, edited with an introduction by Sterling P. Lamprecht, New York: Appleton-Century-Crofts; reprint 1982, Westport, CT: Greenwood Press.
Jensen, M.C. and W.H. Meckling (1976), 'Theory of the firm: managerial behavior, agency-costs and ownership structure', *Journal of Financial Economics*, **3**, 305–60.
Libecap, G.D. (1989), *Contracting for Property Rights*, Cambridge: Cambridge University Press.
Musgrave, R.A. (1987), 'Merit goods', in J. Eatwell, M. Milgate and P. Newan (eds), *The New Palgrave: A Dictionary of Economics*, Vol. 3, London: Macmillan, pp. 452–3.
North, D.C. and R.P. Thomas (1973), *The Rise of the Western World: A New Economic History*, Cambridge: Cambridge University Press.
Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.
Ostrom, E., R. Gardner and J. Walker (1994), *Rules, Games and Common-Pool Resources*, Ann Arbor, MI: University of Michigan Press.

Ostrom, E., J. Walker and R. Gardner (1993), 'Covenants with and without a sword: self-governance is possible', in Anderson and Simmons (eds), pp. 127–56.
Richter, R. and E. Furubotn (1996), *Neue Institutionenökonomik – Eine Einführung und kritische Würdigung* [New Institutional Economics – an Introduction and Critical Assessment], Tübingen: J.C.B. Mohr.
Shavell, S. (1979), 'Risk sharing and incentives in the principal and agent relationships', *Bell Journal of Economics*, **10**, 55–73.
Tietzel, M. (1981), 'Die Ökonomie der Property Rights: Ein Überblick' ['The economics of property rights: a survey'], *Zeitschrift für Wirtschaftspolitik*, **30**, 207–43.
Tietzel, M. (2001), 'In praise of the commons: another case study', in *European Journal of Law and Economics*, **12**, 159–71.
Tietzel, M. and C. Müller (2002), 'Merit goods from a constitutional perspective', in H.G. Brennan, H. Kliemt and R. Tollison (eds), *Method and Morals in Constitutional Economics. Essays in Honor of James M. Buchanan*, Berlin: Springer, pp. 375–400.

3 Legal change in economic analysis

John N. Drobak and Douglass C. North

One school of law and economics analyses legal problems by using economic principles. Scholars working in the new institutional economics reverse that process and incorporate legal analysis in their explanation of economic events. The new institutionalists believe that economic growth cannot be understood with neoclassical theory alone. Neoclassical theory can be a powerful explanatory and predictive tool, but it is also a static theory that often oversimplifies, sometimes erroneously, the dynamic world. In an attempt to bring order to this uncertain and constantly changing world, human beings have used institutions – the rules of the game of a society – to structure human interaction. Institutions provide the framework of incentives that shape economic, political and social organization. They provide a foundation for the formation of property rights. Institutions affect economic performance by determining, together with the technology employed, the transaction and transformation costs that make up the total costs of production. Informal institutions include such things as norms of behaviour, codes of conduct and business conventions. Many formal institutions, such as constitutions, statutes, regulations and decisions of courts, are legal. Some formal institutions are created by non-governmental organizations – religious laws, corporate rules of self-governance and use restrictions imposed by residential groups, for example (see North, 1990).

An institution is defined by how it is enforced, as well as by the written or understood terms of the rule. Enforcement can be carried out by third parties (government enforcement, social ostracism), second parties (retaliation) or the first party (self-imposed conduct). Judicial or bureaucratic enforcement of written rules can give a clearer (and sometimes different) meaning to a written rule. The history of the enforcement of the competition laws in the United States and in the European Union illustrates the importance of understanding the law as applied, not just the law as written. The operative language of the Sherman Act, the principal competition law in the United States, is brief: it proscribes 'contracts, combinations or conspiracies … in restraint of trade' and makes 'monopolization' illegal. Over 100 years of judicial application of those two brief phrases has created a voluminous, intricate set of competitive rules. The European Union has enacted comparable provisions dealing with anti-competitive practices. Articles 85 and 86 of the Treaty of Rome deal respectively with anti-competitive agreements and with dominant

firms having market power. This structure is like that in the Sherman Act. The wording of the two articles, containing lists of prohibited practices, is much more detailed than the Sherman Act, although it is similar to the anti-competitive rules fashioned by the courts in the United States out of the Sherman Act. Notwithstanding these striking similarities, the competition laws in the United States and in the European Union are very different, primarily because the European laws are applied to strengthen the common market in the tradition of the European social market economies.

The law facilitates economic growth in many different ways. As Ronald Coase (1960) demonstrated, the law can provide a baseline for commercial transactions by establishing a default rule, applicable unless the parties specify otherwise. This enables the parties to know *ex ante* their relative rights and obligations; if these are less desirable, they can contract around them. This ability to rearrange the terms of the default rule makes, in one important sense, the terms of the rule unimportant. All that is important is that a default rule exist. There are two instances, however, when the terms of the default rule do matter. The clarity and transparency of any law is relevant to the transaction costs associated with using the law. The clearer and more transparent the law, the less costly it will be to use. Furthermore, there are times when a default rule will be used, whether from a failure to reach the bargain or from a breakdown in the bargain itself. Then the meaning of the rule may sometimes have a strong effect on economic growth. *Bass* v. *Gregory*, one of the series of nuisance cases discussed by Coase, provides an excellent example of the importance of this.

That case involved a suit by the owner of the Jolly Angler pub to establish the legal right to use a ventilating shaft on a neighbour's land. For over 40 years, the pub had operated a brewery in its cellar, venting the production process through a shaft that connected into an old well located in the neighbour's yard. When the neighbour blocked the ventilation through the well, the pub owner sued. The outcome of the lawsuit established a legal rule with both short- and long-term consequences. As Coase (1960, p. 15) saw it, 'The economic problem was to decide which to choose: a lower cost of beer and worsened amenities in adjoining houses or a higher cost of beer and improved amenities'. His rule for making the choice was to maximize the value of production from both parcels. In *Bass* v. *Gregory* the court ruled that the pub had the right to vent its brewing operations through the well. Assuming that the brewery, with its established ventilating system, added more to productive output than the neighbour lost, the value of production would have been maximized. This short-term consequence of the decision was a good economic result in terms of Coase's criteria. Of course, if the court had reached the poorer economic outcome by ruling for the neighbour, the parties could still have reached the desirable result through the neighbour's sale of

the right to the pub owner, assuming no impediments to bargaining. Thus the short-term goal of maximizing productivity can be reached either way.

The long-term consequences of the court's decision – the downstream economic effect – can have an even bigger impact on economic growth. Consider the incentives for future conduct resulting from the decision in *Bass* v. *Gregory*. If the defendant had been the pub's neighbour for years, the court was enforcing a relationship the parties had created over time, in effect enforcing settled expectations. That is an important, albeit not surprising, principle worthy of reinforcement. Suppose the neighbour, instead, was a recent purchaser of the land who was surprised and bothered by the pub's exhaust gases. The court's decision has important lessons for this type of real estate purchaser. First, the buyer must seek a remedy from the seller of the property, not from the pub or a similarly situated neighbour. Second, the buyer is obligated to inspect the property, inquiring about the use of visible aspects of the property, like the well. This prophylactic rule, designed to encourage buyers to prevent problems as in *Bass* v. *Gregory* from ever arising, has positive downstream economic effects. One could argue that the court was wrong, because a ruling for the neighbour would have created an incentive for all easements (including the right to use someone else's property for ventilation or exhaust purposes) to be reduced to writing and then recorded on the public records. This could be said to reduce transaction costs in the aggregate since a buyer need only rely on the land records and not make a physical investigation of the property. A ruling for the neighbour, however, would have actually raised aggregate transaction costs because it would compel the unnecessary recording of untold minor transactions. Since economic growth is furthered by the correct incentives, a court should consider the downstream economic effect as one of the primary factors driving its decision. There will naturally be cases where this goal will not be helpful, as where the downstream effect is too uncertain or where the costs of the competing outcomes are indeterminate. These were not problems in *Bass* v. *Gregory*. Judging the result by both the short-term productivity goal and the downstream economic effects, the court's ruling for the pub owner was the one that did the most to further economic growth.

The nuisance cases discussed by Coase are but one small part of a large body of commercial law that provides the framework for a market system. Some types of rules for exchange are important even in primitive economies. As economies became more complex, with increased specialization and division of labour, bringing more and complex transactions, commercial law had to become more specialized. A modern market-based economy depends on bodies of law dealing with property, contracts, debtor–creditor and bankruptcy. The recent experiences of the formerly communist countries in Central Europe in their attempts to move to a market system reinforce the importance

of some type of securities law governing emerging capital markets. Economic growth also depends upon a criminal law that provides for security of both people and property. It is extremely difficult, perhaps impossible, for a sustained economic growth to occur in countries where extortion, theft and other violent crimes undermine market transactions. The warlord economies of some African countries and the disruption of market activities in Russia by the mafia are examples of this.

Both commercial and criminal law regulate relations between individuals. Sustained economic growth also depends upon legal institutions that constrain government. These need to take two forms. First, it is important that the law somehow prevent governments from acting as a mafia and extracting wealth for the rulers, without any consideration of the economic well-being of citizens. Democracy is one solution to this problem, since citizens can vote out rulers who disregard their economic well-being. The constitutional structure of government also advances this interest, with checks and balances between various parts of government and allocation of various powers to different branches of government serving to limit any attempts by rulers to govern only for their self-interest. Second, the history of economic growth in the Western economies has shown that market participants need to be able to trust the promises of governments (North and Weingast, 1989). There are many incentives for a government to renege on its commitments. This is especially true for financial obligations, where the short-term financial gain to the state (or to certain groups of voters) will be seen as outweighing uncertain long-term consequences. In the United States, for example, the contracts clause in the Constitution prohibits state and local governments from repudiating debt obligation (Drobak, 1997). Most nations prohibit outright expropriation of property. Governments can also 'take' property through various kinds of regulations that fall short of expropriation. The United States and Germany, to cite two examples, have constitutional provisions limiting the taking of private property through excessive regulation (Epstein, 1985; Kommers, 1997, pp. 241–97). These kinds of provisions deter government actions that would otherwise undermine the credibility of the government and harm the workings of the market.

The effectiveness of all these laws hinges on an unbiased, honest judiciary and bureaucracy and on a dispute resolution system that is relatively efficient. Confidence of the market participants in the judiciary and the bureaucracy is crucial to the smooth running of a market system and to economic growth. All of these legal institutions, and the means of enforcement, make up the 'rule of law', an essential component of sustained economic growth. Real economies cannot operate without these legal underpinnings. Likewise, realistic economic analysis has to incorporate legal institutions. The merging of law and neoclassical theory has been an important contribution of the new institutional economics.

References

Articles 85–6, Treaty Establishing European Economic Community, entered in force, 1 January 1958, 298 U.N.T.S. 11.

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

Drobak, John N. (1997), 'Credible commitment in the United States: substantive and structural limits on the avoidance of public debt', in John N. Drobak and John V.C. Nye (eds), *The Frontiers of the New Institutional Economics*, San Diego and London: Academic Press, pp. 247–66.

Epstein, Richard A. (1985), *Takings: Private Property and the Power of Eminent Domain*, Cambridge, MA and London: Harvard University Press.

Kommers, Donald P. (1997), *The Constitutional Jurisprudence of the Federal Republic of Germany*, 2nd edn, Durham, NC and London: Duke University Press.

North, Douglass C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

North, Douglass C. and Barry R. Weingast (1989), 'Constitutions and commitment: evolutions of the institutions of public choice in 17th century England', *Journal of Economic History*, **59** (4), December, 803–32.

Sherman Act, 15 U.S.C. §§ 1, 2 (1997).

4 Positive, normative and functional schools in law and economics

Francesco Parisi

Various important methodological questions have accompanied the growth and evolution of law and economics. Economists and jurists alike have debated the appropriate role of economic analysis in the institutional design of lawmaking and the limits of methods of evaluation of social preferences and aggregate welfare in policy analysis. In many respects, these methodological debates have contributed to the growing intellectual interest and to the diversification of methodologies in the economic analysis of law.

The origins and the evolved domain of law and economics

Law and economics is probably the most successful example of the recent surge of applied economics into areas that were once regarded as beyond the realm of economic analysis and its study of explicit market transactions. Methodologically, law and economics applies the conceptual apparatus and empirical methods of economics to the study of law.

The origins of modern law and economics

Extensive research has been carried out to identify the historical and antecedents to modern law and economics. Indeed, this volume contains several biographical entries devoted to precursors and early European exponents of the law and economics movement. It is interesting to see that, although the recognition of law and economics as an independent field of research is the result of studies carried out in the United States after the 1970s, most of the precursors can be found in Europe. Notable antecedents to law and economics include the work of Adam Smith on the economic effects of legislation (1776), and Jeremy Bentham's theory of legislation and utilitarianism (1782 and 1789).

In the United States, it was not until the mid-twentieth century – through the work of Henry Simon, Aaron Director, Henry Manne, George Stigler, Armen Alchian, Gordon Tullock and others – that the links between law and economics became an object of serious academic pursuit. The regulation of business and economic law fell within the natural interest of the first American scholars of law and economics. Early research concentrated on areas related to corporate law, tax law and competition law. In so doing, the first

generation of law and economics scholars paralleled the efforts of other economists, trying to explain the functioning of explicit economic markets and the impact of alternative legal constraints, such as taxes and regulation, on the market.

In the 1960s the pioneering work of Ronald Coase and Guido Calabresi brought to light the pervasive bearing of economics in all areas of the law. The methodological breakthrough occasioned by Coase and Calabresi allowed immediate extensions to the areas of tort, property and contract. The analytical power of their work was not confined to these fields, however, and subsequent law and economics contributions demonstrate the explanatory and analytical reach of its methodology in a number of other areas of the law.

A difference in approach is detectable between the law and economics contributions of the early 1960s and those that followed in the 1970s. While the earlier studies appraise the effects of legal rules on the normal functioning of the economic system (that is, they consider the impact of legal rules on the market equilibrium), the subsequent generation of studies utilizes economic analysis to achieve a better understanding of the legal system. Indeed, in the 1970s a number of important applications of economics to law gradually exposed the economic structure of basically every aspect of a legal system: from its origin and evolution, to its substantive, procedural and constitutional rules.

Despite some resistance to the application of economics to non-market behaviour, the important bonds between legal and economic analysis, as well as the social significance of the object of study, were in themselves a guarantee of success and fruitfulness for law and economics.

An important ingredient in the success of law and economics research has come from the establishment of specialized journals. The first such journal, the *Journal of Law and Economics*, appeared in 1958 at the University of Chicago. Its first editor, Aaron Director, should be credited for this important initiative, successfully continued by Ronald Coase. Other journals emerged in the following years: in 1972, the *Journal of Legal Studies*, also housed at the University of Chicago, was founded under the editorship of Richard Posner; in 1979, *Research in Law and Economics*, under the editorship of Richard Zerbe, Jr; in 1981, the *International Review of Law and Economics* was established in the United Kingdom under the editorship of Charles Rowley and Anthony Ogus (later joined by Robert Cooter and Daniel Rubinfeld); in 1982, the *Supreme Court Economic Review*, under the editorship of Peter Aranson (later joined by Harold Demsetz and Ernest Gellhorn); in 1985, the *Journal of Law, Economics and Organization*, under the editorship of Jerry Mashaw and Oliver Williamson (later joined by Roberta Romano); in 1994, the *European Journal of Law and Economics* was launched under the editorial direction of Jürgen Backhaus and Frank Stephen; in 1999, the *Ameri-*

can Law and Economics Review, under the editorship of Orley Ashenfelter and Richard Posner; and, most recently, in 2004 the *Journal of Empirical Legal Studies* under the editorship of Theodore Eisenberg, Jeffrey J. Rachlinski, Steward J. Schwab, and Martin T. Wells; and in 2005 the *Review of Law and Economics* under the editorship of Robert Cooter, Ben Depoorter, Lewis Kornhauser, Gerrit De Geest, Nuno Garoupa, and Francesco Parisi. These specialized journals provided – and continue to provide – an extremely valuable forum for the study of the economic structure of law.

In many respects, the impact of law and economics has exceeded its planned ambitions. One effect of the incorporation of economics into the study of law was to irreversibly transform traditional legal methodology. Legal rules began to be studied as a working system – a clear change from the Langdellian tradition, which had relied almost exclusively on the self-contained framework of case analysis and classification, viewing law as little more than a filing system. Economics provided the analytical rigour necessary for the study of the vast body of legal rules present in a modern legal system. This intellectual revolution came at an appropriate time, when legal academia was actively searching for a tool that permitted critical appraisal of the law, rather than merely strengthening the dogmatic consistencies of the system.

The marriage of law and economics has also affected the economic profession, contributing to the expansion of the original domain of microeconomic analysis – the study of individual and organizational choices in the market – to the study and understanding of other institutions and non-market phenomena.

The evolved domain of law and economics

Despite the powerful analytical reach of economics, it was clear from the outset that the economist's competence in the evaluation of legal issues was limited. While the economist's perspective could prove crucial for the positive analysis of the efficiency of alternative legal rules and the study of the effects of alternative rules on the distribution of wealth and income, economists generally recognized the limits of their role in providing normative prescriptions for social change or legal reform.

Recognition of the positive nature of the economic analysis of law was not sufficient to dispel the many misunderstandings and controversies in legal academia engendered by the law and economics movement's methodological revolution. As Coase (1978) indicated, the cohesiveness of economic techniques makes it possible for economics to move successfully into another field, such as law, and dominate it intellectually. But methodological differences played an important part in the uneasy marriage between law and economics. The Popperian methodology of positive science was in many respects at odds with the existing paradigms of legal analysis. Rowley (1981)

characterizes such differences, observing that positive economics follow the Popperian approach, whereby testable hypotheses (or models) are derived by means of logical deduction and are then tested empirically. Anglo-American legal analysis, on the other hand, is generally inductive: lawyers use individual judgments to construct a general premise of law. Much work has been done in law and economics despite these methodological differences, with a reciprocal enrichment of the analytical tools of both disciplines.

Law and economics relies on the standard economic assumption that individuals are rational maximizers, and studies the role of law as a means for changing the relative prices attached to alternative individual actions. Under this approach, a change in the rule of law will affect human behaviour by altering the relative price structure – and thus the constraint – of the optimization problem. Wealth maximization, serving as a paradigm for the analysis of law, can thus be promoted or constrained by legal rules.

The early years of law and economics were characterized by the uneasiness of some traditional legal scholars in the acceptance of the notion of wealth maximization as an ancillary paradigm of justice. Although most of the differences gradually proved to be largely verbal – and many others were dispelled by the gradual acceptance of a distinction between paradigms of utility maximization and wealth maximization – two objections continue to affect the lines of the debate. The first relates to the need for specifying an initial set of individual entitlements or rights, as a necessary prerequisite for operationalizing wealth maximization. The second springs from the theoretical difficulty of defining the proper role of efficiency as an ingredient of justice, *vis-à-vis* other social goals.

In his well-known defence of wealth maximization as a guide for judicial action, Posner (1985) distinguishes wealth or expected utility from market prices. While market prices may not always fully reflect idiosyncratic valuations, they avoid an undertaking of interpersonal utility comparisons, with the opportunity for *ex post* rationalization of positions taken on emotional grounds. Posner's view is sympathetic to the premises of a property rights approach to legal relationships, and he stresses the importance of an initial distribution of property rights prior to any calculation of wealth maximization. His paradigm of wealth maximization serves as a common denominator for both utilitarian and individualist perspectives. By combining elements of both, Posner provides a theory of wealth maximization that comes closer to a consensus political philosophy than does any other overarching political principle.

In contrast, Calabresi (1980) claims that an increase in wealth cannot constitute social improvement unless it furthers some other goal, such as utility or equality. Denying that one can trade off efficiency against justice, he argues instead that efficiency and distribution are ingredients of justice, which

is a goal of a different order than either of these ingredients. Calabresi thus defends law and economics as a worthy examination of certain ingredients of justice, rather than a direct examination of justice itself.

The intellectual resistance that has characterized the birth of law and economics can only be temporary. Both legal practitioners and policy makers are becoming aware of the important role of economic analysis in their discipline, and we have already mentioned notable contributions to mainstream economic theory from lawyers in the law and economics movement. Likewise, as Coase (1978) noted, economists have come to realize that the other social sciences are so intertwined with the economic system as to be part of the system itself. For this reason, law and economics can no longer be appraised as a branch of applied microeconomics; rather, it must be seen as contributing to a better understanding of the economic system itself. The study of the effects of other social sciences on the economic system will, Coase predicts, become a permanent part of the field of economics.

Coase also examines the reasons for the movement of economists into the other social sciences, and attempts to predict the future of this phenomenon. Groups of scholars are bound together by common techniques of analysis, a common theory or approach to the subject, and/or a common subject matter. In the short run, Coase maintains, one group's techniques of analysis may give it such advantages that it is able to move successfully into another field and maybe even dominate it. In the long run, however, the subject matter tends to be the dominant cohesive force. While the analytical techniques employed by economists – such as linear programming, quantitative methods and cost–benefit analysis – may recently have aided the entry of economists into the other social sciences, Coase predicts that such a movement can only be temporary. After all, the wisdom possessed by economists, once its value is recognized, will be acquired by some of the practitioners in these other fields (as is happening in the field of law).

As the domain of law and economics continues to expand, its perspective on methodological issues has not been stagnant. While this chapter emphasizes the wide range of substantive applications, some degree of controversy still surrounds several of the methodological, normative and philosophical underpinnings of the economic approach to law. Most of the ideological differences tend to lose significance because their operational paradigms often lead to analogous results when applied to real cases. Some scholars, however, perceive that the current state of law and economics as comparable to the state of economics prior to the advent of public choice theory, in so far as an understanding of 'political failures' was missing from the study of market failures (Buchanan, 1974; Rowley, 1989) Public choice may indeed inject a sceptical – and at times disruptive – perspective into the more elegant and simple framework of neoclassical economics, but this added element

may well be necessary to better understand a complex reality. In a way, the systematic incorporation of public choice theory into the economic approach to law has contributed to bridging the conflicting normative perspectives in law and economics, at least by bringing the debate onto the more solid ground of collective choice theory.

Economics is a powerful tool for the analysis of law. If humans are rational maximizers of their utility, wealth or well-being then they respond rationally to changes in exogenous constraints, such as laws. This rationality assumption provides the basic foundation for much law and economics literature. Building upon the standard economic assumption that individuals are rational maximizers, the sophisticated tools of price theory become a useful aid in the study and choice of legal rules (Cooter, 1984). While there is much consensus on the value of economic theory in the study of legal rules, important methodological differences arise with respect to the choice of the appropriate instruments of legal analysis and the choice of method for evaluation of social preferences. I shall briefly discuss these methodological issues in turn.

Schools and intellectual perspectives in law and economics

Most practitioners of law and economics believe that there is an important common ground that unifies all scholars in the discipline, regardless of their ideological creed: a search for new insights in the law by applying economic concepts and theories (MacKaay, 2000). Despite this common statement of purpose, various schools of law and economics can be identified, each with an elaborate research programme and a distinct methodological approach.

The Chicago and Yale schools: positive versus normative approaches to law and economics

During the early period of the discipline, law and economics scholarship was labelled 'Chicago' or 'Yale' style. These labels made reference to the respective positive or normative approach utilized by each school. The origins of the Chicago and Yale schools of law and economics are attributable to the early work of a handful of scholars, including the pioneering work of Coase and Calabresi in the early 1960s.

At this point, methodological differences came to the surface with substantive practical differences. The Chicago school laid most of its foundations on the work carried out by Posner in the 1970s. An important premise of the Chicago approach to law and economics is the idea that the common law is the result of an effort – conscious or not – to induce efficient outcomes. This premise is known as the efficiency of the common law hypothesis. According to this hypothesis, first intimated by Coase (1960), and later systematized and greatly extended by Ehrlich and Posner (1974), Rubin (1977) and Priest (1977), common law rules attempt to allocate resources in either a Pareto or

Kaldor–Hicks efficient manner. Further, according to the positive school, 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.

The primary hypothesis advanced by positive economic analysis of law is thus the notion that efficiency is the predominant factor shaping the rules, procedures and institutions of the common law. Posner contends that efficiency is a defensible criterion in the context of judicial decision making because 'justice' considerations – on the content of which there is no academic or political consensus – introduce unacceptable ambiguity into the judicial process.

In arguing for positive use of economics, Ehrich and Posner (1974) is not denying the existence of valuable normative law and economics applications. In fact, law and economics often has many objective things to say that will affect one's normative analysis of a policy.[1]

Despite the powerful analytical reach of economic analysis, Chicago scholars acknowledged from the outset that the economist's competence in the evaluation of legal issues was limited. While the economist's perspective could prove crucial for the positive analysis of the efficiency of alternative legal rules and the study of the effects of alternative rules on the distribution of wealth and income, Chicago-style economists generally recognized the limits of their role in providing normative prescriptions for social change or legal reform.

On the contrary, the Yale school of law and economics, often described as the 'normative' school, believes that there is a greater need for legal intervention in order to correct for pervasive forms of market failure.[2] Distributional concerns are central to the Yale-style literature. The overall philosophy of this group is often presented as more value tainted and more prone to policy intervention than the Chicago law and economics school.

Unlike its Chicago counterpart, the Yale school has attracted liberal practitioners who employ the methodology of the Chicago school but push it to formulate normative propositions on what the law ought to be like (MacKaay, 2000). Given the overriding need to pursue justice and fairness in distribution through the legal system, most Yale-style scholars would suggest that efficiency, as defined by the Chicago school, could never be the ultimate end of a legal system.

The Virginia school: the functional approach and the return to normative individualism

In recent years, a new generation of literature – developed at the interface of law, economics and public choice theory – pushes the methodological bound-

aries of economic analysis of law. The resulting approach is in many respects functional in its ultimate mission, cutting across the positive and normative distinction and unveiling the promises and pitfalls of both the normative and the positive alternatives. This approach to legal analysis has the potential of shedding light on the traditional conception of lawmaking, suggesting that the comparative evaluation of alternative sources of law requires an appropriate analysis of the incentive structure in the originating environment. This line of research is attentive to the identification of political failures in the formation of law, stressing the importance of market-like mechanisms in the creation and selection of legal rules.

The functional approach to law and economics is still in its initial phase of development and far from a point of theoretical maturity, but this approach is unquestionably successful in raising some crucial questions regarding the difficult link between individual preferences and social outcomes, with an emphasis on institutional mechanism design and individual choice. The resulting approach is quite sceptical of both the normative and the positive alternatives.[3] Public choice theory provides strong methodological foundations for the functional school of law and economics: the systematic incorporation of the findings of public choice theory into the economic analysis of law may serve to bridge the conflicting normative perspectives in law and economics, at least by bringing the debate onto the more solid ground of collective choice theory.

The functional approach is wary of the generalized efficiency hypotheses espoused by the positive school. In this respect, the functionalists share some of this scepticism of 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 sceptical of a general efficiency hypothesis when applied to sources of the law other than common law (for example, legislation or administrative regulations).

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. Thus, functionalists think it is often dangerous to utilize such tools to design corrective or interventionist policies. 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.

An important premise of the functional approach to law and economics is its reliance on methodological individualism. According to this paradigm of analysis, only individuals choose and act (see, for example, Buchanan, 1990 and the various contributions of the Virginia school of political economy). The functional approach to law and economics is informed by an explicit

recognition that whatever social reality we seek to explain at the aggregate level, ought to be understood as the result of the choices and actions of individual human beings who pursue their goals with an independently formed understanding of the reality that surrounds them (Vanberg, 1994: 1). Normative individualism further postulates that only the judgment of single individuals can provide a relevant benchmark against which the merits of alternative rules can be evaluated.

The findings of public choice theory, while supporting much of the traditional wisdom, pose several challenges to neoclassical law and economics. In spite of the sophisticated mathematical techniques of economic analysis, judges and policy makers in many situations still lack the expertise and methods for evaluating the efficiency of alternative legal rules. Courts and policy makers should thus undertake a functional analysis. Such an analysis requires them to first inquire into the incentives underlying the legal or social structure that generated the legal rule, rather than directly attempting to weigh the costs and benefits of individual rules.[4] In this way, the functionalist approach to law and economics 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.

Pareto, Bentham and Rawls: the dilemma of preference aggregation

The need to make comparative evaluations between different rules motivates much of law and economics. Consequently, the second methodological problem in law and economics concerns the choice of criteria for carrying out such comparative analysis. In practical terms, this problem concerns the method of aggregation of individual preferences into social preferences. This problem is not unique to law and economics. It is part of a much larger methodological debate in economic philosophy and welfare economics.

Already in the late nineteenth century, F.Y. Edgeworth (1881: 7–8) stated the moral dilemma of social welfare analysis, observing that a moral calculus should proceed with a comparative evaluation of 'the happiness of one person with the happiness of another. … Such comparison can no longer be shirked, if there is to be any systematic morality at all'. The problem obviously arises from the fact that economists do not have any reliable method for measuring individuals' utility, let alone make interpersonal comparisons of utility.

Economic analysis generally utilizes one of the three fundamental criteria of preference aggregation.

Ordinal preferences and the Pareto criterion

The first criterion of social welfare is largely attributable to Italian economist and sociologist Vilfredo Pareto. The Pareto criterion limits the inquiry to *ordinal* preferences of the relevant individuals. According to Pareto, an optimal allocation is one that maximizes the well-being of one individual relative to the well-being of other individuals being constant.[5] In normal situations, there are several possible solutions that would qualify for such a criterion of social optimality. For example, if the social problem is that of distributing a benefit between two parties, any hypothetical distribution would be Pareto optimal, since there is no possible alternative redistribution that would make one party better off without harming another one.

The Pareto criterion has been criticized for two main reasons: (a) it is *status quo* dependent, in that different results are achieved depending on the choice of the initial allocation; and (b) it only allows *ordinal* evaluation of preferences, since it does not contain any mechanism to induce parties or decision makers to reveal or evaluate *cardinal* preferences (that is, the intensity of preferences). As a result of these shortcomings scholars (for example, Calabresi, 1991), have questioned the usefulness of the Pareto criterion in its applications to law and economics.

Utilitarian tests: Bentham and Kaldor–Hicks

In the nineteenth and early twentieth centuries, economists and philosophers developed welfare paradigms according to which the degree of all affected individuals had to be taken into account in any comparative evaluation of different states of the world. This methodological trend, related to utilitarian philosophy, is best represented by philosophers and jurists such as Bentham (1839) and later economists such as Kaldor (1939) and Hicks (1939), who in different ways formulated criteria of social welfare that accounted for the *cardinal* preferences of individuals.

In *Principles of Morals and Legislation*, Bentham (1789) presents his theory of value and motivation. He suggests that mankind is governed by two masters: 'pain' and 'pleasure'. The two provide the fundamental motivation for human action. Bentham notes that not all individuals derive pleasure from the same objects or activities, and not all human sensibilities are the same.[6] Bentham's moral imperative, which has greatly influenced the methodological debate in law and economics, is that policy makers have an obligation to select rules that give 'the greatest happiness to the greatest number'. As pointed out by Kelly (1998: 158) this formulation is quite problematic, since it identifies two maximands (that is, degree of pleasure and number of individuals) without specifying the tradeoff between one and the other. Bentham's utilitarian approach is thus, at best, merely inspirational for policy purposes.

Later economists, including Kaldor (1939), Hicks (1939) and Scitovsky (1941), formulated more rigorous welfare paradigms which avoided the theoretical ambiguities of Bentham's proposition. However, these formulations presented a different set of difficulties in their implementation. The core idea of their approach is that state A is to be preferred to state B if those who gain from the move to A gain enough to compensate those who lose. This is generally known as the Kaldor–Hicks test of potential compensation. It is one of 'potential' compensation because the compensation of the losers is only hypothetical and does not actually need to take place.[7] In practical terms, the Kaldor–Hicks criterion requires a comparison of the gains of one group and the losses of the other group. As long as the gainers gain more than the losers lose, the move is deemed efficient. Mathematically, both the Bentham and the Kaldor–Hicks versions of efficiency are carried out by comparing the aggregate payoffs of the various alternatives and selecting the option that maximizes such summation.

Non-linear social preferences: Nash and Rawls

Other paradigms of social welfare depart from the straight utilitarian approach, suggesting that social welfare maximization requires something more than the maximization of total payoffs for the various members of society. Societies are formed by a network of individual relations and there are some important interpersonal effects that are part of individual utility functions. Additionally, human nature is characterized by diminishing marginal utility, which gives relevance to the distribution of benefits across members of the group.

Imagine two hypothetical regimes: (a) in which all members of society eat a meal a day; and (b) in which only a random one-half of the population eat a double meal while the other unlucky half remains starving. From a Kaldor–Hicks perspective, the two alternatives are not distinguishable from the point of view of efficiency because the total amount of food available remains unchanged. In a Kaldor–Hicks test, those who get a double meal have just enough to compensate the others and thus society should remain indifferent between the two allocational systems. Obviously, this indifference proposition would leave most observers unsatisfied. In the absence of actual compensation, the criterion fails to consider the diminishing marginal benefit of a second meal and the increasing marginal pain of starvation. Likewise, the randomized distribution of meals fails to consider the interpersonal effects of unfair allocations. Fortunate individuals suffer a utility loss by knowing that other individuals are starving while they enjoy a double meal. Because of the diminishing marginal utility of wealth and interpersonal utility effects, from an *ex ante* point of view, no individual would choose allocation system (b), even though the expected return from (b) is equal to the return from (a).

Scholars who try to evaluate the welfare implications of distributional inequalities generally do so by invoking Rawls's (1971)[8] theories of justice or by utilizing Nash's (1950)[9] framework of welfare.

The intuition underlying these criteria of welfare is relatively straightforward: the well-being of a society is judged according to the well-being of its weakest members. The use of an algebraic product to aggregate individual preferences captures that intuition. Like the strength of a chain is determined by the strength of its weakest link, so the chain of products in an algebraic multiplication is heavily affected by the smallest multipliers. Indeed, at the limit, if there is a zero in the chain of products, the entire grand total will collapse to zero. This means that the entire social welfare of a group approaches zero as the utility of one of its members goes to zero.

In the law and economics tradition, these models of social welfare have not enjoyed great popularity. This is not so much for an ideological preconception but rather for a combination of several practical reasons. These include the general tendency to undertake a two-step optimization in the design of policies, and the difficulties of identifying an objective criterion for assessing interpersonal utility and diminishing marginal utility effects. From a methodological point of view, distributional concerns are generally kept separate from the pursuit of efficiency in policy making. Such separation has been rationalized on the basis that the legal system is too costly an instrument for distribution, given the advantage of the tax system for wholesale reallocation of wealth (for example, Kaplow and Shavell, 1994).

Wealth, utility and revealed preferences: the choice of maximand

There is an important methodological question that has openly engaged the attention of prominent law and economics scholars: what should the legal system try to maximize? In this debate, even strict adherents to the instrumentalist view of the law may question whether the objective of the law should be the maximization of aggregate wealth, aggregate utility, or merely provide the conditions for free individual choice.

If the scholars involved in these debates could look at the issue as neutral spectators, consensus could be reached on the idea that the ultimate policy goal is the maximization of human happiness and well-being. But regardless of such an observation, economic analysis of law rarely uses utility-based methods of evaluation. The reason for this is, once again, mostly pragmatic. Unlike wealth (or quantities of physical resources), utility cannot be objectively measured. Furthermore, interpersonal comparisons of utility are impossible, rendering any balancing across groups or individuals largely arbitrary. These limitations make utility maximization unviable for practical policy purposes.

Given the above limitations, following Posner, several practitioners of economic analysis of law have departed from the nineteenth-century utilitarian

ideal of utility maximization.[10] Rather, they have increasingly used a paradigm of wealth maximization. Several scholars in law and economics remain uneasy in accepting the notion of wealth maximization as an ancillary paradigm of justice. Although several of the differences prove to be largely verbal, two objections continue to affect the lines of the debate.

The first objection relates to the need for specifying an initial set of individual entitlements or rights as a necessary prerequisite for operationalizing wealth maximization. In this context, one can think of the various criticisms of wealth maximization by property rights advocates who perceive the social cost of adopting such criterion of adjudication as very high, given wealth maximization's instrumentalist view of individual rights and entitlements. These critics argue that rights have value that must be accounted for outside of how useful they might be to the accumulation of wealth (Buchanan, 1974; Rowley, 1989).

The second objection springs from the theoretical difficulty of defining the proper role of efficiency as an ingredient of justice, *vis-à-vis* other social goals. Legal scholars within the law and economics tradition (see, for example, Calabresi, 1980) have claimed that an increase in wealth cannot constitute social improvement unless it furthers some other social goal, such as utility or equality. Denying that one can trade off efficiency against justice, these scholars argue instead that efficiency and distribution are equally essential elements of justice, which is seen as a goal of a different order than either of its constitutive elements.

The functional school of law and economics provides a third alternative by identifying individual choice and revealed preferences as the fundamental criterion for evaluation. The design of metarules that are aimed at fostering free individual choice by eliminating strategic and transactional impediments to the revelation of true preferences becomes an explicit objective of the functional school. As discussed above, the evaluation of alternative sources of law requires an appropriate analysis of the incentive structure in the originating environment and is aimed at introducing market-like mechanisms in the creation and selection of legal rules, with an emphasis on institutional mechanism design and individual choice. The recent literature on reciprocity (Smith et al., 1998; Fon and Parisi, 2003), social norms and customary law (Parisi, 1998; Cooter, 2000), choice of law (Parisi and Ribstein, 1998; Romano, 1999; Ribstein and O'Hara, 2000), federalism (Ribstein and Kobayashi, 2001) and freedom of contract (Trebilcock, 1994; Buckley, 1999) are examples of the growth and value of functional approaches in law and economics.

Future generations of law and economics scholars should be cognizant of the important methodological debates that have engaged their precursors, taking full advantage of the insights developed by the different methodological traditions when appraising legal rules and institutions.

Notes

1. Ehrich and Posner (1974) offers crime as an example. Positive law and economics can help explain and predict how various punishments will affect the behaviour of criminals. It might determine that a certain sanction is more likely to deter a certain crime. While this analysis does not by itself mean that the law should be adopted, it can be used to influence normative analysis on whether the law would be beneficial to society.
2. MacKaay (2000) observes that the Yale school considers market failures to be more pervasive than Chicago scholars are willing to admit. Legal intervention is believed to be the appropriate way of correcting such failures, although it may not succeed in all circumstances.
3. For a brief intellectual history of the three approaches to law and economics, see Posner and Parisi (1998).
4. On this point, see Cooter (1994) introducing the similar idea of structural adjudication of norms.
5. As a corollary, a change to a Pareto superior alternative makes someone better off without making anyone worse off.
6. See Posner (1998) for an interesting discussion on Bentham and his influence on the law and economics movement.
7. One should note that, if actual compensation was carried out, any test satisfying the Kaldor–Hicks criterion of efficiency would also satisfy the Pareto criterion.
8. 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 (that is, 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 Ullmann-Margalit (1977); Sen (1979); and Gauthier (1986).
9. According to the Nash criterion, social welfare is given by the product of the utility of the members of society (Nash, 1950). See Mueller (1989: 379–82), attributing the multiplicative form of the social welfare function to Nash.
10. Posner is the most notable exponent of the wealth-maximization paradigm. Under wealth-maximization principles, a transaction is desirable if it increases the sum of wealth for the relevant parties (where wealth is meant to include all tangible and intangible goods and services). Bentham (1839) had already challenged the use of objective factors, such as wealth or physical resources, as a proxy for human happiness. Despite the difficulties in quantification of values such as utility or happiness, the pursuit of pleasure and happiness and the avoidance of and pain are the motivating forces of human behaviour. Wealth, food and shelter are mere instruments to achieve such human goals.

References

Bentham, Jeremy (1782), *Of Laws in General*.
Bentham, Jeremy (1789), *An Introduction to the Principles of Morals and Legislation*.
Bentham, Jeremy (1839), *A Manual of Political Economy*, New York: G.P. Putnam.
Buchanan, James (1974), 'Good economics – bad law', 60 *Virginia Law Review* 483–92.
Buchanan, James M. (1990), 'The domain of constitutional economics', 1 (1) *Constitutional Political Economy* 1–18.

Buckley, Francis H. (1999), *The Fall and Rise of Freedom of Contract*, Durham, NC: Duke University Press.
Calabresi, Guido (1980), 'About law and economics: a letter to Ronald Dworkin', 8 *Hofstra Law Review* 553–62.
Calabresi, Guido (1991), 'The pointlessness of Pareto: carrying Coase further', 100 *Yale Law Journal* 1211.
Coase, Ronald H. (1960), 'The problem of social cost', 3 *Journal of Law and Economics* 1–44.
Coase, Ronald H. (1978), 'Economics and contiguous disciplines', 7 *Journal of Legal Studies*, 201–11.
Cooter, Robert (1984), 'Prices and sanctions', 84 *Columbia Law Review* 1523–60.
Cooter, Robert D. (1994), 'Structural adjudication and the new law merchant: a model of decentralized law', 14 *International Review of Law and Economics* 215.
Cooter, Robert D. (2000), 'Three effects of social norms on law: expression, deterrence, and internalization', 79 (1) *Oregon Law Review* 1.
Edgeworth, Francis Y. (1881), *Mathematical Psychics*, London: C. Kegan Paul.
Ehrich, Isaac and Richard A. Posner (1974), 'An economic analysis of legal rule making', 3 *Journal of Legal Studies*, 257–86.
Fon, Vincy and Francesco Parisi (2003), 'The limits of reciprocity for social cooperation', George Mason Law and Economics Research Paper No. 03-08, George Mason University, Fairfax, VA.
Gauthier, D. (1986), *Morals by Agreement*, Oxford: Clarendon.
Harsanyi, J.C. (1955), 'Cardinal welfare individualistic ethics, and interpersonal comparisons of utility', 63 *Journal of Political Economy* 315.
Hicks, J.R. (1939), 'The foundations of welfare economics', 49 *Economic Journal* 696–712.
Kaldor, N. (1939), 'Welfare propositions of economics and inter-personal comparisons of utility', 49 *Economic Journal* 549–52.
Kaplow, Louis and Steven Shavell (1994), 'Why the legal system is less efficient than the income tax in redistributing income', 23 *Journal of Legal Studies*, 667–81.
Kelly, Paul (1998), 'Jeremy Bentham', 1 in P. Newman (ed.), *The New Palgrave Dictionary of Economics and the Law*, London: Macmillan.
MacKaay, Ejan (2000), 'History of law and economics', 1 *Encyclopedia of Law and Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Mueller, Dennis (1989), *Public Choice II*, Cambridge: Cambridge University Press.
Nash, John (1950), 'The bargaining problem', 18 *Econometrica* 155.
Parisi, Francesco (1998), 'Customary law', 1 in P. Newman (ed.) *The New Palgrave Dictionary of Economics and the Law*, London: Macmillan, pp. 572–8.
Parisi, Francesco and Larry Ribstein (1998), 'Choice of Law', 1 in P. Newman (ed.), *The New Palgrave Dictionary of Economics and the Law*, London: Macmillan, pp. 387–96.
Posner, Richard A. (1985), 'Wealth maximization revisited', 2 *Notre Dame Journal of Law, Ethics, and Public Policy* 85–105.
Posner, Richard A. (1988), 'The ethics of wealth maximization: reply to Malloy', 36 *Kansas Law Review* 261.
Posner, Richard A. (1998), 'Bentham's influence on the law and economics movement', 51 *Current Legal Problems* 425.
Posner, Richard A. and Francesco Parisi (1998), 'Scuole e Tendenze nella Analisi Economica del Diritto', 147 *Biblioteca della Liberta'* 3–20.
Priest, George L. (1977), 'The common law process and the selection of efficient rules', 6 *Journal of Legal Studies* 65–82.
Rawls, John (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.
Ribstein, Larry and Bruce Kobayashi (2001), 'A recipe for cookies: state regulation of consumer marketing information', George Mason Law and Economics Research Paper No. 01-04, George Mason University, Fairfax, VA.
Ribstein, Larry and Erin A. O'Hara (2000), 'From politics to efficiency in choice of law', 67 *University of Chicago Law Review* 1151.
Romano, Roberta (1999) 'Corporate law as the paradigm for contractual choice of law', in F.

Buckley (ed.), *The Fall and Rise of Freedom of Contract*, Durham, NC: Duke University Press, p. 390.
Rowley, Charles K. (1981), 'Social sciences and the law: the relevance of economic theories', 1 *Oxford Journal of Legal Studies* 391–405.
Rowley, Charles K. (1989), 'The common law in public choice perspective: a theoretical and institutional critique', 12 *Hamline Law Review* 355–83.
Rubin, Paul H. (1977), 'Why is the common law efficient?', 6 *Journal of Legal Studies* 51–63.
Scitovsky, T. (1941), 'A note on welfare propositions in economics', 9 (1) *Review of Economic Studies* 77–88.
Sen, A.K. (1979), 'Rational fools: a critique of the behavioural foundations of economic theory', in F. Hahn and M. Hollis (eds), *Philosophy and Public Affairs*, **6**, pp. 317–44.
Smith, Adam (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*.
Smith, Vernon, Kevin McCabe and Stephen Rassenti (1998), 'Reciprocity, trust and payoff privacy in extensive form bargaining', 24 *Games and Economic Behavior* 10–24.
Trebilcock, Michael J. (1993), *The Limits of Freedom of Contract*, Cambridge, MA: Harvard University Press.
Ullmann-Margalit, E. (1977), *The Emergence of Norms*, Oxford: Clarendon Press.
Vanberg, Viktor J. (1994), *Rules and Choice in Economics*, London: Routledge.

5 Commons and anticommons

Francesco Parisi and Ben Depoorter

Commons and anticommons problems are the consequence of symmetric structural departures from a unified conception of property, and are the consequence of a lack of conformity between use and exclusion rights (Parisi et al. 2004).

Commons and anticommons: two tragedies on common grounds

Recently, a new term has gained acceptance among law and economics scholars of property law: the 'anticommons'. The concept, first introduced by Michelman (1982) and then made popular by Heller (1998, 1999), mirror images in name and in fact of Hardin's (1968) well-known 'tragedy of the commons'.

In situations where multiple individuals are endowed with the privilege to use a given resource, without a cost-effective way to monitor and constrain each other's use, the resource is vulnerable to overuse, leading to a problem known as the tragedy of the commons. Symmetrically, when multiple owners hold effective rights to exclude others from a scarce resource, and no one has an effective privilege of use, the resource might be prone to underuse, leading to a problem known as the 'tragedy of the anticommons'. As pointed out by Buchanan and Yoon (2000), the effects of the two problems are in many respects symmetrical.

The commons problem

If a depletable resource is open to access by more than one individual, incentives for overutilization will emerge. As the number of individuals enjoying free access grows larger relative to the capacity of the common resource, overutilization will approach unsustainable levels and the utilizers will risk the complete destruction of the common good. Although Hardin (1968) calls this destruction the 'tragedy of the commons', he credits a mathematical amateur, William Forster Lloyd (1794–1852), for formalizing it in a little-known pamphlet published in 1833 on population growth.

Since Lloyd, other economists have identified the problems associated with the common ownership of resources exploited under conditions of individualistic competition. Most notably, Gordon (1954) pointed out that without controls on entry, common resources will be exploited even at levels of negative marginal productivity. This is because external effects are not fully

internalized within the choice of each individual decision maker. The sources of externalities in a commons problem are twofold. First, there are static (or current) externalities, in that the use of the resource reduces the benefit from usage to others. Second, there are possible dynamic (or future) externalities because the use of a renewable resource today bears its consequences into the future. Due to the lack of conformity between use and exclusion rights, individuals do not have to consider the full social costs of their activities. Private and social returns diverge and total use by all parties exceeds the social wealth-maximizing point.

The anticommons problem

The term 'anticommons' was coined by Michelman in his article 'ethics, economics and the law of property' (1982). Michelman defined the anticommons as 'a type of property in which everyone always has rights respecting the objects in regime, and no one, consequently, is ever privileged to use any of them except as particularly authorized by others' (Michelman, 1967), which had almost no counterpart in real-world property relations. The hypothetical example would be that of a wilderness preserve where any person has the authority to enforce the wilderness conservation laws and regulations.

Heller (1998) revitalized the concept in an article on the transition to market institutions in contemporary Russia, where he discusses the intriguing prevalence of empty storefronts in Moscow. Storefronts in Moscow are subject to underuse because there are too many owners (local, regional and federal government agencies, mafia and so on) holding the right to exclude. The definition of the anticommons as employed by Heller is 'a property regime in which multiple owners hold effective rights of exclusion in a scarce resource' (ibid.: 668).

In the tragedy of the anticommons, the coexistence of multiple exclusion rights creates conditions for suboptimal use of the common resource. If the common resource is subject to multiple exclusion rights held by two or more individuals, each co-owner will have incentives to withhold resources from other users to an inefficient level. In the presence of concurrent controls on entry exercised by individual co-owners acting under conditions of individualistic competition, exclusion rights will be exercised even when the use of the common resource by one party could yield net social benefits. In other words, some common resources will remain idle even in the economic region of positive marginal productivity. Again, this is because the multiple holders of exclusion rights do not fully internalize the cost created by the enforcement of their right to exclude others.

As with the commons problem, the sources of externalities in an anticommons problem are also twofold. First, there are static (or current)

externalities, in that the exercise of a right of exclusion by one member reduces or eliminates the value of similar rights held by other individuals. In price theory terms, one can think of this externality as the cross-price effect of the various exclusion rights. Second, the withholding of productive resources may create dynamic (or future) externalities, because the underuse of productive inputs today bears its consequences into the future, as standard growth theory suggests.

In search of a common ground: a unified conception of property

The symmetrical features of commons and anticommons cases are the result of the same underlying problem. In both situations there is a misalignment of the private and social incentives of multiple owners in the use of a common resource. This misalignment is due to the presence of externalities that are not captured in the calculus of interests of the users (commons situations) and excluders (anticommons situations).

The unitary basis of the problem can be understood in terms of the traditional structure of a property right. According to the traditional conception of property, owners enjoy a bundle of rights over their property, which include, among other things, the right to use their property and the right to exclude others from it. In such a framework, the owner's rights of use and exclusion are exercised over a similar domain. Right to use and right to exclude are, in this sense, complementary attributes of a unified bundle of property rights.

The commons and anticommons problems can be seen as deviations in symmetric directions. In commons situations, the right to use is stretched beyond the effective right (or power) to exclude others. Conversely, in anticommons situations, the co-owners' rights of use are compressed, and potentially eliminated, by overshadowing rights of exclusion held by other co-owners. In both commons and anticommons cases, rights of use and rights of exclusion have non-conforming boundaries. Such lack of conformity causes a welfare loss due to the forgone synergies between those complementary features of a unified property right.

This conceptualization of the commons and anticommons suggests a link between the welfare losses of the two cases and a dual model of property. As noted above, welfare losses are produced by a discrepancy between the rights of use and the rights of exclusion held by the various owners. The problem is in this way detached from the usual understanding of the tragedy of the commons as a consequence of poorly defined or absent property rights (Cheung, 1987). Common and anticommons problems are not confined to situations of insufficient or excessive fragmentation of ownership, but result from the dismemberment and resulting non-conformity between the internal entitlements of the property right.

It follows that the qualitative results of the commons and anticommons models represent limit points along a continuum, each characterized by different levels of discrepancy between use and exclusion rights, with varying welfare losses from commons or anticommons problems.

Anticommons, property fragmentation and the laws of entropy

Building upon the recent literature on anticommons and property fragmentation, Parisi (2002a) considers the proposition that property is subject to a fundamental law of entropy. With this metaphor Parisi refers to the second law of thermodynamics, according to which every process that can occur spontaneously will go in one direction only and will result in a release of energy that cannot be recaptured, so that the amount of entropy in the universe will continually increase. In the property context, entropy induces a one-directional bias which leads towards increasing property fragmentation. The law of entropy further indicates that only in the purely abstract case of (both internally and externally) reversible transformations will the overall net change in entropy be zero. In the property context, this indicates that only in a world of zero transaction costs would there be no such tendency towards fragmentation.

The economic forces that induce entropy in property are quite straightforward. Property division creates one-directional inertia: unlike ordinary transfers of rights from one individual to another, reunifying fragmented property rights usually involves transaction and strategic costs higher than those incurred in the original deal. Consider the case of unified property as the starting point: a single owner faces no strategic costs when deciding how to partition his/her property. Conversely, the reunification of fragmented rights requires the participation of multiple parties, with an unavoidable increase in transaction and strategic costs. For example, even reversing a simple property transaction can result in monopoly pricing by the buyer-turned-seller; reunifying property that has been split among multiple parties engenders even higher costs, given the increased difficulty of coordination among the parties.

Thus the move from unified property to fragmented property and vice versa poses an interesting situation of asymmetric transaction costs. The presence of such asymmetry is due to the fact that fragmented owners are faced with a strategic problem, given the interdependence of their decisions. The equilibrium pricing (or quantity supply) of fragmented owners impedes the optimal reunification of non-conforming fragments into a unified bundle.

A model of duopolistic anticommons

In the context of property, Posner (1998: 76) first recognized the costs of excessive property fragmentation. Heller argued that it is often harder to

regenerate separated bundles than to fragment them (1998, 1999); Buchanan and Yoon (2000); Schulz et al. (2002) and Parisi et al. (2004) restated this thesis with formal economic models.

For the purpose of illustrating the problem of the anticommons and the resulting entropy in property, we can thus briefly restate the results of such literature, considering a simple model of property rights fragmentation. Suppose that agent 1 owns a large estate of land which he/she uses as a commercial farm. Agent 2 acquires from agent 1 the right to use the estate for recreational hunting. As a result, the unitary property right is fragmented, giving the two agents partial property rights and reciprocal exclusion privileges. The property right of agent 1 is constrained by the real interests acquired by agent 2. Agent 1 holds a right to exclude any use of agent 2 other than recreational hunting. Agent 2 conversely holds a right to exclude any use of the land by agent 1, which is in conflict with his/her acquired rights. In this sense, the previously unitary proprietary interest over the land is now fragmented. Such fragmentation will remain beneficial for all parties as long as the mixed use of the land for the respective activities of the fragmented owners remains the most valuable allocation of the land for the parties.

Suppose now that a third party sees an opportunity which would generate more value than the current use. Take, for example, the construction of a hotel resort on the estate. The construction would obviously compress the property rights held by the two agents. Each agent could thus withhold his/her consent to the transformation of the land and exercise his/her veto right impeding the value-enhancing transformation. However, as the opportunity is supposed to be more valuable than the current use, it would be rational for the various agents to agree to the proposed transformation. Yet, each fragmented owner would rationally attempt to maximize his/her profit from the sale of his/her fragmented property right. We should thus consider the likely price mechanism that would lead to the development of the land and compare it to the alternative scenario of a property transformation controlled by a single unified property owner.

An application of the Buchanan and Yoon (2000) model could illustrate our problem. In the presence of development opportunities, a third party who wishes to utilize property fragments for the reunified development of the land needs to obtain the consent of all fragmented property holders. Because in the face of a redevelopment opportunity, each property fragment constitutes a strictly complementary input for the achievement of reunified property, the demand for each property fragment depends not only on the price set by the individual fragmented owner, but also on the price charged by the other property right holders. This implies that any change in the price or quantity supply of the complementary good by one duopolistic property seller will have external effects for the other property seller. Each

party maximizes his/her profits, without regard to the effect of his/her own pricing strategy on the profits of other property owners. When one seller decreases output and raises the fragment price, the demand curve faced by the other property owners will be negatively affected, and vice versa. A concentrated monopolistic seller of the land would instead fully internalize these price or output externalities.

A simple illustration is useful. Suppose our two fragmented owners, A and B, each own two fragments of property the reunification of which is essential for the development of the land. Assume that each fragmented property owner must make a decision about price without knowing what the other seller will do. To simplify, suppose there are only three pricing options: the price a single monopolistic seller would demand, P_M, a price greater than P_M, or a price smaller than P_M. The game matrix in Figure 5.1 illustrates the incentives facing each seller. Each cell would contain the payoff (profit) to seller A and seller B from the corresponding combination of their pricing decisions. Seller A is a row player, and its Nash strategy given each of seller B's choices is indicated with the dotted, vertical arrows. Seller B is a column player, and its Nash strategies given each of A's potential choices are indicated with the solid, horizontal arrows.

Here, given the cross-price effects present in this complementary anticommons, both sellers would have a dominant strategy, with a single

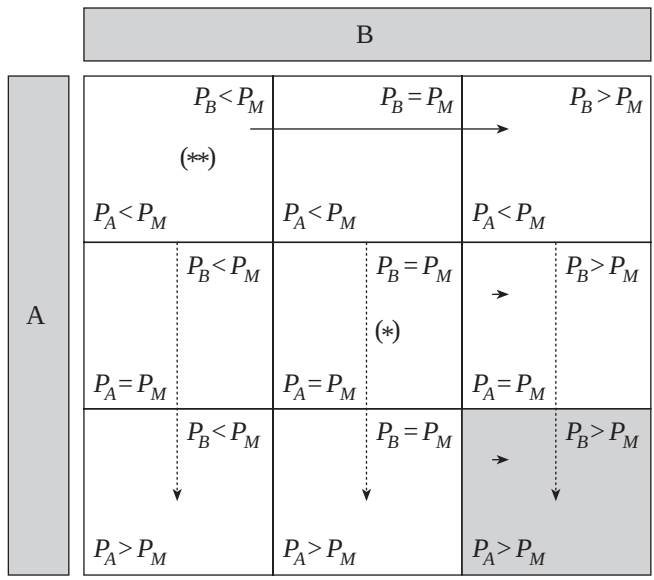


Figure 5.1 Game matrix

Nash equilibrium, indicated by the shaded areas in Figure 5.1. The sellers will choose to price above P_M, to the detriment of both the developer's profits and the overall (that is, developer's plus sellers') welfare. The cells corresponding to the profit-maximizing prices and the welfare-maximizing prices are respectively marked with a single asterisk (*) and a double asterisk (**), in the figure.

It should be noted that in our anticommons problem strategies $P_A = P_B > P_M$ obtain in equilibrium. The sellers' strategic pricing renders the optimal property reunification unobtainable in equilibrium. The uncoordinated pricing of the two fragmented property owners results in a higher total cost of land development, and therefore to a potential underutilization of the land, beyond what either of them would expect as a unified monopolistic owner of the estate, in order to maximize their own profit. Interestingly, the 'competitive' (that is, fragmented) supply of land development rights leads to higher prices than those that would be charged by a single 'monopolistic' (that is, unified) owner.

As pointed out by Schulz et al. (2002), the differences between the two equilibria are due to the presence of negative externalities in the independent choices of the fragmented property rights. This result should not come as a surprise. The position of multiple property owners in the face of a new opportunity, which requires a reunification of their fragmented property rights, creates a strategic problem similar to the well-known hold-up problem. Suboptimal final use of resources may result from such fragmentation.

A model of oligopolistic anticommons

The above model of property fragmentation can be extended to show that an increase in the extent of fragmentation exacerbates the result of final underutilization of the resource. Recalling our example, imagine that the estate was partitioned among a larger number of agents, *n*. Let us further assume that the property fragments are controlled by independent agents and that the development of the land necessitates the agreement of all *n* individuals. What would be the equilibrium price of the land lease if the fragmented property owners are pricing their fragments independently from one another?

The model of duopolistic anticommons illustrated above can be extended to illustrate the more general problem of oligopolistic anticommons. We shall thus develop a simple model of oligopolistic anticommons with *n* fragmented property owners, showing that the extent of the deadweight loss also depends on the extent of property fragmentation.

Suppose that *n* individuals hold property rights over *n* fragments, which can be used as inputs of production for a composite good, *Q*. Because of their strict complementarity as inputs for the reunification of land, the demand for each depends on the price of all others. P_Q is the sum of the prices of the *n*

separate fragmented property rights $\sum_{i=1}^{n} P_i$. Each owner of a specific input of production thus has a profit function that can be written as:

$$\Pi_j = P_j D(P_Q) = P_j D\left(\sum_{i=1}^{n} P_i\right). \tag{5.1}$$

Differentiating the profit functions with respect to the corresponding price variable yields these first-order conditions:

$$\frac{\partial \Pi_j}{\partial P_j} = P_j D'(P_Q) + D(P_Q) = 0. \tag{5.2}$$

Summing the first-order conditions yields the equilibrium price for the unified land when the fragmented property rights are held by independent individuals, operating in an oligopolistic anticommons:

$$P_Q D'(P_Q) + nD(P_Q) = 0. \tag{5.3}$$

We can now compare these conditions with those that characterize the supply of a single owner of unified property with monopolistic control over his/her property (or by separate owners, who can effectively coordinate prices). In the case where a single individual owns all property fragments, the profit function will take the following form:

$$\Pi = P_Q D(P_Q). \tag{5.4}$$

By differentiating this profit function with respect to the price, we determine the first-order conditions for the single owner:

$$\frac{\partial \Pi}{\partial P_Q} = P_Q D'(P_Q) + D(P_Q) = 0. \tag{5.5}$$

The interesting comparison is between the optimal prices in equations (5.3) and (5.5). One finds that the optimal price under unified property sold by a single monopolist (equation (5.5)) is actually lower than the total price that would be necessary in order to reunify the land under an oligopolistic anticommons (equation (5.3)). It is also interesting to look at the comparative statics of equation (5.3) with respect to the number of property fragments. By inspection, it is possible to see that both overall price, and overall deadweight loss, increases in *n*. In the case of oligopolistic anticommons, the strategic pricing of the fragmented property owners leads to higher

prices: both the fragmented property sellers and the third-party developer see their respective surplus diminished compared to the alternative monopoly outcome.

Conclusions

The results of the anticommons in the context of property do not strictly depend on the legal or physical nature of property fragmentation. Property fragmentation merely indicates the existence of multiple rights held by different individuals to control or veto a change in the use of their land. As shown in our example, suboptimal final allocations of resources may be the consequence of fragmented decision rights, even when such fragmentation concerns a unitary physical asset. Even in the face of value-enhancing opportunities, multiple right holders may face incentives to employ their veto power to maximize the private return from the joint enterprise. The combined effect of the various agents' strategies leads to an inefficient outcome.

The outcome of this model of fragmented property is perhaps most easily understood if it is further recognized that each agent exerts a positive externality on the other agent. Hence, the above result is consistent with conventional wisdom, according to which in situations of positive externalities the use of some resource is less than optimal. These results were more extensively formalized by Schulz et al. (2002) and show that the severity of the deadweight losses from dysfunctional property fragmentation increases monotonically with the number of independent fragmented owners. The larger the number of individuals who can independently price an essential input for the land development project, the higher the equilibrium price that each of these individuals will demand for his/her own fragment. At the limit, as the number of fragmented owners approaches very large numbers (or infinity), complete abandonment of the land will result.

References

Buchanan, J. and Y.J. Yoon (2000), 'Symmetric tragedies: commons and anticommons property', *Journal of Law and Economics* **43** (1), April, pp. 1–13.

Cheung, S.N.S. (1987), 'Common property rights', 1, in J. Eatwell, M. Milgate and P. Newman (eds), *The New Palgrave: A Dictionary of Economics*, London: Macmillan, 504–5.

Gordon, H.S. (1954), 'The economic theory of a common-property resource: the fishery', *Journal of Political Economy*, **62**, 124–42.

Hardin, G. (1968), 'The tragedy of the commons', *Science*, **162**, 1243–8.

Heller, M.A. (1998), 'The tragedy of the anticommons: property in the transition from Marx to markets', *Harvard Law Review*, **111**, January, 621.

Heller, M.A. (1999), 'The boundaries of private property', *Yale Law Journal*, **108**, 1163–223.

Michelman, F.I. (1967), 'Property, utility and fairness: comments on the ethical foundations of "just compensation" law', *Harvard Law Review*, **80** (6), April, 1165–258.

Parisi, F. (2002a), 'Entropy in property', *American Journal of Comparative Law*, **50**, 701–38.

Parisi, F. (2002b), 'Freedom of contract and the laws of entropy', *Supreme Court Economic Review*, **10**, 65–90.

Parisi, F., N. Schulz and B. Depoorter (2004), 'Simultanous versus sequential anticommons', *European Journal of Law and Economics*, **17** (March) (2), 175–90.
Posner, R.A. (1998), *Economic Analysis of Law*, 5th edn, New York: Aspen.
Schulz, N., F. Parisi and B. Depoorter (2002), 'Fragmentation in property: towards a general model', *Journal of Institutional and Theoretical Economics*, **158**, December, 594–613.

PART II

PRIVATE LAW AND ECONOMICS

6 The economics of tort law

Giuseppe Dari Mattiacci and Francesco Parisi

The relatively simple structure of a tort problem provides one of the most fertile areas for the application of economic analysis to law. The positive economic theory of tort law maintains that the common law of torts is best explained as though judges were trying to promote efficient resource allocation, that is, maximize efficiency. The Coase (1960) theorem shows that if parties are allowed to negotiate and transaction costs are sufficiently low, legal entitlements will be reallocated efficiently. In the case of tort accidents, transaction costs are high. This is easily understood because the parties potentially involved in an accident are not easily identifiable *ex ante*, and the cost of acquiring the relevant information for bargaining can be high. This renders contractual arrangements *à la* Coase impracticable. In most tort situations the legal system thus needs to provide rules to give potential injurers and potential victims appropriate incentives to act as if they had to bear the total social cost of their activities. This is an important goal of tort law. Tort law is therefore justified when bargaining is not possible because high transaction costs are present, and banning an activity is undesirable given the social value of the risk-creating activity (Calabresi and Melamed, 1972).

In order to create optimal incentives, liability rules need to induce parties to minimize the total social cost of accidents. The relevant variables for this tort problem are the cost of accidents, the cost of accident avoidance (precaution), and the administrative costs of the justice system. Every legal system chooses from various liability rules (for example, negligence, strict liability and so on) and safety standards to minimize the overall cost of accidents.

The goals of tort law

A first intuitive end of tort law is to compensate the victims for losses due to accidents. This is indeed an important task of tort adjudication but it is not the central issue concerning the design of tort rules. It has been shown that tort law is a very expensive means of compensating harms, because it involves high administrative cost due to the functioning of the judicial system. Insurance, to the contrary, is a much cheaper and quicker system (Shavell, 1987: 263): if the only goal were to compensate victims, first-party insurance would be preferable over tort liability. Moreover, the cost of insurance can be paid by the potential injurers, shared among potential victims or financed by taxpayers, in order to redistribute the costs (McEwin, 2000).

On the contrary, economic analysis suggests that the primary reason for utilizing the tort system is to allow risk-creating activities to be carried out only if the social value of the activity justifies the risk created. This balancing of costs and benefits is currently endorsed by North American tort doctrine and is clearly summarized by the *Restatement (Second) of Torts* § 291:

> Where an act is one which a reasonable man would recognize as involving a risk of harm to another, the risk is unreasonable and the act negligent if the risk is of such magnitude as to outweigh what the law regards as the utility of the act or of the particular manner in which it was done.

(The *Restatement (Second) of Torts* §§ 292 and 293, indicate the criteria for verifying utility of conduct and magnitude of risk.)

More specifically, economic analysis suggests that tort law should be designed in such a way as to provide potential injurers and victims with appropriate incentives to avoid the accident by internalizing the externalities created by their activities. In the absence of tort liability, potential tortfeasors would bear the private cost of their precaution without internalizing any of the benefits thereof. The benefits (of precaution) are external with respect to the decision (on how much precaution). This would lead to suboptimal levels of care and excessive accident rates. Through tort liability, a potential tortfeasor internalizes the benefits of his/her precaution, that is, the reduction in expected liability. Tort rules should thus be designed to induce parties to internalize the external costs of their activities and to adopt optimal levels of precaution.

In addition, tort law gives parties incentives to acquire information about the accident. With respect to risk, the tort law system should enhance an optimal allocation of the risk between victim and injurer, but this goal can be reached via insurance. With respect to transaction costs, the goal of the tort law system is to minimize the administrative cost associated with the functioning of the system itself (mainly the costs of courts and lawyers and the indirect costs borne by litigants). Calabresi (1970) presented the first formulation of the ends of liability in those terms, while Brown (1973) formalized an economic model of accidents. We shall focus on incentives towards optimal precaution and discuss the other aspects in passing.

A taxonomy of liability rules

There are several ways in which legal systems can apportion liability between parties. Historically, a broad variety of liability rules has been developed by legal systems. Most early legal systems adopted liability rules that did not depend on the fault of the tortfeasor. This feature of early legal systems has been explained as instrumental to promoting appeasement between the parties and to avoiding cumbersome and controversial ascertainment of the

subjective elements of a tort (Parisi, 1992, 2001). Gradually, legal systems began to recognize fault as a viable basis for liability and in modern legal regimes strict liability is seen as an exception to the norm. Liability for accidents should arise only in the case of tortfeasor fault (including both negligent and intentional wrongdoing).

We shall proceed with the presentation of some of the most common liability rules, starting from strict liability to simple negligence and more complex legal regimes. In our analysis we shall utilize the conventional terminology by which the injurer is defined as the individual who does not suffer harm in an accident and the victim as the individual who suffers such harm. In this survey, we shall focus on two-party accidents.

There are two fundamental possibilities in a two-party accident. The first occurs when both parties have to take precaution in order to avoid the accident (*bilateral precaution*). The second is given by situations in which either party can take precaution and successfully avoid an accident (*alternative precaution*). In the second case, there is a waste of precaution cost if both parties take precaution, since one party's precaution would already have been enough. A particular and common case of alternative precaution is unilateral precaution. As in alternative precaution, one party's precaution is enough to prevent the accident, but only one party has the actual possibility of avoiding the accident. We shall consider all such possibilities when referring to the effect of alternative legal rules on the parties' behaviour.

Strict rules: no liability and strict liability

Strict liability can be thought of as the mirror image of no liability. A party who occasions harm to another will compensate the victim regardless of who is at fault. This rule is the converse of no liability. No liability can in fact be thought of as strict liability on the part of the victim, in that the victim always bears the loss regardless of the parties' fault. No liability and strict liability can thus be considered the limit points in the range of possible liability rules. The choice between strict liability and no liability has obvious distributive effects, in that strict liability results in the victim always being compensated, while no liability makes the victim bear all accident costs.

The different allocation of accident costs has clear incentive effects. In a strict liability system, the injurer has to bear both the cost of precaution and the expected accident cost and, hence, he/she will minimize the sum of those costs. This will lead to the efficient level of precaution. On the contrary, a no-liability rule does not achieve an efficient result because the injurer would bear the cost of precaution without internalizing the benefit of such precaution. In the absence of liability, the injurer would adopt no precaution at all, which is an inefficient result. On the other hand, if we look at the victim's incentives to take precaution, we see that the opposite is true. Strict liability

creates no incentives for victim precaution, while no liability would shift the entire residual liability on the victim, inducing optimal victim care. It follows that strict liability and no liability can give incentives to take efficient precaution only to one party, respectively either the injurer or the victim. Strict liability will fail to produce an efficient outcome when the avoider is the victim, and no liability will fail when the avoider is the injurer.

With respect to alternative precaution, the result is slightly different. In the case of unilateral precaution, if the tort law system fails to target the avoider, he/she will take no precaution at all, while in the case of alternative precaution, either party can take precaution; therefore, imposing liability on the party who is not the least-cost avoider will result in suboptimal precaution levels and excessive precaution costs. Strict liability and no liability can thus yield efficient results only in the case of unilateral or alternative precaution, provided that liability is allocated on the least-cost avoider. In the case of bilateral precaution, both strict liability and no liability fail to generate optimal incentives, because neither rule can simultaneously threaten both parties with liability in a Nash equilibrium. In bilateral-precaution situations a different rule is therefore needed to induce both parties to adopt the necessary precautions.

Negligence rules in general

Fault can be seen as a way of creating optimal incentives on both tortfeasors and victims and also of achieving efficiency in the case of bilateral precaution. Negligence rules draw a line between liability and no liability by identifying a level of due care and verifying whether the relevant party adopted that level of due care. American case law in a sense anticipated the economic definition of negligence, adopting the simple and formal logic of cost–benefit analysis to adjudicate tort cases. Already in 1947, Judge Learned Hand, in the celebrated decision of *United States* v. *Carroll Towing Co.* (159 F.2d 169 (2d Cir. 1947)), clarified the tradeoffs between the costs and benefits of risk and prevention activities using a mathematical formula. This rule became a milestone in the law of torts, and it is now known as the Hand formula of negligence. The formula defines negligence as a function of three variables: (a) the probability of a harmful event occurring (magnitude of risk); (b) the seriousness of the damage that may result from this event (gravity of harm); and (c) the cost of preventing the occurrence of the harmful event (burden of prevention). In the original formula, (P) indicates the magnitude of risk; (L) indicates the gravity of the loss; and (B) indicates the burden of prevention (that is, the cost of adequate precautions). According to the Hand formula, conduct is negligent if the cost of adequate precautions is less than the cost of the injury multiplied by the probability of its occurrence, that is, if $(B) < (PL)$.

Although the Hand formula does not directly consider the social value of risk-creating behaviour, it produces the proper incentives for the evaluation of such behaviour. By imposing a balance between risk and prevention, the result in *Carroll Towing* encourages individuals to weigh the cost of prevention against the utility of the behaviour. When deciding whether to engage in an activity, the reasonable person will consider whether the utility derived from the activity justifies the risk of liability and/or the cost of prevention (this is, indeed, the question of the *Restatement (Second) of Torts* § 291, comment (a), which asks whether 'the game is worth the candle'). According to this logic, individuals will respond to liability rules by undertaking the socially optimal level of precaution. A vast region of law and economics literature has explored the wisdom of this tort doctrine, often with the use of formal economic models, bringing to light the importance of using marginal (rather than total) values in the assessment of liability. Along similar lines, after establishing a positive economic model of tort law, Landes and Posner (1982) conclude that the Hand formula of negligence, as applied, coincides with the economic model of due care.

Introducing fault means setting a due level of precaution, defined by the legislator or by the judge. The due level of precaution should be set to be equal to the efficient level of precaution. Under any negligence rule the judge has to perform such a test by confronting the level of precaution actually taken by the parties with the due level of precaution. This increases the administrative cost of adjudication compared to strict liability rules and generates some complexities.

Among such complexities is the fact that while some forms of precaution are easily observable *ex post*, others are very difficult, or even impossible to assess and to compare with the legal standard of precaution. In the presence of non-observable precautions, it is clear that individuals would rationally limit their investment to observable precaution to avoid negligence and refrain from investing in non-observable precautions, since they could not draw much benefit from such investment.

In the law and economics literature the case of non-observable precautions is generally treated under the discussion of care versus activity levels. The most common example of activity level is the repetition of a dangerous action, such as driving. Although courts may occasionally take into account the frequency of an activity in their assessment of negligence, often no threshold of 'optimal frequency' can easily be utilized by legal rules as a liability allocation mechanism, given the difficulty of pinpointing a critical value to separate efficient from excessive activity. Since courts cannot be asked to balance unascertainable costs and benefits and cannot be asked to evaluate non-observable precaution levels, it is clear that the types of precautions that are evaluated for the finding of negligence are generally confined to

care levels, not activity levels. Therefore, the introduction of the criterion of negligence introduces a dichotomy between care-type and activity-type precaution investments (Shavell, 1980a). No such distinction between care and activity level is relevant in regimes of strict liability and no liability.

Negligence rules under which the victim is the residual bearer
Hereafter we shall analyse those rules that are generally referred to as negligence rules. It will soon be clear that we can think of them as being constructed by adding a negligence defence to a rule of no liability.

Simple negligence Within negligence regimes, the most straightforward rule is simple negligence. Under simple negligence an injurer is liable for damages only if he/she is found negligent. The victim bears the so-called 'residual liability', in the sense that he/she has to bear the consequences of the accident if the injurer cannot be blamed for negligence. In this sense, simple negligence is analogous to a no-liability rule, because it leaves residual liability on the victim.

With unilateral-precaution accidents, when the victim is the avoider, the injurer cannot be declared negligent since it is not possible for him/her to take effective precautions. Therefore, the victim bears the cost of the accident and he/she will have incentives to take the optimal level of precautions (care and activity level) as under no liability. If the avoider is the injurer, he/she will have to pay only when he/she does not take at least the due level of care. If the injurer is negligent, he/she has to bear the cost of care and the expected accident cost (pay damages to the victim). On the contrary, if the injurer takes due care, he/she avoids liability and bears only the cost of care. If due care is set at the efficient level, the injurer will have incentives to take due care. We can conclude that in unilateral-precaution cases simple negligence produces the right incentive to take optimal care when either the injurer or the victim is the avoider.

However, with respect to activity level, only the victim, as a residual bearer, has incentives to take the optimal level of precaution. In equilibrium, in fact, the injurer will adopt due care and avoid liability, so that any investment in non-observable precautions would yield him/her no private benefit. Simple negligence thus gives efficient incentives with respect to activity level only to the victim, since he/she bears the full cost of the accident in equilibrium. The same logic allows us to show that also in the case of alternative precaution both parties face incentives to adopt optimal care levels. However, we know that in such situations efficiency requires that only the least-cost avoider to take care. If the other party or both parties adopt precautions there is an inefficient result. One way to avoid obtaining such an inefficient outcome is to formulate the negligence criterion in light

of such requirement, so that negligence could be found only when the injurer is the least-cost avoider.

The introduction of the requirement of negligence improves the performance of the rule in bilateral-precaution situations. The negligence criterion makes both parties take the optimal level of care in all situations (unilateral, alternative and bilateral precaution), but gives incentives to choose an optimal activity level only to the residual bearer, the victim.

Contributory and comparative negligence Under contributory negligence, the injurer is liable to compensate his/her victim only if he/she was negligent and the victim was careful. In all remaining cases the victim remains the residual bearer and receives no compensation for his/her loss. Therefore, the victim does not have the right to compensation when both were negligent, when both were careful, and obviously when the injurer was careful and the victim negligent. Similar results are reached with a rule of comparative negligence. In a regime of comparative negligence, however, victim negligence does not constitute a complete bar to recovery but leads to a reduction of liability in proportion to the parties' respective levels of negligence.

In both regimes, the injurer can escape liability by taking the due level of care. This creates the appropriate incentive for the injurer to comply with the legal standard of care. Given that the injurer can reasonably expect to bear the entire residual loss, he/she would also face incentives to behave carefully, since he/she would internalize the full benefit of his/her precaution investment. Contributory and comparative negligence thus create efficient care incentives for both parties, but only the victim, as a residual bearer, would have incentives to undertake an optimal activity level. The injurer is able to avoid liability with the adoption of due care and therefore would have no incentive to invest in non-observable precautions. Contributory and comparative negligence thus produce the same set of incentives generated by a rule of simple negligence, but they have possible distributive effects, because they would either foreclose or reduce compensation when the victim is found negligent.

Strict liability with negligence defences: the injurer as the residual bearer
A negligence rule can also be applied in conjunction with strict liability. In these cases, the residual bearer is the injurer. Under strict liability with the defence of contributory negligence, there is a test on the victim's fault. If the victim is at fault, he/she is barred from obtaining compensation. When there is no fault on the part of the victim, the injurer is strictly liable, regardless of his/her fault. A regime of strict liability with a defence of dual contributory negligence encompasses a double test on fault; the negligence criterion is applied to both the victim and the injurer. In this case the victim bears the

accident loss only if he/she was negligent and the injurer careful: in all the remaining cases he/she is entitled to compensation. The injurer has to pay damages if both were negligent, if only the victim was careful and if both were careful. This rule generates the same incentive effects as strict liability with a defence of contributory negligence.

Under both variants of this rule of strict liability with negligence defences, the victim has incentives to take the due level of care to avoid losing his/her right to compensation in the case of an accident. If the victim is careful, the injurer bears the expected accident cost and will take the level of precaution (care and activity level) that minimizes the total cost of accidents, a level that would correspond to the socially optimal level of precaution.

Under these regimes, it is sufficient for the victim to take due care in order to be compensated for the accident loss, so that he/she does not have any incentive to take (additional) unobservable precaution. On the contrary, since the injurer is the residual bearer, he/she will have incentives to invest in both observable and unobservable precautions.

Comparative causation and loss-sharing rules

Under the liability regimes examined above, if neither party is at fault the loss is either entirely borne by the victim (negligence rules under which the victim is the residual bearer) or is shifted entirely on the tortfeasor (strict liability with negligence defences). These rules lack explicit ways for apportioning the loss between a faultless victim and a faultless tortfeasor. Historically, comparative causation emerges in the midst of legal systems based on negligence, in response to the conviction that, in the absence of fault, there is no obvious reason to let the loss fall on the innocent victim, just like there is no reason to shift it on the tortfeasor. The criterion of comparative causation allows the spreading of an accident loss among a faultless tortfeasor and an innocent victim on the basis of the relative causal contribution of the parties to the loss. The principle of comparative causation only operates as a residual basis for liability in the presence of faultless parties, avoiding the all-or-nothing allocation of liability generated by traditional rules.

In terms of levels of care, a rule of comparative causation under negligence may induce both victims and tortfeasors to adopt socially optimal levels. Comparative causation differs from traditional regimes in this respect, since both parties face positive shares of the accident loss in equilibrium. This results in the spreading of expected accident loss and activity level incentives between the parties, rather than the concentration of such losses and incentives on one or the other party. As a result, under comparative causation the activity level chosen by one party improves at the expense of the other. Thus, neither version of comparative causation dominates traditional negligence and strict liability rules on both activity-level margins.

The loss-sharing and resulting dilution of activity level incentives may or may not increase total net benefits. Loss spreading in equilibrium may promote optimal risk allocation among risk-averse agents when insurance is not readily available. Loss spreading may similarly minimize distortion of incentives deriving from truncated liability when tortfeasors face large potential losses. However, comparative causation is also likely to exacerbate administrative costs, given the need to ascertain relative causation and the need to adjudicate cases even in situations where neither party is at fault. This may explain the limited spread of this rule in contemporary legal systems.

Conclusion: the difficult design of tort law

All liability rules based on negligence struggle with a common dilemma. An increase in care level or a reduction in activity level for one party makes an accident less likely to occur. However, each party's precautions also make the accident less likely for the other party. There is no feasible and cost-effective mechanism in tort law to induce victims and tortfeasors to internalize the benefits and costs of their behaviour in all dimensions.

Tort rules can only direct efficient incentives with respect to activity level towards the residual bearer, thus failing to enhance the other party's efficient behaviour. This is Shavell's theorem on activity level (Shavell, 1980a) according to which no negligence rule exists which can give both parties efficient incentives with respect to activity level. This follows from the fact that the distinction between care (precaution the judge can observe *ex post*) and activity level (precaution the judge cannot observe *ex post*) is due to the introduction of the negligence criterion. The party who can escape liability by simply taking the due level of care will not invest in other unobservable precautions, while the other, the residual bearer, will.

This point can be generalized by observing that a point of discontinuity in the liability curves faced by the parties must be created to entice both parties to choose optimal care and activity levels. With respect to care, this is generally done by identifying a socially optimal care level and by utilizing that level to mark the boundaries between diligence and negligence. Landes and Posner (1987: 70–71) and Gilles (1992) suggest that courts take into account activity levels in their assessment of negligence whenever it is feasible to do so. However, no threshold of 'optimal activity level' is generally invoked by legal rules as a liability allocation mechanism. The reason for this omission is due to the difficulty of pinpointing a critical value to separate efficient from inefficient activity levels. Without this critical threshold, no discontinuity in the parties' expected liability can be created.

Optimal activity levels are difficult to specifiy because the value of such activities can only be ascertained from private information of the parties. Unlike optimal levels of care, which largely depend on the objective cost of

precaution and the expected gravity of harm, optimal activity levels rely on values that are harder to ascertain by a third-party decision maker since they include the subjective value of the individual that carries out the risk-creating (or risk-bearing) activity. In the absence of such a threshold it is difficult to induce both parties to internalize the full social cost of their activity levels in equilibrium.

This leads us to point out a general characteristic of tort law. Since it is not possible for both parties to bear the accident loss in equilibrium, traditional legal rules concentrate activity-level incentives on one or the other party. Negligence rules under which the victim is the residual bearer (simple negligence, contributory and comparative negligence) give efficient incentives with respect to activity level only to the victim, the residual bearer in those cases, while the strict-liability-based negligence rules (under which the injurer is the residual bearer: strict liability and strict liability with negligence defences) give efficient incentives with respect to activity level only to the injurer, the residual bearer in those remaining cases.

In theory, a rule of decoupled liability could give both parties efficient incentives with respect to care and activity levels. Decoupling liability (Polinsky and Che, 1991) means making both the injurer and the victim the residual bearers by denying the victim any compensation (as under no liability) and having the injurer pay a fine equal to compensatory damages (as under strict liability), regardless of their level of precaution. However, other functions of tort law (for example, compensatory and so on) would be compromised by such a decoupling mechanism.

A guided tour through the literature

In this section we shall provide a pathfinder through the existing law and economics literature on torts. The listings are by necessity limited to some of the more representative contributions.

Textbooks on tort law and economics

Shavell (1987) and Landes and Posner (1987) were the first systematic treatments of the topic. Although outdated, they remain the fundamental reference for tort law and economics. Miceli (2004, chs 2 and 3) provides a more recent, simple and rigorous formal treatment of the theory. Cooter and Ulen (2004, chs 8 and 9) add a discussion of contemporary issues.

The origin of the economic approach to tort law

Coase (1960) yielded an intellectual revolution in the way scholars considered the problem of externalities (accidents) in two ways. First, it put forward the reciprocal nature of accidents as both victims and injurers are to be considered as joint inputs to the externality. Hence, simply making the injurer

pay may not be the optimal solution. Second, it raised the question of why we need tort law if market exchange can do the job. Calabresi (1961, 1970), whose work may be considered as the intellectual response to the first problem, analysed different liability rules against the goals of providing incentives to reduce the total accident costs (precaution costs plus expected harm), the risk-bearing cost and the administrative costs of the system. Calabresi and Melamed (1972) provided an answer to the second problem, by arguing that tort law is needed in situations in which transaction costs prevent parties from bargaining (but the discussion on this point is open to date). Brown (1973) formalized this framework in the now standard economic model of torts.

Incentives to take precaution: the fundamental results

Under the simple assumptions of the Brown model, two main results have been derived with respect to incentives: first, Landes and Posner (1980) showed that any liability rule that features a negligence defence leads to both the injurer(s) and the victim taking optimal care; second, Shavell (1980a) proved that no such rule can induce both parties to take the optimal level of activity (defined as including all precautionary measures not explicitly included in the negligence inquiry). Gilles (1992) analysed the actual ability of American courts to include issues concerning the frequency or repetition of certain dangerous actions in the determination of negligence.

Incentives to acquire information about risk

Liability rules also serve another important goal besides those indicated by Calabresi (1970): they induce the residual bearer to acquire information in order to reduce the loss he/she bears. These incentives are distinct from the incentives to take optimal precaution. Posner (1973) raised the issue, Shavell (1992) analysed it in a formal model.

Risk allocation and insurance

Concerning the allocation of risk, the liability system is said to be comparatively more expensive than insurance, which is in general desirable even if it partially dilutes the incentives towards optimal precaution, as proven in Shavell (2000).

Administrative costs

To date, the economic analysis offers no satisfactory theory concerning the administrative cost of different liability rules. However, a particularly common rule, comparative negligence, seems to be more expensive than other likewise efficient (in the standard model) rules. This puzzling waste of administrative costs calls for scholarly attention.

Comparative negligence
The literature has moved in the direction of relaxing some of the standard assumptions. Comparative negligence seems to improve incentives when judges make random errors in comparing the due level of care to the level of care actually taken by the parties (evidentiary uncertainty, Cooter and Ulen, 1986) when the standard of care is uniform for all parties but the individual costs of care differ (Rubinfeld, 1987), and when judges err regarding the level of care cost actually borne by parties (Haddock and Curran, 1985). Bar-Gill and Ben-Shahar (2003) criticize part of this approach. The literature is vast and the academic discourse remains open.

Comparative causation
While the apportionment of losses among negligent actors (comparative negligence) is a reality that is difficult to explain, apportioning damages among non-negligent is a profitable solution that is rarely found in actual legal systems, as argued in Parisi and Fon (2004).

Errors, uncertainty and accuracy
The incentive effects of liability rules crucially depend on their correct implementation. Errors or uncertainty in the determination of the damage award, the causal link, or the issue of negligence might distort incentives. See Diamond (1974), Calfee and Craswell (1984) and Craswell and Calfee (1986). However, when this does not happen, a certain degree of inaccuracy might help save administrative costs. See Kaplow and Shavell (1994, 1996).

Insolvent and disappearing injurers
If the injurer's assets are not sufficient to cover the victim's compensation or if there is a chance that the injurer will not be identified or sued, incentives to take precaution might be diluted. Summers (1983) identified this problem; Shavell (1986) analysed it in connection with insurance; Dari Mattiacci and De Geest (2005) show that the level of precaution that the injurer takes depends on the precaution technology available.

Vicarious liability and mandatory insurance
In response to the two problems mentioned above, liability may be shifted from the insolvent or disappearing injurer to his/her principal in order to give the latter incentive to exercise a delegated control function on the former. Sykes (1981) and Kornhauser (1982) provided the first analysis. Mandatory insurance may serve the same purpose. Dari Mattiacci and Parisi (2003) analyse different such systems of delegated control in a unitary framework.

Punitive damages
If injurers can escape liability, their incentives may be diluted. Their incentives may be corrected by increasing the damages they pay when they are actually apprehended. Punitive damages are hence seen as a corrective mechanism for disappearing injurers. See Cooter (1982) and Polinsky and Shavell (1998).

Pure economic loss
Some losses consist of the victim's forgone profits. In this case, the market mechanism might cause a third party (for example, a competitor of the victim) to increase his/her profits as a consequence of the accident. It has been said that such losses do not correspond to a socially relevant loss, hence the victim does not deserve compensation. The issue is approached differently in different legal systems and is often discussed. The issue originated from Bishop (1982) and Rizzo (1982). Bussani et al. (2003) provide a comparative analysis.

Non-pecuniary loss and compensation for pain and suffering
The economic model of torts is based on the assumption that the victim's loss (as well as the parties' precaution costs) may easily be expressed in monetary terms and that monetary compensation can restore the victim to the pre-accident situation. Both assumptions are often not satisfied, raising the two related problems of whether and how much compensation to award, also in connection with the injurer's incentives. See Arlen (2000, section B), for a survey of both the empirical literature and the economic arguments in favour of and against compensation of such loss.

Product liability
Accidents that occur in connection with products are of a different type from ordinary torts. In fact, the victim (the consumer) and the injurer (the producer) are parties to a contractual relationship. The Coase theorem applies and, if its conditions are satisfied, the liability rule is irrelevant to the outcome, as parties will bargain around it. However, producers will in general enjoy an informational advantage compared to consumers and, hence, the liability rule might make a substantial difference. Strict liability is in general preferred as it puts the informational burden on the producer (see also subsection on incentives to acquire information). See Spence (1977).

Joint and several liability
When more than one injurer is responsible for the loss suffered by the victim, a problem arises of how to apportion damages among them. In the standard model, the apportionment rule does not affect the outcome (Landes and Posner, 1980). However, in connection with insolvency and the possibility of

settling out of court, the problem becomes relevant and the rules that govern the apportionment of the loss affect the injurers' incentives. See Kornhauser and Revesz (1990).

Causation

The issue of causation is controversial in economics, as both the victim and the injurer may be seen as joint inputs in the production of the accident loss in a Coasean perspective. The analysis has mainly focused on ascertaining the effects of causation on the functioning of the negligence rule (Grady, 1983; Kahan, 1990), on the allocation of damages when there is uncertainty over the causal contribution of several injurers (Shavell, 1985), and on the optimal restriction of the scope of liability (Shavell, 1980b).

Tort liability and regulation

Tort liability as a way of producing incentives to optimal precaution may be compared to the regulatory system, which serves the same purpose. Regulation may substitute or complement tort liability when the latter is impaired by problems related to an insolvent or disappearing defendant. Wittman (1977), Shavell (1984a, 1984b), Kolstad et al. (1990), Burrows (1999) and Schmitz (2000).

Litigation

Liability as a system of providing parties with incentives to take precaution relies on the enforcement of the duty to pay damages. The way in which the judicial system functions may affect the incentives produced by tort liability as it affects the victims' ability to collect from injurers and hence the injurers' internalization of the victims' loss. A survey of the ongoing research on the topic may be found in Cooter and Rubinfeld (1989) and Kobayashi and Parker (2000).

History and evolution of tort liability

Economic analysis may also be applied to the study of the genesis and evolution of tort liability in response to changes in society and technology, which in turn affected the nature and the probability of accidents. Posner (1980 and 1981) and Parisi (1992, 2001).

References

Arlen, Jennifer (2000), 'Tort damages', in Boudewijn Bouckaert and Gerrit De Geest (eds), *Encyclopedia of Law and Economics*, vol. II, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 682–734.

Bar-Gill, Oren and Omri Ben-Shahar (2003), 'The uneasy case for comparative negligence?', 5 *American Law and Economics Review*, 433–69.

Bishop, William (1982), 'Economic loss in tort', 2 *Oxford Journal of Legal Studies*, 1–29.

Brown, John Prather (1973), 'Toward an economic theory of liability', 2 *Journal of Legal Studies*, 323–49.
Burrows, Paul (1999), 'Combining regulation and legal liability for the control of external costs', 19 *International Review of Law and Economics*, 227–44.
Bussani, Mauro, Vernon Palmer and Francesco Parisi (2003), 'Liability for pure financial loss in Europe: an economic restatement', *American Journal of Comparative Law*, **51**, 113–63.
Calabresi, Guido (1961), 'Some thoughts on risk distribution and the law of torts', 70 *Yale Law Journal*, 35–46.
Calabresi, Guido (1970), *The Costs of Accidents*: *A Legal and Economic Analysis*, New Haven, CT: Yale University Press.
Calabresi, Guido and A. Douglas Melamed (1972), 'Property rules, liability rules and inalienability: one view of the cathedral', 85 *Harvard Law Review*, 1089–28.
Calfee, John E. and Richard Craswell (1984), 'Some effects of uncertainty on compliance with legal standards', 70 *Virginia Law Review*, 965–1003.
Coase, Ronald H. (1960), 'The problem of social cost', 3 *Journal of Law and Economics*, 1–44.
Cooter, Robert D. (1982), 'Economic analysis of punitive damages', 56 *Southern California Law Review*, 79–101.
Cooter, Robert D. and Daniel L. Rubinfeld (1989), 'Economic analysis of legal disputes and their resolution', 27 *Journal of Economic Literature*, 1067–97.
Cooter, Robert D. and Thomas S. Ulen (1986), 'An economic case for comparative negligence', 81 *New York University Law Review*, 1067–110.
Cooter, Robert D. and Thomas S. Ulen (2004), *Law and Economics*, 4th edn, Reading, MA: Addison-Wesley.
Craswell, Richard and John E. Calfee (1986), 'Deterrence and uncertain legal standards', 2 *Journal of Law, Economics, and Organization*, 279–303.
Dari Mattiacci, Giuseppe and Gerrit De Geest (2005), 'Judgment proofness under four different precaution technologies', *Journal of Institutional and Theoretical Economics*, **161**, 1–19.
Dari Mattiacci, Giuseppe and Francesco Parisi (2003), 'The cost of delegated control: vicarious liability, secondary liability and mandatory insurance', *International Review of Law and Economics*, **23**, 453–75.
Diamond, Peter A. (1974), 'Single activity accidents', 3 *Journal of Legal Studies*, 107–64.
Gilles, Stephen G. (1992), 'Rule-based negligence and the regulation of activity levels', 21 *Journal of Legal Studies*, 319–63.
Grady, Mark F. (1983), 'A new positive economic theory of negligence', 92 *Yale Law Journal*, 799–829.
Haddock, David D. and Christopher Curran (1985), 'An economic theory of comparative negligence', 14 *Journal of Legal Studies*, 49–72.
Kahan, Marcel (1989), 'Causation and incentives to take care under the negligence rule', 18 *Journal of Legal Studies*, 427–47.
Kaplow, Louis and Steven Shavell (1994), 'Accuracy in the assessment of liability', 37 *Journal of Law and Economics*, 1–15.
Kaplow, Louis and Steven Shavell (1996), 'Accuracy in the assessment of damages', 39 *Journal of Law and Economics*, 191–209.
Kobayashi, Bruce and Jeffry S. Parker (2000), 'Civil procedure: general', in Boudewijn Bouckaert and Gerrit De Geest (eds), *Encyclopedia of Law and Economics*, vol. V, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 1–26.
Kolstad, Charles D., Thomas S. Ulen and Gary V. Johnson (1990), 'Ex post liability for harm vs. ex ante safety regulation: substitutes or complements?', 80 *American Economic Review*, 888–901.
Kornhauser, Lewis A. (1982), 'An economic analysis of the choice between enterprise and personal liability for accidents', 70 *California Law Review*, 1345–92.
Kornhauser, Lewis A. and Richard L. Revesz (1990), 'Apportioning damages among potentially insolvent actors', 19 *Journal of Legal Studies*, 617–51.
Landes, William M. and Richard A. Posner (1980), 'Multiple tortfeasors: an economic analysis', 9 *Journal of Legal Studies*, 517–55.

Landes, William M. and Richard A. Posner (1982) 'The positive economic theory of tort law', 15 *Georgia Law Review*, 851–924.
Landes, William M. and Richard A. Posner (1987), *The Economic Structure of Tort Law*, Cambridge, MA: Harvard University Press.
McEwin, R. Ian (2000), 'No-fault compensation systems', in Boudewijn Bouckaert and Gerrit De Geest (eds), *Encyclopedia of Law and Economics*, vol. II, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 753–63.
Miceli, Thomas J. (2004), *The Economic Approach to Law*, Stanford, CA: Stanford University Press.
Parisi, Francesco (1992), *Liability for Negligence and Judicial Discretion*, 2nd edn, Berkeley, CA: University of California Press.
Parisi, Francesco (2001), 'Genesis of liability in ancient law', 3 *American Law and Economics Review*, 82–124.
Parisi, Francesco and Vincy Fon (2004), 'Comparative causation', *American Law and Economics Review*, **6**, 345–68.
Polinsky, A. Mitchell and Yeon-Koo Che (1991), 'Decoupling liability: optimal incentives for care and litigation', 22 *Rand Journal of Economics*, 562–70.
Polinsky, A. Mitchell and Steven Shavell (1998), 'Punitive damages: an economic analysis', 111 *Harvard Law Review*, 869–962.
Posner, Richard A. (1973), 'Strict liability: a comment', 2 *Journal of Legal Studies*, 205–21.
Posner, Richard A. (1980), 'A theory of primitive society, with special reference to law', 23 *Journal of Law and Economics* 1.
Posner, Richard A. (1981), *The Economics of Justice*, Cambridge, MA: Harvard University Press.
Rizzo, Mario J. (1982), 'A theory of economic loss in the law of torts', 11 *Journal of Legal Studies*, 281–310.
Rubinfeld, Daniel L. (1987), 'The efficiency of comparative negligence', 16 *Journal of Legal Studies*, 375–94.
Schmitz, Patrick W. (2000), 'On the joint use of liability and safety regulation', 20 *International Review of Law and Economics*, 371–82.
Shavell, Steven (1980a), 'Strict liability versus negligence', 9 *Journal of Legal Studies*, 1–25.
Shavell, Steven (1980b), 'An analysis of causation and the scope of liability in the law of torts', 9 *Journal of Legal Studies*, 463–516.
Shavell, Steven (1984a), 'Liability for harm versus regulation for safety', 13 *Journal of Legal Studies*, 357–74.
Shavell, Steven (1984b), 'A model for optimal use of liability and safety regulation', 15 *Rand Journal of Economics*, 271–80.
Shavell, Steven (1985), 'Uncertainty over causation and the determination of civil liability', 28 *Journal of Law and Economics*, 587–609.
Shavell, Steven (1986), 'The judgment proof problem', 6 *International Review of Law and Economics*, 45–58.
Shavell, Steven (1987), *Economic Analysis of Accident Law*, Cambridge, MA: Harvard University Press.
Shavell, Steven (1992), 'Liability and the incentives to obtain information about risk', 21 *Journal of Legal Studies*, 259–70.
Shavell, Steven (2000), 'On the social function and the regulation of liability insurance', 25 *Geneva Papers on Risk and Insurance: Issues and Practice*, 166–79.
Spence, A. Michael (1977), 'Consumer misperceptions, product failure and product liability', 64 *Review of Economic Studies*, 561–72.
Summers, John (1983), 'The case of the disappearing defendant: an economic analysis', 132 *Pennsylvania Law Review*, 145–85.
Sykes, Alan O. (1981), 'An efficiency analysis of vicarious liability under the law of agency', 91 *Yale Law Journal*, 168–206.
Wittman, Donald (1977), 'Prior regulation versus post liability: the choice between input and output monitoring', 6 *Journal of Legal Studies*, 193–212.

7 Family

Margaret Brinig

Introduction

The idea of thinking of families in economic terms is not new, but dates back at least to the time of Aristotle, whose *Econœmeia* meant 'management of the household' (Spiegel, 1983, p. 25), and whose views on affection between the generations are cited even today by law and economics scholars such as Richard Posner (Posner, 1996). The patriarchal family was used as a metaphor for the monarchy by William Filmer in the seventeenth century (Filmer, 1653), a theory debunked by John Locke, who, in writing his *Second Treatise on Government* (Locke, 1689, pp. 179–87), laid out much of the foundation for the later work of William Blackstone, a lawyer, whose contractarian notion of the implicit contract between parent and child appears in much of the law and economics literature on family relationships (Blackstone, 1765). Similarly, David Hume's (1761) writing about the need for marital exclusivity, particularly on the part of the wife, sounds an economic chord, for he bases his suggestion on the requirement of the husband to support the wife's offspring. Another contractual writing about the family comes from Sir Matthew Hale, a British jurist whose statement about the impossibility of spousal rape stemmed from the wife's having, by her marriage, given an irrevocable consent to intercourse with her husband. The nineteenth-century economic writings of British and American authors Harriet Martineau (1889) and Catharine Beecher (1841) relate women's participation in the home economy to their political and social roles. Beecher's *A Treatise on Domestic Economy* (1841) became the primer for the adoption of the married woman's 'separate sphere', in which she was to specialize in the education and upbringing of children while her husband toiled in the labour market.

In the twentieth century, family law suffered from an intellectual stigma that probably came from the tawdriness of the staged suits for fault divorce and the equally unsavoury 'heartbalm actions' of breach of promise, seduction and alienation of affections. Except for the forays of some members of the Chicago school in the late 1970s, writing about a market in babies (Landes and Posner, 1978) and the economic basis for alimony (Landes, 1978), family law lagged behind other legal fields in developing economic scholarship. Recently, however, quite a few pieces and several books have featured the law and economics of the family, particularly involving marriage and divorce.

Family law is perhaps different from other areas because of its tremendous social importance and the ever-presence of externalities, in the form of children. Although there has been very little work published that systematically looks at all of family law (Brinig, 2000), one might begin with a theoretical construction such as that shown in Table 7.1.

Table 7.1 A systematic representation of family law

Forms			Children not present	Children present
Social unit	Legal analysis	Economic		
Individuals	Contract paradigm	Consumer economics	Courtship and marriage creation	Fertility and adoption
		Economics of the firm	Husband and wife	Parent and child
Families	Covenant paradigm	Law and economics of franchise	Divorcing couples	Emancipation
				Divorcing couples with children
				Adoptive and birth families

A complete study of the law and economics of the family could thus be organized along three well-known economic models. Some family relationships, which might be called 'pre-families', can be described using a consumer economics model. When this model sets the appropriate theoretical framework, law and economics scholars suggest that unregulated, or private, contracts ought to be encouraged, unless one or both of the contracting parties is incompetent, conditions suggest information asymmetries, the situation involves substantial negative externalities or rent extraction (or hold-up) is likely to take place (Epstein, 1995). In these exceptional circumstances, the most efficient contracting cannot take place. Therefore circumstances require some type of governmental intervention in the marketplace.

This analysis indicates a free marriage market except where there are substantial information problems (annulment) or substantial negative externalities (void marriages or unenforceable agreements, both of which occur because the social benefits of families are threatened (Brinig and Alexeev, 1995). A contract will be closely scrutinized when it frequently leads to rent extraction (as with antenuptial agreements, where courts are particularly

concerned with inequality of bargaining power: Schultz, 1980). Before children become part of a family, externalities are always present because any contracts are made between adults, while at the same time they affect the children. Moreover, the state has interests because of the importance of parenting. Much of the litigation involves problems of information (about what a birth parent is giving up, or what an adoptive parent is receiving) and market power (wealth extraction by third parties: Brinig, 1995a). Some involves the costs of transacting adoptions. Other externalities (commodification) are particularly feminist concerns, as are problems with the 'pricing' or sale of children (Radin, 1987), because of incommensurability (inability to value or measure: Sunstein, 1994).

Once the family has been formed, the relationships may be analogized to the economic problem faced by the firm. The closest legal relationship will not be contract, but covenant (Brinig, 1994a, 2000). The goal of a firm's profit parallels the individual's utility. The factors of production critically involve transaction costs as well as the costs of externalities. The principal–agent problem poses the greatest transaction cost difficulties for the firm (Ross, 1973). The modern firm particularly requires investments in capital. In the family, profit is seen by mainstream economists as a maximization of household production (Becker, 1991). More recently, feminist economists challenge the appropriateness (and accuracy) of that goal and suggest alternatives of happiness, intimacy and security (Singer, 1995). Becker's household production model fails to take account of all things involving the family, and particularly ignores what makes up people's tastes or preferences. The household production function may be re-examined, with particular attention paid to transaction costs and also to the specialization and the division of labour.

The family firm can also be examined in the context of children. A prominent subject for economic analysis and, recently, law and economic analysis, involves investment in human capital, the knowledge and skills that will build future citizens, workers and social beings (Becker, 1975). How these investments are made – whether by father, mother or state – is of tremendous social importance as well as interest here.

When families change through ageing, adoption, separation or divorce, two models appear (Brinig, 1996). Choice between them will dominate law reform choices. One model springs naturally from an emphasis on contract: this is the *sovereign nation* model (Kronman, 1985). The other is an outgrowth of the principal–agent discussions found in the covenant or firm: the *franchise* model (Hadfield, 1990). Dissolution of marriage, adoption and emancipation might constitute a return to contractualization of the family. But can a legal decree, or a statute, really terminate a permanent relationship? What are the consequences of doing so? Transaction costs, bargaining power and equality come to the forefront again, as do the question of externalities

(in terms of the effect on children and the stability of relationships themselves) and the extraction of quasi-rents.

A final concern might surround the role of law and lawyers. Legal academics have long puzzled over which comes first, social change or law reform. What is not generally discussed is the question of what happens if the law does not accurately reflect what is empirically true. An example of this problem in the family law context is the duties and obligations of fathers: the legal consequences flow from the biological relationship, while recent sociological evidence shows that emotional allegiance follows the man's attachment to his current mate. Thus, if law were to follow inclination, support and custodial responsibilities would be owed by the stepfather, not the birth father, following divorce. A final set of questions revolves around the goals society sets for families. Should we return to an older (more covenantal) rule of law, with the hope that society will follow? For example, we may question whether laws should be drawn (or reformed) to make divorce more difficult, to make care for the elderly a family, rather than a societal or individual, responsibility, or to discourage out-of-wedlock child bearing.

There are thus three main parts of a law and economics of the family, roughly dividing family law into studies of market behaviour (family formation), the family firm (functioning families or husband and wife and parent and child), and the family as franchise (post-divorce and post-emancipation relationships). At the beginning of relationships, market principles dominate, while after families are formed the better paradigm is covenant, for franchise relationships may be described as the vestiges of covenant. The law and economics of the family should be constant through various countries and economies. The transition economies of Europe, while perhaps having a different property tradition from that of the English-speaking countries, have the same family structures. As the transitional European countries choose or reform their family laws, it may be helpful to look at the experience of other Western nations, particularly in terms of divorce and custody laws.

The marriage market

When we examine what we might call the marriage market, we find that economists have already written, as with Stigler's (1961) economics of information, about the search for an ideal mate. Although Telser (1927) coined the term 'assortative mating', England and Farkas (1986) describe a 'D' or desirability factor that the individual uses to assess his or her own chances in the marriage market. The young adult holds each potential spouse up to this standard, deciding whether to settle down with a particular person or continue to search. Allen (1992b) notes that people will marry realizing that they will share during their marriage relationship, and therefore will be attracted to people much like themselves. Bishop (1984) writes about the signalling

effect of marriage on others who remain in the market. Becker et al. (1978) discuss courtship in terms of its impact upon divorce: when a person cannot continue to search because, for example, she already has a child, she is likely to make a less successful match. Similarly, mismatches in terms of religion or education also lead to a greater probability of divorce. Brinig and Alexeev (1993) further develop the search concept, writing about the 'lemons' problem that may occur because some traits of a prospective spouse cannot be discovered until after marriage. They note that annulments on the ground of fraud, which are granted when the deception goes to the essence of the marriage relationship, increase when the divorce law eliminates consideration of fault. Brown (1995) discusses the impact of the legalization of same-sex marriages upon the economy of Hawaii, an American state that seemed a good candidate to be the first in performing them. Baker and Emery (1993) doubt that couples accurately assess the probabilities of their own divorce when they contemplate marriage. They are therefore unable to plan accurately in terms of marital contracts. Brinig (1990) discusses the use of diamond engagement rings as bonding devices when states abolished the breach of promise action, which had equalized the losses between men and women from a broken engagement. Schultz (1982) discusses the pros and cons of a complete marital contingent claims contract. Contracting, while desirable for many reasons, itself may inhibit the trust and intimacy that should develop in a marriage (Singer, 1992).

There has been some recent work on precommitment strategies, in terms of choosing, at the time of marriage, whether to accept the prevalent no-fault divorce regime or specify a more stringent, fault-based private rule, in the couple's own case. For example, Scott (1990) argues that, if the couple plan to invest heavily in their relationship, for example by having children, they ought to be able to opt for a more difficult exit option. Likewise, Stake (1992) would require the choice of a more or less onerous set of marital obligations. Marriage involves a set of state-required default terms, which cannot be varied even at the parties' behest, seemingly contrary to a liberal view of contract. These may exist because they are the sort of terms couples would not think of at the time of the marriage ceremony, while they are nevertheless important (Baker and Emery, 1993). They may also be seen as state provision for children, who are incapable of contracting (particularly before the fact) on their own behalf (Becker and Murphy, 1988).

In the last several years, some feminist writers (not from a law and economics tradition) have advocated a re-evaluation of marriage, with the idea that movement towards the centrality of the mother–child dyad might better provide for the societal needs of child rearing while compensating for the dependency that is inevitable when people take care of children or elderly parents (Dowd, 1994; Fineman, 1995). Brinig and Buckley (1999b) criticize

such ideas because they do not account for the complementary functions that two parents perform in families.

Becoming parents: the market for children

Scholars with interdisciplinary interests are increasingly turning to the new field of sociobiology, which explains human behaviour, in important part, in terms of people's desire to maximize the success of their own genetic heritage. Popular accounts which are important for students of the family include Ridley (1992) and Wright (1994). The biologists were not the first to discuss this important ingredient of tastes. For example, Anderson and Tollison (1991) write about children's behaviour, which they describe as rational because each child attempts to get as much parental attention and care as possible. Even earlier, Gary Becker and others (Becker and Lewis, 1973; Becker and Tomes, 1976) have described the interaction between the number of children and the time or goods that may be devoted to each child, hypothesizing that the available resources parents can allocate are fixed.

But even before parents can invest in their children, they occasionally must bring children from other families into their own through adoption or, more recently, surrogacy. Landes and Posner (1975) wrote a controversial essay suggesting that adoption laws, which forbid any explicit payment for children, be relaxed so that the large and inelastic demand for children can be met by a larger supply of adoptable children. Many unexpected pregnancies are terminated through elective abortions in countries where the latter are legal. Other law and economics scholars have supported this and later Posner work on adoption (Posner, 1987, 1992), urging that a legal 'market in babies' be extended to surrogacy arrangements, in which a married father's sperm are implanted in a woman other than his wife who bears the child for the married couple (Cass, 1987; Epstein, 1995). Cass, while realizing possible objections to surrogacy arrangements, noted that surrogacy allows some relief for couples who increasingly find it difficult to adopt, and make parties on both sides of the transaction happy. Epstein suggested specific enforcement of surrogacy contracts since damage remedies would probably not be effective and since certainty would benefit all the parties to the contract. Prichard (1984), however, disapproved of a 'market in babies' for largely philosophical reasons, while Trebilcock and Keshvani (1991) approved of surrogacy in general, but only if the contracts involved a waiting period after the child's birth before consent would be final. Brinig (1995a) suggests that surrogacy contracts should not be specifically enforced because of information asymmetries, third-party externalities (the effect on the surrogate's other children) and market asymmetries allowing middlemen to exploit parties on both sides of the contract. In other papers, Brinig (1993, 1994d) discusses the negative effect on the adoption rate of increasing transaction costs by allowing birth

mothers to revoke consent. Like Cass (1987), Prichard (1984) and J. Cohen (1987), she is concerned about the effect of surrogacy on hard-to-place children who currently can expect lengthy placement in foster homes.

Foster families are technically hybrids between biological families and paid child carers. They therefore involve particularly complex principal–agent problems (Schneider and Brinig, 1996b). Because foster parents are not guaranteed permanent relationships with the children for whom they care, they do not have the incentives to treat them as well as they would their own, as Grubb describes with the historical practice of indentured servitude (Grubb, 1985). However, because many foster parents are, in effect, repeat players, meaning that they have several sets of foster children, they have reputation effects to consider. This may cause their behaviour to remain within the acceptable level, but at the same time may keep them from forming strong psychological bonds with the foster children. Further, particularly in the minority community, they are often blood relatives of their foster children.

The family firm: marriage

Once people marry, in many ways their individualistic, or market-like, behaviour changes so that it more resembles behaviour within the firm. They are concerned with what Becker (1991) calls 'maximization of household production', which broadly includes not only children, consumption goods and leisure time to enjoy them, but also values like intimacy and love. As with commercial firms, there will be no day-to-day accounting: as Ben-Porath (1980) noted, large outstanding balances will be tolerated. As within the firm, specialization will become important (Becker, 1991). The balance between household and labour force participation will change with changes in the spouses' opportunity costs (England and Farkas, 1986) though, as Becker (1991) suggests, women for biological reasons always have the comparative advantage in child rearing and, because child care and other household jobs may be done at the same time, may be the only spouse to work in both the labour force and household production. Legally, the paradigm is the covenant, far more difficult to dissolve than the ordinary contract and requiring state action (divorce) to do so (Brinig, 1993, 1994a, 1996).

During marriage, couples remain free to invest in their own human or other capital, but more often will choose to invest specifically in the marriage or in each other's careers. Specific investment in marriages includes retailoring cooking, cleaning or entertainment tastes to suit the other spouse (Lloyd Cohen, 1987). More importantly, it includes the creation and raising of children and the mingling of assets (Zelder, 1993; Brinig and Crafton, 1994). Fuchs (1992) suggests that women may assume more of these roles even in an otherwise egalitarian marriage because they, more than their husbands, prefer children. Investment in the other spouse's human capital is increas-

ingly important, and has become the focus for much litigation on divorce (Parkman, 1992; Brinig, 1997). In most marriages couples also specialize, and frequently the specialization is along gendered lines (Becker, 1991; Brinig, 1994c). Although most married women in the United States currently remain in the labour force, they continue to assume a disproportionate share of the household production functions (Manser and Brown, 1980) and are particularly involved with the rearing of children (Zick and Geurer, 1991). Men, on the other hand, tend to specialize in jobs outside the home, including not only labour force participation, but also work on the home exterior, the lawn and the family car. This specialization has the effect of making divorces more costly to women, who typically cannot carry enhanced earning capacity outside the marriage and into other relationships (Landes, 1978; Lloyd Cohen, 1987). Children, who possess some of the attributes of public goods, make some divorces 'inefficient' because their relationships with non-custodial parents will necessarily be different following separation (Zelder, 1993).

Marriage also functions as insurance (Becker, 1993; Scott and Scott, 1998). It protects the spouses against the vagaries of the marriage market and also from possible setbacks in the labour force. How important this function is may depend upon the spouses' relative risk aversion. Evidence that suggests that women are more risk-averse than their husbands points to reasons why they may tend to remain in a less than happy marriage, as well as the amount they are likely to invest in the relationship, rather than just in themselves (Brinig, 1995b; Parkman, 1996).

A good deal of recent work involves bargaining within the marriage setting. For example, Manser and Brown (1980) note that the threat of divorce, if credible, may encourage a shirking spouse to perform marital duties. More recently, Lundberg and Pollack (1993) write that, instead of divorce, many quarrelling couples are likely to reach a stalemate in which each spouse will perform the minimum expected, with a resulting distribution of labour along very traditional gender lines. Bergstrom (1995) does the same type of analysis from a game-theoretic approach. Allen and Brinig (1998) extend bargaining to consensual sexual intercourse, with a 'property right' belonging to the spouse who is least interested in coitus at any given time during the marriage. As the property right shifts from wife to husband, more adultery and divorces are likely to result.

The family firm: parent and child

Phillipe Aries (1962), the noted historian of childhood, linked parental investment in children to the infant mortality rate. Thus parents could not afford, emotionally or otherwise, to become attached to children, or to allow childhood to develop as a separate stage of life, until most children would live to adulthood, becoming economic advantages for their parents. The British

jurist Blackstone (1765), writing in the eighteenth century, writes of an implicit contract between parents and children, one in which parents supply the child's material wants and educate the child (morally as well as intellectually) in return for current obedience and wages and, later in life, support in the parents' old age. This implicit contract becomes the basis for the legal system which protects the child during minority, gives the parent control over education, discipline and training, promotes the 'best interests of the child', allows the parent to consent to the child's marriage before majority, and places the duty of the aged parents' support on the adult child. Rubin et al. (1979) tie this set of reciprocal duties to holdings in land, supposing that there will be more parental investments in a child's human capital as eventual inheritance of the family farm stops being an important inducement to adhere to the implicit contract. Because their children cannot adequately fend for themselves, parents act as fiduciaries for them (Scott and Scott, 1995), with many of the principal–agent problems that relationship typically produces (Cooter and Freedman, 1991).

Once the children are born, as rational beings they manoeuvre to increase their share of the scarce resources of parental attention (Anderson and Tollison, 1991). The state assumes some of the functions that are difficult for parents to perform if they expect, as part of an implicit contract, to be provided for in their own old age by the children they now raise (Becker and Murphy, 1988). Not only the state, but parents as well, invest in the children's specific capital, attempting to forge bonds that will reward the parent in dotage (Becker, 1993; Brinig, 1994b). On the other hand, if they do not expect to have adult relationships with their children, parents may dissipate this most valuable human resource through abuse and neglect (Brinig and Buckley, 1999b). Cox and Stark (1993) say that parents may use their own behaviour towards ageing parents to model the way they would like to be treated by their children.

Alternatively, some of what parents do may be seen as a biological response to the drive towards prolonging the species. Epstein (1989) uses biology to account for differences in child rearing by men and women as well as the investments parents make, while Bergstrom (1995) uses it to explain the difference in felt duty towards children versus other relatives, and Jones (1997) explains step-parent abuse through the absence of a biological tie with the child in question. Becker, in a series of famous papers, writes of the tradeoff between time or other resources parents spend on children and the number of children a family has (Becker and Lewis, 1973; Becker and Tomes, 1976). Two recent papers cast some doubt that biology can drive all of the parental relationship. Seltzer (1998) describes fathers' relationships with children as being dependent on the man's relationship with the children's mother: once the father begins a new relationship, it is that woman's children that are

most important to him, sometimes to the exclusion of the man's own offspring. Likewise, a number of family law scholars, including Woodhouse (1992, 1996) and Bartlett (1984), write of parent substitutes who assume the reciprocally fulfilling roles of parents in children's lives. Lupu (1994) and Brinig (1995a) describe parents as complementary factors in the children's upbringing: Lupu from a political science (balance of powers) perspective; Brinig from an economic one.

Changing families

The firm model for families is necessarily incomplete. It does not, for example, explain the legal and other ties that exist after families are legally dissolved through adoption, emancipation or divorce (Brinig, 1996). Because the love that flows between family members is unconditional and permanent, divorce becomes difficult and frequently painful. What may be left after the law pronounces family relationships ended is akin to the economic concept of a franchise, while what people involved in such transitions may experience is an attempt to return to the market unfettered, again thinking of their relationships contractually. Couples divorce, according to Becker et al. (1978), when the gains from remaining married are outweighed by the anticipated benefits of returning to the single life. Frequently, this occurs when couples have married in relative haste and are therefore surprised by negative traits in each other. Sometimes it occurs even after a more lengthy marriage when one spouse can extract the quasi-rents of the other (Lloyd Cohen, 1987) or the other does not have enough to offer to keep the marriage viable (Zelder, 1993; Allen and Brinig, 1998).

The divorce rate, after rising steadily throughout the twentieth century, exploded during 1960–80 in the United States, and has increased in parallel, though usually not at such a high level, throughout most of the world (Goode, 1993). Allen (1990) notes that no-fault divorce statutes seem to increase across cultures as more married women participate in the paid labour force, and South (1985) indicates the relationship between divorce and lack of strong social cohesion in the party's home state. Scholars debate whether no-fault divorce itself causes more divorces, though there is little doubt that the rates increase around the time no-fault is introduced (Nakoneszy et al., 1995). Peters (1986) used a panel data study to suggest that, under the Coase theorem, unhappy spouses would reach an efficient level of divorce. The relaxation of divorce grounds would negatively affect women, who would be forced to trade tangible property rights to keep a marriage together. Peters's results were disputed for a variety of reasons by Duncan and Hoffman (1981), Allen (1992a) and Brinig and Buckley (1998a), all of whom suggest that, because divorce transactions costs decreased when no-fault was introduced, there ought to be a permanent effect on the divorce rate. In no-fault states, divorces

cost far less and there is generally much less litigation (Brinig and Alexeev, 1994), as the Priest–Klein (Priest and Klein, 1984) model would predict. Mediation has become increasingly prevalent as an alternative to litigated awards, and has the advantage, not only of being less costly, but also of allowing the parties to reach freely their own settlements (Brinig, 1995b).

Peters (1986) predicted substantial differences in divorce settlements before and after no-fault, and her fears were echoed in the sociological literature (Weitzman, 1981) and the legal literature (Mnookin and Kornhauser, 1979). In fact, many divorced women with children do live in poverty, but perhaps mostly because of the loss of economies of scale as the marital household divides in two (Duncan and Hoffman, 1981). Despite the great variety in American divorce laws, the settlements people receive in terms of child custody and financial assets seem quite constant across states (Maccoby and Mnookin, 1992; Brinig and Alexeev, 1993; Fox and Kelly, 1993; Garrison, 1993; Weiss and Willis, 1993; Brinig and Buckley, 1998a).

There seems to be little doubt about the adverse effects of divorce on children, both psychologically (Hetherington et al. 1982; Wallerstein, 1991) and financially (Weiss and Willis, 1985; Brinig and Buckley, 1998b). Whether because they are forming new relationships, as Seltzer (1998) suggests, or because they can no longer monitor how their former wives are spending child support contributions (Weiss and Willis, 1993), or whether time with children has some of the properties of an addictive good (Becker and Murphy, 1988), more than half the child support payments ordered in the United States remain uncollected.

Although the parents may be more comfortable ending their relationship, the contracting that they do has the negative external effect typical of a Kaldor–Hicks rather than a Pareto-optimal solution. Perhaps for this reason, divorce laws may be more stringent for couples with children (Scott, 1990) or parents may be ordered to pay for children's college education, as they are not while married. The other troubling external aspect of divorce is its effect on other marriages. As the divorce rate approaches 50 per cent, many couples can rationally expect an end to their relationship. Husbands, who are typically in a better financial position, may use their advantage to extract quasi-rents from their wives (Lloyd Cohen, 1987; Brinig and Crafton, 1994). Some couples may choose not to marry at all, while in others women may work 'too hard' (Parkman, 1996). To some extent, these problems were relieved by the institution of alimony (Landes, 1978). As states make permanent alimony the exception rather than the rule, spouses, at least theoretically, may engage in 'trading' of custodial time with financial assets (Mnookin and Kornhauser, 1979). Brinig and Alexeev (1993) note that, in the states they studied, people seemed to bargain towards a standard custody pattern, with men giving up property until reaching about a 25 per cent custody share, and

afterwards 'paying' not to have more time with the children. This result seems to be true nationwide (Brinig and Buckley, 1998b) even when state statutes presume an even division of custody (Maccoby and Mnookin, 1992). Child custody adjudications are difficult and, like the termination of parental rights that takes place if parents abuse their children, pose all the economic problems of the rules and discretion debate (Kydland and Prescott, 1977) and the confusion introduced when it is difficult to identify which error is of Type I (Schneider and Brinig, 1996b) (the error the state most wants to avoid).

In relationships with joint custody, child support seems to be easier to collect (Maccoby and Mnookin, 1992; Fox and Kelly, 1993), suggesting that, if the relationship with the non-custodial parent continues to be strong, not only will the hoped-for psychological stability remain, but also the child will be better off financially. Similarly, although adoption legally terminates relationships with birth parents, many family members seem reluctant to declare the ties severed. Grandparents, for example, are litigating to keep visitation rights at an ever-increasing rate, as are foster parents, natural siblings of the adopted child and even birth mothers (Brinig, 1996). Some of this litigation is no doubt produced by the monitoring problems Weiss and Willis (1985) discuss. Some may be simply an attempt to preserve the family franchise: the reputation and other assets of the extended relationship (Brinig, 1996).

References

Allen, Douglas (1990), 'An inquiry into the state's role in marriage', *Journal of Economic Behavior and Organization*, **13**, 171–91.

Allen, Douglas (1992a), 'Marriage and divorce: a comment', *American Economic Review*, **82**, 679–85.

Allen, Douglas (1992b), '"What does she see in him?" the effect of sharing on the choice of spouse', *Economic Inquiry*, **11**, 57–67.

Allen, Douglas and Margaret Brinig (1998), 'Sex, property rights and divorce', *European Journal of Law and Economics*, **5**, 211–29.

Anderson, Gary and Robert Tollison (1991), 'A theory of rational childhood', *European Journal of Political Economy*, **2**, 199–213.

Aries, P. (1962), *Centuries of Childhood: A Social History of Family Life*, New York: Knopf.

Baker, Lynn and Robert Emery (1993), 'When every relationship is above average: perceptions and expectations of divorce at the time of marriage', *Journal of Law and Human Behavior*, **17**, 439–50.

Bartlett, Katherine (1984), 'Rethinking parenthood as an exclusive status: the need for legal alternatives when the premise of the nuclear family has failed', *Virginia Law Review*, **70**, 879–963.

Becker, G. (1975), *Human Capital: With Special Reference to Education*, 3rd edn, Chicago: University of Chicago Press.

Becker, G. (1991), *A Treatise on the Family*, 2nd edn, Cambridge, MA: Harvard University Press.

Becker, Gary (1993), 'Nobel lecture: the economic way of looking at behavior', *Journal of Political Economy*, **101**, 385–409.

Becker, Gary, Elizabeth Landes and Robert Michael (1978), 'An economic analysis of marital instability', *Journal of Political Economy*, **85**, 1141–87.

Becker, Gary and J. Gregg Lewis (1973), 'On the interaction between the quantity and quality of children', *Journal of Political Economy*, **2**, S279–88.
Becker, Gary and Kevin Murphy (1988), 'The family and the state', *Journal of Law and Economics*, **31**, 1–18.
Becker, Gary and Nigel Tomes (1976), 'Child endowments and the quantity and quality of children', *Journal of Political Economy*, **84**, S23–39.
Beecher, C. (1841), *A Treatise on Domestic Economy*; reprint, New York: Shocken Books, 1977.
Ben-Porath, Yoram (1980), 'The F-connection: families, friends, firms and the organization of exchange', *Population Development Review*, **3**, 18–30.
Bergstrom, Ted (1995), 'On the evolution of altruistic ethical rules of siblings', Working Paper, Department of Economics, University of Michigan.
Bernheim, B. Douglas, Andrei Shleifer and Lawrence Summers (1985), 'The strategic bequest motive', *Journal of Political Economy*, **93**, 1045–76.
Bishop, William (1984), '"Is he married?" marriage as information', *University of Toronto Law Review*, **34**, 245–62.
Blackstone, W. (1765), *Commentaries on the Law of England*, 15th edn, London: A. Strahan, 1809.
Brinig, Margaret (1990), 'Rings and promises', *Journal of Law, Economics and Organization*, **6**, 203–15.
Brinig, Margaret (1993), 'The law and economics of no-fault divorce', *Family Law Quarterly*, **26**, 453–70.
Brinig, Margaret (1994a), 'Status contract, and covenant', *Cornell Law Review*, **79**, 1573–602.
Brinig, Margaret (1994b), 'Finite horizons: the American family', *International Journal of Children's Rights*, **2**, 293–315.
Brinig, Margaret (1994c), 'A comment on Jana Singer's alimony and efficiency', *Georgetown Law Journal*, **84**, 2461–79.
Brinig, Margaret (1994d), 'The effect of transaction costs in the adoption market', *Seton Hall Journal on Legislation*, **18**, 553–79.
Brinig, Margaret (1995a), 'A maternalistic approach to surrogacy', *Virginia Law Review*, **81**, 2377–99.
Brinig, Margaret (1995b), 'Does mediation systematically disadvantage women?', *William and Mary Journal of Gender and the Law*, **2**, 1–34.
Brinig, Margaret (1996), 'The family franchise: elderly parents and adult siblings', *Utah Law Review*, 393–428.
Brinig, Margaret (1997), 'Property distribution physics, the talisman of time and middle class law', *Family Law Quarterly*, **32**, 93–118.
Brinig, Margaret (2000), *From Contract to Covenant: Beyond the Law and Economics of the Family*, Cambridge, MA: Harvard University Press.
Brinig, Margaret and Michael Alexeev (1993), 'Trading at divorce: preferences, legal rules and transaction costs', *Ohio State Journal on Dispute Resolution*, **8**, 279–97.
Brinig, Margaret and Michael Alexeev (1995), 'Fraud in courtship: annulment and divorce', *European Journal of Law and Economics*, **2**, 45–78.
Brinig, Margaret and F.H. Buckley (1996), 'The market for deadbeats', *Journal of Legal Studies*, **25**, 201–32.
Brinig, Margaret and F.H. Buckley (1998a), 'No-fault laws and at fault people', *International Review of Law and Economics*, **16**, 325–40.
Brinig, Margaret and F.H. Buckley (1998b), 'Joint custody, bonding and monitoring theories', *Indiana Law Journal*, **73**, 393–427.
Brinig, Margaret and F.H. Buckley (1999a), 'Child abuse and the ugly duckling', *Journal of Law and Family Relations*, **1**.
Brinig, Margaret and F.H. Buckley (1999b), 'The price of virtue', *Public Choice*, **98**, 111–29.
Brinig, Margaret and Steven Crafton (1994), 'Marriage and opportunism', *Journal of Legal Studies*, **23**, 869–94.
Brown, Jennifer G. (1994), 'Competitive federalism and the legislative incentives to recognize same sex marriages', *Southern California Law Review*, **68**, 745–839.

Buchanan, James M. (1983), 'Rent seeking, noncompensated transfers and the laws of succession', *Journal of Law and Economics*, **26**, 71–85.
Cass, Ronald (1987), 'Coping with life, law and markets: a comment on Posner and the law and economics debate', *Boston University Law Review*, **67**, 73–97.
Cohen, Jane (1987), 'Posner, pessimism, pluralism', *Boston University Law Review*, **67**, 105–75.
Cohen, Lloyd (1987), 'Marriage, divorce and quasi-rents, or, "I gave him the best years of my life"', *Journal of Legal Studies*, **16**, 269–303.
Cooter, Robert and Barry Freedman (1991), 'The fiduciary relationship: its economic character and legal consequences', *New York University Law Review*, **66**, 1045–75.
Cox, Donald and Oded Stark (1993), 'The demonstration effect', Working Paper, Boston College, Department of Economics.
Dowd, Nancy (1994), 'A feminist analysis of adoption', *Harvard Law Review*, **107**, 913–36.
Duncan, Greg and Saul Hoffman (1981), 'A reconsideration of the economic consequences of marital dissolution', *Demography*, **22**, 485–99.
Ellman, Ira (1990), 'The theory of alimony', *California Law Review*, **77**, 1–81.
England, P. and G. Farkas (1986), *Households, Employment and Gender*, Hawthorne, NY: Aldine Press.
Epstein, Richard (1989), 'The varieties of self-interest', *Philosophy and Policy*, **8**, 102–23.
Epstein, Richard (1995), 'Surrogacy: the case for full contractual enforcement', *Virginia Law Review*, **81**, 2305–42.
Filmer, R. (1653), *Patriarcha and Other Writings*; reprint, edited by Peter Laslett, Oxford: Basil Blackwell, 1949.
Fineman, M. (1995), *The Neutered Mother, the Sexual Family and Other Twentieth Century Tragedies*, New York: Routledge.
Fox, Greer and Robert Kelly (1993), 'Determinants of alimony awards: an empirical test of current theories and a reflection on public policy', *Syracuse Law Review*, **44**, 641–722.
Fuchs, V. (1992), *Women's Quest for Economic Equality*, Cambridge, MA: Harvard University Press.
Garrison, Marsha (1993), 'Good intentions gone awry: the impact of New York's equitable distribution law on divorce outcomes', *Brooklyn Law Review*, **57**, 679–754.
Goode, W. (1993), *World Changes in Divorce Patterns*, New Haven, CT: Yale University Press.
Grubb, Farley (1985), 'The market for indentured immigrants', *Journal of Economic History*, **45**, 855–68.
Hadfield, Gillian (1990), 'Problematic relations: franchising and the law of incomplete contracts', *Stanford Law Review*, **42**, 927–92.
Hale, Matthew (1678), *Pleas of the Crown*; reprint, London: Professional Books, Ltd, 1982.
Hetherington, Mavis, George Hetherington, Roger Cox and Martha Cox (1982), 'Effects of divorce on parents and children', in Michael Lamb (ed.), *Nontraditional Families: Parenting and Child Development*, Hillsdale, NJ: Erlbaum Associates.
Hume, David (1761), 'Of polygamy and divorces', in Eugene F. Miller (ed.), *Essays, Moral, Political and Literary*, Indianapolis, IN: Liberty Classics, 1987.
Jones, Owen (1997), 'Evolutionary analysis in law: an introduction and application to child abuse', *Carolina Law Review*, **75**, 1117–81.
Kronman, Anthony (1985), 'Contracts and the state of nature', *Journal of Law, Economics and Organization*, **1**, 1–22.
Kydland, F.W. and E.C. Prescott (1977), 'Rules rather than discretion: the inconsistency of optimal plans', *Journal of Political Economy*, **85**(3), 473–91.
Landes, Elisabeth (1978), 'The economics of alimony', *Journal of Legal Studies*, **7**, 35–78.
Landes, Elisabeth and Richard Posner (1975), 'The economics of the baby shortage', *Journal of Legal Studies*, **5**, 323–50.
Locke, J. (1689), *Second Treatise on Government*; reprint, Indianapolis, IN: New American Library, 1992.
Lundberg, Shelly and Robert Pollack (1993), 'Separate spheres bargaining and the marriage market', *Journal of Political Economy*, **101**, 933–1010.

Lupu, Ira (1994), 'Mediating institutions: beyond the public/private distinction: the separation of powers and the protection of children', *University of Chicago Law Review*, **61**, 1317–73.
Maccoby, E. and R. Mnookin (1992), *Dividing the Child*, Cambridge, MA: Harvard University Press.
Manser, Marilyn and Murray Brown (1980), 'Marriage and household decision-making: a bargaining analysis', *International Economic Review*, **21**, 31–44.
Martineau, Harriet (1889), *Household Education*, Boston, MA: Houghton-Mifflin.
Marvell, Thomas (1989), 'Divorce rates and the fault requirement', *Law and Society Review*, **23**, 543–60.
Mnookin, Robert and Lewis Kornhauser (1979), 'Bargaining in the shadow of the law: the case of divorce', *Yale Law Journal*, **79**, 950–98.
Nakoneszy, P.A., R.D. Snull and J.L. Rodgers (1995), 'The effect of no-fault divorce law on the divorce rate across the 50 states and its relation to income, employment and religiosity', *Journal of Marriage and the Family*, **57**, 477–88.
Palmer, John (1986), 'The social cost of adoption agencies', *International Review of Law and Economics*, **6**, 189–203.
Parkman, A. (1992), *No-Fault Divorce: What Went Wrong*, Boulder, CO: Westview Press.
Parkman, Allen (1995), 'Human capital as property in celebrity divorces', *Family Law Quarterly*, **29**, 141–69.
Parkman, Allen (1996), 'Why are married women working so hard?', Working Paper, Anderson School of Business, Albuquerque, University of New Mexico.
Peters, Elizabeth (1986), 'Marriage and divorce: informational constraints and private contracting', *American Economic Review*, **76**, 437–54.
Posner, Richard (1987), 'The regulation of the market in adoptions', *Boston University Law Review*, **67**, 59–72.
Posner, R. (1992), *Sex and Reason*, Cambridge, MA: Harvard University Press.
Posner, R. (1996), *Aging and Old Age*, Cambridge, MA: Harvard University Press.
Prichard, J. Robert S. (1984), 'A market for babies', *University of Toronto Law Journal*, **45**, 341–57.
Priest, George and Benjamin Klein (1984), 'The selection of disputes for litigation', *Journal of Legal Studies*, **13**, 1–55.
Radin, Margaret (1987), 'Market inalienability', *Harvard Law Review*, **100**, 1849–937.
Ramseyer, Mark (1995), 'The market for children: evidence from early modern Japan', *Journal of Law, Economics and Organization*, **11**(1), 127–49.
Ridley, Matt (1992), *The Red Queen: Sex and the Evolution of Human Nature*, New York: Macmillan.
Ross, Stephen (1973), 'The economic theory of agency: the principal's problem', *American Economic Review*, **63**, 134–49.
Rubin, Paul, James Kau and Edward Meeker (1979), 'Forms of wealth and parent-offspring conflict', *Journal of Social Biologic Structures*, **2**, 53–64.
Schneider, C. and M. Brinig (1996a), *An Invitation to Family Law*, St Paul, MI: West Publishing Company.
Schneider, C. and M. Brinig (1996b), *Teacher's Manual to Accompany An Invitation to Family Law*, St. Paul, MI: West Publishing Company.
Schultz, Marjorie (1982), 'Contractual ordering of marriage: new directions for state policy', *California Law Review*, **70**, 204–349.
Scott, Elizabeth (1990), 'Rational decisionmaking in marriage and divorce', *Virginia Law Review*, **76**, 9–94.
Scott, Elizabeth and Robert Scott (1995), 'Parents as fiduciaries', *Virginia Law Review*, **81**, 2401–76.
Scott, Elizabeth and Robert Scott (1998), 'Marriage as relational contract', *Virginia Law Review*, **84**, 1225–334.
Seltzer, Judith (1998), 'Legal custody arrangements and children's economic welfare', *American Journal of Sociology*, **96**, 895–929.
Singer, Jana (1992), 'The privatization of family law', *Wisconsin Law Review*, 1443–1567.

Singer, Jana (1995), 'Alimony and efficiency: The gendered costs and benefits of the economic justification for alimony', *Georgetown Law Journal*, **82**, 2923–60.
South, Scott (1985), 'Economic conditions and the divorce rate: a time-series analysis of the postwar United States', *Journal of Marriage and the Family*, **47**, 31–46.
Spiegel, H. (1983), *The Growth of Economic Thought*, rev. edn, Durham, NC: Duke University Press.
Stake, Jeffrey (1992), 'Mandatory planning for divorce and marriage', *Vanderbilt Law Review*, **45**, 397–456.
Stigler, George (1961), 'The economics of information', *Journal of Political Economy*, **69**, 213–25.
Sunstein, Cass (1994), 'Incommensurability and valuation in law', *Michigan Law Review*, **92**, 779–861.
Terman, L. (1930), *Mental and Physical Traits of a Thousand Gifted Children*, Vol. 3, Stanford, CA: Stanford University Press.
Trebilcock, Michael and Rosamin Keshvani (1991), 'The role of private ordering in family law: a law and economics perspective', *University of Toronto Law Journal*, **42**, 533–90.
Wallerstein, Judith (1980), *Surviving the Breakup: How Children and Parents Cope with Divorce*, New York: Basic Books.
Wallerstein, Judith (1991), 'The long-term effects of divorce on children: a review', *Journal of the American Academy of Child and Adolescent Psychiatry*, **30**, 349–60.
Weiss, Yoram and Robert Willis (1985), 'Children as collective goods in divorce settlements', *Journal of Labor Economics*, **3**, 268–92.
Weiss, Yoram and Robert Willis (1993), 'Transfers among divorced couples', *Journal of Labor Economics*, **11**, 629–79.
Weitzman, Lenore (1981), 'The social and economic consequences of divorce', *University of California–Los Angeles Law Review*, **28**, 1181–268.
Woodhouse, Barbara (1992), 'Hatching the egg: a child-centered perspective on parents' rights', *Cardozo Law Review*, **13**, 1747–865.
Woodhouse, Barbara (1996), 'It depends what you mean by home', *Utah Law Review*, 569–612.
Wright, Robert (1994), *The Moral Animal*, New York: Pantheon Books.
Zelder, Martin (1993), 'Inefficient divorce as a consequence of public goods', *Journal of Legal Studies*, **22**, 503–20.
Zick, Cathleen and Jennifer Geurer (1987), 'Family compensation and investment in household capital: contrasts in the behavior of husband–wife and female headed households', *Journal of Consumer Affairs*, **21**, 21–39.
Zick, Cathleen and Jennifer Geurer (1991), 'Trends in married couples' time use: evidence from 1977–78 and 1987–88', *Sex Roles*, **24**, 459–87.

8 Inheritance

Richard E. Wagner

Private property, freedom of contract and personal liability provide the central legal framework for a market economy, as Walter Eucken (1952) explains in his well-known statement of *Ordnungstheorie*. While there is little dispute about the principal features of this framework, there is much dispute about the status of property when the owner dies. There are two polar regimes in this regard, free inheritance and collective inheritance. Under free inheritance, an owner of assets would have the same right to dispose of his assets upon death as he had during life. Any state involvement would be minimal, as illustrated by such things as the usual recordation fees charged when certain asset titles are transferred. Under collective inheritance, an owner of assets would have the full use of his assets only during his lifetime, and those assets would become state property upon his death. Collective inheritance would entail the imposition of a 100 per cent tax on all assets that were held by a decedent at the time of death.

A pure regime of collective inheritance is almost certainly impossible in modern societies. There are several related reasons why any effort to tax estates at 100 per cent would collect little, if any, revenue. The effort to impose such a tax would induce people who had accumulated wealth during their lifetime to consume it before their death, by doing such things as converting that wealth into annuities. It would also induce such people to transfer more of their wealth while they were alive, though this possibility would also lead the state to attempt to tax gifts, which in turn would further induce prospective donors to seek methods of doing this none the less. Furthermore, deaths typically are known to heirs before they are known to tax officials, and collective inheritance would surely induce heirs to scavenge among the decedent's assets, particularly among those assets for which titles are not recorded publicly, before officers of the state even arrive on the scene to press their claims. For these reasons, among others, a regime of collective inheritance would bring into the state's possession, at most, only a small portion of the assets that would have been transferred from decedents to heirs under a regime of free inheritance.

To avoid such an impossible situation, states stop well short of 100 per cent rates of taxation and generally confine tax liability to what are considered to be comparatively large accumulations of wealth. A state that sought to maximize the revenues it could collect from a tax that it imposed on the

wealth of decedents would be well advised to keep its rate of tax well below 100 per cent and to limit the base to relatively large estates. This is not to claim that such a revenue-maximizing approach would be desirable, but is only to say that there is a pragmatic limit to the extent to which inheritance can be collectivized. Total collectivization is simply impossible and the maximally attainable extent of collectivization will still involve some semblance of free inheritance.

Despite this pragmatic recognition of the limits to the collectivization of inheritance, conceptual clarity concerning alternative approaches to inheritance is promoted by considering the polar alternatives of free and collective inheritance. What does it matter whether inheritance is free or collectivized, since in any event the extent of collectivization will necessarily be incomplete? Is free inheritance socially beneficial, in which case inheritance taxation could be socially destructive, or might free inheritance be socially destructive, in which case inheritance taxation might be socially beneficial? For the most part, the arguments for collective and free inheritance mirror one another, with each claiming to be part of a programme for human flourishing.

Collective inheritance and equal opportunity

There seem to be two sources of support for some collectivization of inheritance through taxation, one grounded in claims of self-interest and the other in claims of social benefit. With respect to self-interest, the best tax is surely always one that someone else pays. Whether the count is made in terms of decedents or in terms of heirs, liability for inheritance taxes is confined to a small minority of the population. If the alternative to the taxation of inheritance is some broad-based tax on income or consumption, the taxation of inheritance would seem to offer tax reductions for a majority of people. To be sure, there is good reason and strong evidence in support of the proposition that the burdens of inheritance taxes are distributed more widely throughout the population, owing to such things as negative effects on capital accumulation and wages (see Wagner, 1993). None the less, some measure of support for inheritance taxation may derive from a belief held by most people that it is a tax that someone else pays.

Most of the support for inheritance taxation, however, has been based on some claim of fairness joined with claims about the characteristics of a good society. This support for some collectivization of inheritance is often stated with reference to the deformities that are alleged to be generated by free inheritance, and which are claimed to be amenable to mitigation through the collectivization of inheritance. Through its ability to magnify and transmit material inequality across generations, a regime of free inheritance is alleged to inject elements of a caste system into society. People become wealthy, not because of what they have accomplished, but because their parents were

wealthy. Others become poor, not so much because of failings on their part, but because the posts of accomplishment in society will have been foreclosed to them by the transmission of material position through inheritance. The claim in this respect is that free inheritance impedes the failure of children from rich families while concomitantly impeding the success of children from poor families. It is as if there are only so many corporate chief executive positions in a society, and for every such position that is filled through inheritance an opportunity to attain such a position is closed to those without such inheritances.

By reducing the advantages that parents can transmit to their children, some collectivization through taxation is argued to be a means of promoting some measure of equality of opportunity within a society. To be sure, more than material wealth is transmitted from parents to children, so the ability of collective inheritance to promote equal opportunity will be similarly limited (Meade, 1973). None the less, it would be easy for proponents of collective inheritance to argue that some effort to promote equal opportunity along those dimensions that are susceptible to such promotion is surely better than failing to do so, simply because there happen to be other dimensions that are not so susceptible to such promotion.

In a commonly used analogy, the receipt of an inheritance is treated as being similar to the receipt of a head start in a footrace. The collectivization of inheritance is construed as a means of helping to promote equality of opportunity, which in turn is construed as a situation where everyone starts the race from the same position. The popularity of this footrace analogy surely lies in its transparent simplicity. If economic life is analogized to a relay race, free inheritance gives head starts to those racers in any cohort who receive the baton ahead of the pack. Free inheritance creates a relay race where the time of departure from the starting line for any particular runner is governed by the speed at which his forebears had run their legs. By contrast, collective inheritance would convert this relay race into a series of independent stages, one for each generation. Or at least this would be the idealized accomplishment of collective inheritance, keeping in mind the caveat that there are many forms of inheritance besides material wealth transmitted at death.

The very simplicity of the footrace analogy seems often to overshadow its dubious relevance. A footrace is a zero-sum game. Whatever increases the odds that one particular racer will win must necessarily decrease the odds that other racers will win. Economic life, however, is positive sum and not zero sum. The increased wealth that accrues to the inventor of a new industrial process does not come at the expense of everyone else, but rather is a genuinely new creation of something that did not previously exist and which, moreover, generates increased wealth elsewhere in society as well. There are

not a fixed number of chief executive officer (CEO) positions available in a society, because the number of such positions will depend on a variety of considerations that govern the creation and success of enterprises. If taxes that impinge heavily upon the successful creation of enterprises diminish such efforts within a society, there will be a shrinkage in the observed number of CEO positions.

Claims about equal opportunity and collective inheritance are often reinforced by claims that earned wealth is morally superior to unearned or inherited wealth. The widely-cited Rignano approach to inheritance, for instance, would apply only a 50 per cent tax to the first generation of inheritance but would confiscate any transfer to a second generation. This represents one effort to institutionalize a belief that inherited wealth is normatively inferior to earned wealth, and with the degree of inferiority rising as time passes. To a large extent, this claim of inferiority involves a presumption that wealth can perpetuate itself without effort, as illustrated by simple analogies based on annuities and compound interest. This formulation makes it seem as though people can live on the unearned incomes that they have received through bequest, instead of earning their own way in society. It is undeniable that a large fortune can support a lot of slothfulness and indulgence for its heirs. It is also undeniable that such a fortune will be a fortune that is on its way to dissipation. While assets can be sold and annuities purchased, wealth will not perpetuate itself without effort, regardless of how that wealth has been attained. Someone might inherit a highly valuable software company. Regardless of the company's value at the time of inheritance, however, that value will plummet should the company choose simply to rest on its past accomplishments. In a competitive market economy, all asset positions are open to continual challenge. The value of the software company at the time it was inherited will have to be earned over and over again, or else it will be lost as customers shift their patronage to the superior offerings of competitors. It will take the same application of energy and creativity to maintain the value of the software company as it took to establish that company.

Moreover, there can be no doubt that the inheritance of material wealth is socially beneficial. We are all vastly better off today than our forebears of yesteryear because of the legacies that they have left us. To be sure, proponents of collective inheritance do not seek to abolish inheritance, but only to subject it to some partial collectivization. In so doing, some tradeoffs would have to be faced. One element of that tradeoff concerns the value placed on equal opportunity relative to that placed on personal liberty. Another element concerns the degree to which the collectivization of inheritance will reduce the value of the legacies that decedents bequeath. Both elements would be part of anyone's evaluation of alternative inheritance regimes. It is possible that someone could think that the aggregate volume of legacies would decline

sharply as tax rates increased and yet support inheritance taxation, because equal opportunity is valued highly relative to liberty. It is also possible that someone could think that inheritance taxation would exert a relatively modest depressant effect on the volume of legacies and yet oppose such taxation, because of the relatively high weight placed on liberty. In any case, a good deal of the controversy over the taxation of inheritance has concerned such things as the rate at which increased taxation reduces capital accumulation and the relative weights to be placed upon equal opportunity and personal liberty. At base, the central concern of this controversy is whether free inheritance or collective inheritance is more consistent with human flourishing.

Free inheritance and human flourishing

Two of the primary institutions of civil society are private property and the family. Free inheritance supports both of these institutions, while collective inheritance subverts both and with the degree of subversion varying directly with the rate of tax. The attack on private property and the family that collective inheritance represents, no matter how incompletely collective it might be, resonates well with the famous controversy between Plato and Aristotle regarding the rearing of children. In *The Republic*, Plato advocated that children be taken away from their parents and raised in common. The argument for doing this was grounded in equal opportunity. If children were raised by their own parents, some children would be advantaged relative to others because of differences in family settings. Plato presumed that, if all children were raised in common, the advantages that particular children derived from being raised in particular families would be abolished, because under the Platonic system all parents would treat all children equally. The problem with this alternative, Aristotle noted in his *Politics*, was that all parents would treat all children with equal indifference. As Aristotle summarized, 'it is better to be own cousin to a man than to be his son after the Platonic fashion'. For children to be raised with parental interest and not indifference, it is necessary to call upon the natural partiality of parents for their own children.

The Platonic scheme may well reduce the variability among the members of a particular generation that arises from differences in family settings, but only at the expense of weakening private property and the family and embracing the consequences that would stem from that weakening. Families differ widely in the legacies they transmit to their children. The taxation of inheritance seeks to pare down the legacies that particularly successful families are able to leave. If flourishing were a zero-sum condition, where more flourishing for some meant less flourishing for others, this might create some greater spread of flourishing throughout a society. But the extent of flourishing within a society is a variable and not a fixed condition, and societies with free

institutions flourish more fully than societies with highly collectivized institutions, as Aristotle recognized relative to Plato.

This raises a fundamental disjunction concerning two ways of approaching equal opportunity and inheritance. The common approach seeks to promote equal opportunity by restricting the ability of the relatively successful to leave bequests. An alternative approach to human flourishing would look to the elimination of negative legacies as an important element in a positive programme for a flourishing society. Rather than seeking to penalize those who were successful in creating positive legacies, it would seek to cultivate conditions that were less conducive to the persistence of negative legacies. Such a programme for a flourishing society would seek to reform those institutions that restrict opportunity, rather than to curtail those institutions, such as private property and freedom of inheritance, that foster it.

What such an approach would look like would go well beyond the bounds of this entry, though a few points can be made briefly in passing. For one thing, such an approach would point in part to territory that is now under examination in the widespread rethinking of the welfare state that is well under way (for an interesting collection of essays in this respect, see Ebeling, 1995). In this rethinking, it is coming to be recognized increasingly strongly that it is impossible to offer guarantees of income in one form or another without at the same time undermining the exercise of those human faculties that are essential to human flourishing. Programmes that encourage what are little more than children to have children do not cultivate conditions conducive to flourishing. Nor do programmes that substitute welfare cheques for the application of effort and the exercise of providence.

A related issue concerning the requisites for human flourishing concerns the place of wealth and inheritance in the creation and support of various publicly beneficial institutions and organizations that typically cannot be supported through normal commercial channels and yet undertake activities that are also undertaken by states. These include a variety of charitable organizations, museums, hospitals, educational establishments, foundations and the like. There is both good reason and credible evidence in support of two related propositions: first, state provision crowds out private provision; second, private provision is often more effective than state provision in promoting the requisites for flourishing. By promoting the substitution of state provision for private provision, the collectivization of inheritance would thus seem to retard human flourishing.

Inheritance, democracy and institutional complementarity

The performance properties of any particular institutional arrangement will depend on the larger institutional framework within which that arrangement is embedded. Within a regime of liberal democracy, free inheritance pro-

motes human flourishing while collective inheritance restrains it. Within the framework of liberal democracy, the state is subject to the same principles of property and contract as are other participants in society. Liberal democracy generates a system of open competition, governed only by the general principles of property and contract. All wealth positions are open continuously to challenge, and a regression towards the mean characterizes relative wealth positions through time, both within the members of any particular generation and across generations.

The institutions of liberal democracy have been under strenuous challenge from those of social democracy throughout the twentieth century. Under a regime of social democracy, the state is not limited by the same principles of property and contract that limit other participants in society. Rather, the state becomes the arena where property rights are determined, revised and extinguished. It is also an arena where the substance of contract can be amended through legislation and regulation. Competition is no longer open, as governed only by the principles of private property and freedom of contract. To a large extent, competition becomes closed and moderated through the state. Established wealth positions attain some shelter from open competition through the political purchase of favourable legislation and regulation that act as a form of insurance against the erosion of those wealth positions through competition. With the state becoming a partisan of those who are established and with the complementarity that results between wealth and power under social democracy, the natural process of regression towards the mean may be slowed. This illustrates how questions concerning the institutional arrangements governing inheritance are ultimately connected with questions concerning the entire system of institutional arrangements for governing our relationships with one another.

Bibliography

Aristotle (1946), *The Politics*, ed. and trans. by E. Barker, Oxford: Oxford University Press, p. 45.

Becker, Gary S. (1981), *A Treatise on the Family*, Cambridge, MA: Harvard University Press.

Becker, Gary S. and Nigel Tomes (1986), 'Human capital and the rise and fall of families', *Journal of Labor Economics*, **4** (3, pt. 2), 1–39.

Brittain, John A. (1978), *Inheritance and the Inequality of Material Wealth*, Washington, DC: Brookings Institution.

Ebeling, Richard M. (ed.) (1995), 'American perestroika: the demise of the welfare state', Ludwig von Mises Lecture Series, no. 23, Hillsdale, MI: Hillsdale College Press.

Eucken, Walter (1952), *Grundsätze der Wirtschaftpolitik* [Principles of Economic Policy], Tübingen: J.C.B. Mohr.

Meade, James E. (1973), *The Inheritance of Inequalities: Some Biological, Demographic, Social and Economic Factors*, London: Oxford University Press.

Plato (1892), *The Republic*, ed. and trans. by B. Jowett, London: Macmillan.

Rignano, Eugenio (1924), *The Social Significance of the Inheritance Tax*, trans. by William Shultz, New York: Knopf.

Tullock, Gordon (1971), 'Inheritance justified', *Journal of Law and Economics*, **14**, 465–74.

Wagner, Richard E. (1977), *Inheritance and the State*, Washington, DC: American Enterprise Institute.
Wagner, Richard E. (1993), *Federal Transfer Taxation: A Study in Social Cost*, Washington, DC: Institute for Research in the Economics of Taxation.
Wedgewood, Josiah (1929), *The Economics of Inheritance*, London: George Routledge & Sons.

9 Intellectual property and the markets of ideas

Giovanni B. Ramello[1]

Introduction

The term 'intellectual property rights' denotes a set of legal doctrines – namely patent, copyright, trademark and trade secret – that differ in their structure, scope and spheres of application, but nevertheless have in common the feature of granting the owner rights over the economic exploitation of an idea or its 'reification' (that is, its expression in any tangible medium, as in the case of copyright–authors' rights). Such rights are generally exclusive, meaning that the owner is given a legal monopoly over the protected idea.[2] From this perspective, therefore, and despite the distinct peculiarities of each, the roots of all intellectual property rights can be traced to the advent of a knowledge economy and the private appropriation of certain types of information. In fact, as the production of knowledge began generating increasing value, in tangible economic terms as well as socially, the question arose as to the appropriability of this new knowledge – or more to the point of the benefits derived therefrom – which was resolved in Western commercial and industrial societies through the attribution of specific property rights. In this respect, intellectual property rights are only one possible solution to the dilemma of the ownership of knowledge.[3]

Different societies have resolved this dilemma in different ways.[4] Of the various possible solutions, those adopted by Western society are especially propitious to the circulation of knowledge within economic contexts and its exploitation in the marketplace.[5] Moreover, it has the effect of specifically stimulating the production of those ideas that the market most readily rewards. Hence, the traditional benefit associated with intellectual property rights, at least from the law and economics perspective, is that it provides an incentive for the creation and/or dissemination of new ideas. But intellectual property rights have yet another feature in common, one that is often neglected in the literature, but central to the economic analysis: namely that, by the very fact of being property rights, they contribute by definition to shaping the market structure, regulating the competitive scenario and determining the rational behaviours of economic agents.

In other words, property rights are not merely static instruments for fine-tuning the market of ideas; they are much more pervasive in their effects, able to shape the features of markets and bring about the emergence of particular actors, with a sometimes significant impact on creative and inventive processes,

as well as on innovation. In the present discussion we shall give equal weight to both perspectives, starting from the traditional analysis and its sources, and then addressing the indirect economic effects on the configuration of the markets. More specifically, the next section will describe the origins of the economic interpretation of intellectual property, presenting the traditional 'static' paradigm. The following sections will discuss, in turn, the application of this paradigm to the various specific rights, illustrating their peculiar features, and the 'dynamic' role of intellectual property rights, showing how these institutions play a decisive role in shaping the markets of ideas. The next section proposes some directions for future development that are currently being explored, albeit fragmentarily, by the contemporary scientific debate, and the final section concludes.

Origins of intellectual property

The economic interpretation of intellectual property rights is a direct descendant of the theory of *labour mixing* and property formulated by John Locke (1632–1704). This theory is in fact the cornerstone of the reasoning that establishes a causal link between creators and ideas, thereby legitimizing the individual appropriation of the latter through the ad hoc institution of property rights.

In his second book of the *Two Treatises on Government* (1690), Locke maintains that every man has a 'natural right' to appropriate the fruits of the labour of his own body. In other words, the English philosopher believed that individuals acquire property rights over assets originally contained in the state of nature, by virtue of the fact that in order to extract them they contribute their own labour, and hence a part of themselves (Drahos, 1996).

Locke did not espouse this position absolutely, however, but considered it subject to some clear-cut and essential derogation criteria. He defined specific limits for appropriation, in the form of two provisos that appear to foreshadow the Pareto optimality criterion: that is, that an individual can extract from the state of nature only that which leaves 'enough and as good left for others', and in any case that 'the same law of nature, that does by this means give us property, does also bound that property too … Nothing was made by God for man to spoil or destroy' (Locke, 1690, sect. 27).

Locke thus implicitly asserted for the first time that there exists a tradeoff between private appropriation and the public sphere, an idea taken up by all the subsequent literature on intellectual property, which has sought in various ways to preserve and enrich the shared resource in question – the 'common' and later the 'creative common' as defined in some of the literature (Hardin, 1968; Lessig, 2001) – while at the same time guaranteeing a sufficient private incentive through individual appropriation. In any case, the two Lockean provisos provide a simple but effective rule of reason for regulating the

extension of such rights, and for resolving any conflicts between private and public interest that might arise. The various criteria for derogation from intellectual property rights – temporal limit, right exhaustion in European Union (EU) jurisprudence or first sale doctrine in US law, fair use doctrine in the case of copyright, minimum threshold for attribution and so on – can thus be viewed as descendants of the Lockean prescriptions, whose ultimate aim is to minimize the negative effects of private appropriation and favour the return of new knowledge to the public sphere.

Nevertheless, economic analysis proper, as an independent discipline, only entered into the intellectual property rights debate later, with Adam Smith (1723–90),[6] who questioned the natural right of individuals over created ideas, while upholding the importance of (limited) *ex lege* protection 'as an encouragement to the labours of learned men'.

Smith's position is only mildly in favour of intellectual property, probably reflecting a disinclination to uphold any type of monopoly – even a legal one – that interferes with competitive process. He thus admits lukewarm support for intellectual property rights 'as they can do no harm and may do some good, are not to be altogether condemned' (Smith 1762, *Lectures on Jurisprudence*, in Goldstein, 1994: 173), but ultimately, no clear theoretical or policy indications emerge.

The challenge of developing a more robust paradigm was picked up by Jeremy Bentham (1748–1832), who provided the theoretical groundwork for centuries to come by first formulating the utilitarian theory of an 'incentive to create', widely adopted by the ensuing literature up until the present day. In particular, the English economist notes that 'he who has no hope that he shall reap, will not take trouble to sow' (Bentham, 1839: 31). In fact Bentham (p. 71) observes:

> [T]hat which one man has invented, all the world can imitate. Without the assistance of the laws, the inventor would almost always be driven out of the market by his rival, who finding himself, without any expense, in possession of a discovery which has cost the inventor much time and expense, would be able to deprive him of all his deserved advantages, by selling at a lower price

Intellectual property can therefore offer a practicable solution to this problem.

Note that Bentham's assertion in a sense prefigures the dilemma of the appropriability of public goods, and poses an implicit challenge to the Smith model of perfect competition: the invisible hand alone is not able to govern markets of ideas which, without specific intervention by the legislator, will therefore be doomed to failure.[7] A strong awareness of the problem of *free-riding* and its negative effects on the remuneration of creators prompted Bentham to openly side with an institution that provides an incentive for

creative activities. The creator, he asserts, must be protected against opportunistic individuals who would otherwise 'without any expense, in possession of a discovery which has cost the inventor much time and expense, be able to deprive him of all his deserved advantages, by selling at a lower price' (ibid.).

Now, the above-cited theory has played a 'totemic' role in the law and economics literature on intellectual property, providing a universal reference framework for nearly all subsequent contributions, which have incrementally fine-tuned the analysis, systematized and updated it, occasionally levelled some criticisms, but always held the central paradigm intact.[8] In other words, we can draw a direct line of continuity between the model proposed by Bentham and the literature of law and economics. In like manner we can argue that the contribution of twentieth-century economic theory has essentially been to extend and reinforce the Benthamian paradigm, by providing further arguments in support of the general incentive-creating role of intellectual property rights.

In effect, we have on one side an entire branch of literature grown out of Schumpeter's (1943) contribution, which posits the centrality of innovative activities to economic systems and their growth, and implicitly attributes to intellectual property rights a determining role in this dynamic.[9] And on the opposing side we have the theory of intellectual property rights as an efficient solution to the problem of public goods and externalities, which has provided the standard argument for the specific case of intellectual property, upholding the pro-efficiency role of intellectual property rights (Coase, 1960).[10]

Intellectual property rights and incentives

Although there are only four types of intellectual property right, in practice there are as many variants as there are nations that recognize them.[11] Some of the distinctions are minor, while others reflect differences in the underlying regulatory frameworks. This is true, for example, in the case of copyright in common law systems and authors' rights in civil law systems, where the different designations indicate, at least in theory, a different conception of the role of the author. In practice, however, recent convergence of international regulations has eroded these distinctions, to the point that some observers claim that, notwithstanding the differences, the two institutions have evolved through a dialectical process and thus exhibit common traits (Strowel, 1993).

Further diminishing the diversity are the effects of regulatory dynamics impressed at an international level by the TRIPS agreement (the Agreement on Trade-Related Aspects of Intellectual Property Rights), defined and ratified by the World Trade Organization (1994), and which essentially seeks to bring about a gradual international unification of the doctrine for enforcement of intellectual property rights.[12] The various types of intellectual property rights are instead differentiated according to the type of information that is

protected (the so-called 'subject matter'), the attribution criteria, the type of exclusive right granted to the owner (in terms of its structure and duration) and the incentive conferred.[13]

Patents generally protect ideas relating to the technological and scientific spheres (that is, new products or novel production processes) and grant the inventor/discoverer – or his/her employer – an exclusive right of limited duration (which varies from country to country) as a reward for shouldering the risk and investments connected with the research and development of the new idea.[14]

Transferring original ownership of the patent from the inventor to the employer – who becomes for all intents and purposes the inventor – should not, at least in theory, undermine the incentive mechanism, because it fulfils the criterion of risk transfer: the company agrees to take on the risk, and is thus entitled to benefit from the success of the venture, while the inventor receives a guaranteed remuneration in return. Now, in a market characterized by perfect information, in which there are no financial or other types of imperfections, such a mechanism will effectively be able to efficiently allocate the risk and stimulate inventive activities at the same time. Otherwise, if the conditions listed above are not met, the outcome will not be as expected and the goal of efficiency will not be achieved (see Scotchmer, 1991, 1998).

Because a patent application procedure requires revealing the invention, the institution of patents also creates an incentive to disclose new information. This twofold character of the resultant incentive – to create and to disclose – does not, however, have strong implications for the economic analysis, because in this case undisclosed information will not permit exploitation of a legal monopoly, and hence provides no incentive to create. In consequence, any profits deriving from exploitation of the exclusive right necessarily entail disclosure of the patented idea, and hence an equivalence between the two incentives.

The granting of a patent is not automatic, but subject to fulfilment of three criteria: the criterion of *novelty*, which calls for a substantial advance with respect to preceding inventions, the criterion of *non-obviousness*, which requires that the advance should not be trivial and, finally, the criterion of *utility*, which requires the invention to have some application and therefore not to be an end in itself. Complying with these criteria should ensure that strong exclusive rights are granted only to those ideas that effectively constitute a real technical or scientific advance.

Copyright–authors' rights, on the other hand, protect expressions of ideas, that is, information fixed in any tangible medium such as books, CDs, films, software and the like.[15] They are granted based on a criterion of minimal originality, meaning that the new expressions of ideas need only be marginally different from pre-existing ones.[16] In reality, there is no verification

procedure involved in the granting of copyright, and it is plagiarized authors who must contest the damage caused by some other, insufficiently original, expression of idea.

Moreover, because fixing of a copyright-protected idea in a tangible medium is in a sense instrumental to the consumption of individual units, while the information contained therein remains fluid and easily duplicated, copyright includes provision for a bundle of rights that also regulate its reproduction and dissemination.[17] We can therefore say that if technology somehow makes it possible to package information (Thomas Edison eloquently dubbed the phonograph, his invention for recording and playing back sound, 'canned sound'[18]), converting it into private unit goods that can be exchanged on the market like traditional goods, copyright provides the legal glue for 'sealing the can' created by technology, and preventing undue appropriation.

Copyright is thus an exclusive right, granted to the author as an incentive to create, in agreement with the premises of the Benthamian paradigm. It should in theory have limited duration, to permit the subsequent enrichment of public knowledge. In practice, though, following various amendments considerably strengthening the right (with the US Copyright Term Extension Act 1998 it can last up to 70 years *post mortem autoris*, or 95 years in the case of a company owning the right, while the European Duration Directive 93/98/EEC has extended protection to 70 years *post mortem autoris*) its duration has become virtually infinite, at least for the purposes of economic analysis.[19] It is this last-mentioned aspect that raises some serious questions concerning the global efficiency of the right (Antill and Coles, 1996; Lessig, 2001).

The incentive of copyright stimulates the creation of new expressions of ideas, by guaranteeing to the owner, who is generally the author (in the US case of work made for hire, as in the case of patents, the company that takes it on can be considered to be the author), exclusive rights over the economic exploitation of the intellectual property in all its forms, both direct and indirect.

Trade secrets similarly seek to stimulate the creativity of individuals by guaranteeing rights of secrecy to those who produce certain forms of information, generally pertaining to production processes (for example, the formula for a particular soft drink, a customer list and so on).[20] In fact such information, once disclosed, can easily be appropriated and imitated by competitors, thereby reducing the incentive for the creator to invest resources in its development, if he/she will be unable to profit from it later.

Some authors do not concur in assigning to trade secrets the attributes of property, for the reason that a trade secret is not necessarily exclusive, with nothing preventing several individuals from unwittingly owning the same secret.[21] For the purposes of this discussion, though, this observation can be

set aside because a trade secret is still based on the same rationale as all other intellectual property rights: granting a private benefit to the owner of a given information good, in order to create an incentive for its production. The fact that such a right may (unwittingly) be shared by two or more individuals does not compromise the logic of the system, and constitutes at worst an imperfection.[22] In any case, the economic behaviours of the owners are decided independently, irrespective of the possible existence of co-owners, and hence as if they effectively had exclusive rights over that particular information.

A trade secret has the function of conferring a competitive advantage to the owner in exchange for the creation of new information. Hence, the prospective existence of two or more co-owners will in the worst case cancel out the advantage, with some interesting pro-competitive side-effects, for example by fostering price competition. The only downside to this eventuality is that two separate individuals are given an incentive to invent the same information, with the attendant duplication of expense. However, this actually happens much more often in the realm of patents, where the problem is solved by a questionable *ex post* attribution that does not avoid the *ex ante* allocation inefficiency.[23] It should nevertheless be emphasized that such situations are the exception, and not the rule.

The overall balance of welfare resulting from trade secrets is somewhat more uncertain than for other intellectual property rights, because the right can endure up to the time when, for whatever reason, the information is disclosed and enters the public domain. This means that, at least in principle, it can last for an unlimited period. If the protection is effectively prolonged indefinitely, the social benefits will be very slight, because the protected information remains forever private. To a degree, a trade secret does provide a certain incentive to develop new and improved products; however, the fact of non-disclosure prevents appropriability of the knowledge for incremental creation – a fundamental point because the owners of the right may not necessarily be the most efficient creators for subsequent developments. On the opposing side, it does give a competitive advantage to the owner of the right, who could exploit the position thus obtained to pursue rent-seeking behaviours detrimental to general efficiency and welfare.

Finally, on top of the social costs arising from exercise of the right, we must also factor into the welfare balance the costs incurred in keeping the information secret. But these costs are also the key to solving the dilemma of duration: precisely because secrets are difficult and expensive to keep, it can generally be expected that the information will sooner or later be publicly disclosed (Friedman et al., 1991).

Finally, *trademark* is an exclusive right that likewise seeks to foster the creation of new information. However the information in question is, at least in the first instance, only an accessory to goods of other markets. Trademark

generally serves to convey information about the quality of products and the reputation of manufacturers, for the benefit of consumers (Landes and Posner, 1987).

However, some authors have recently put forward the thesis that trademark creation might be an economic end in its own right, serving to confer on the owner reputation inertia – and therefore market power – that is extensible and transferable to other markets, as in the case of 'brand extension and brand stretching' (Choi, 1998; Pepall and Richards, 2002), and that a trademark might even constitute an independent asset (Tadelis, 1999) or a well-defined symbolic good for which consumers exhibit a specific willingness to pay.

By carrying a certain amount of information about a product, a trademark effectively enables a product to be differentiated from its competitors, and ultimately gives the owner some degree of market power, which increases in direct proportion with the impact that the brand itself has on consumers.[24] So trademark has the additional, secondary effect of altering the quality perceived by consumers, who may consider equivalent products to be different solely because they are marketed under different brand names. This naturally has the effect of altering their willingness to pay and hence the balance of welfare.

From the attribution standpoint, because the public goal of trademarks is to facilitate exchanges by giving consumers more information about the goods that they are purchasing, trademark law does not generally admit appropriation of words or symbols denoting specific categories of objects (for example, a car maker will not be allowed to use the term 'car' as a trademark, whereas a manufacturer of garments and fashion accessories can use the word 'diesel'). Accordingly, continued protection is denied to trademarks which, for a variety of reasons, have entered into the common language and come to denote generic product categories (as in the celebrated cases of frigidaire, typewriter, elevator, aspirin, nylon, kerosene, yo-yo and so on, all originally brand names).

Intellectual property rights and markets

Generally speaking, law and economics theory treats all the above-mentioned rights as being neutral in their effects on the markets of ideas: given the economic context, the rights have the sole function of favouring an efficient equilibrium, without significantly altering the nature of the markets. However, this assumption is by definition incorrect: given that intellectual property rights embody specific economic policy measures – namely, an effort to encourage an efficient level of ideas – they are instruments of market regulation and therefore influence the competitive configuration, altering it not just marginally, but drastically (Ramello, 2003). This is a dynamic that has been by and large neglected by the scientific literature. In fact the model implicitly

adopted by much of the economic theory on intellectual property rights is more of a Schumpeterian innovation race, adapted to fit the various creative contexts, which takes the form of a type of perfect intertemporal competition (Evans and Schmalensee, 2001). In other words, according to a large part of the literature, the prospect of enjoying temporary supraprofits – brought by the exclusive rights – gives creators an incentive to put new ideas on the market, which does, however, remain competitive over the long term (see, for example, Besen and Raskind, 1991).

This representation has long been associated with a related view, namely that market success coincides with optimality of ideas, that is, that the market always causes the best ideas to emerge, so that the mechanism described above is efficient from every perspective. But this idea, too, has been discredited by the literature of the past decade, which points out how the short-run profit-maximization goals of firms and the uncertainty of creative activities are fundamentally at odds with the long-run welfare-maximization goal, thereby penalizing and crowding out certain types of creative activities such as, for example, basic research (Dasgupta and David, 1994; Arora and Gambardella, 1997; Scotchmer, 1998).

None the less, the mainstream approach continues to rely on the assumption that the legal framework does not stifle – but on the contrary enhances – the competitiveness of the market,[25] neglecting to note that the specific incentive mechanisms have the effect of significantly altering the payoffs and the maximizing behaviours of economic agents. Such effect emerges clearly from surveys on specific industries, which indicate that the market structure of the investigated sectors is significantly altered with respect to both the competitive and the Schumpeterian structures.[26]

Having said that, the primary aim of intellectual property rights is, by definition, to give the owner a certain amount of market power, and this has clear-cut effects on the competitive structure. This is a key point for the economic analysis, and must be properly understood. In effect, there is a widely held belief that the legal monopoly conferred by exclusive rights does not necessarily confer market power on the owner, or translate into an economic monopoly (see Anderson, 1998). Now, this is an acceptable assertion in general terms, but subject to misinterpretation.[27] It is effectively true that intellectual property right-protected information, if not successful, will not allow the legal monopoly to translate into an economic monopoly: a trademark that means nothing to consumers will not differentiate a product from its competitors. A drug that is ineffective, even though it is patented, will not confer any kind of market power. A CD that nobody is interested in buying can by no means be considered an economic monopoly, even though its copyright protection *does* make it a legal monopoly. Ultimately, the legal monopoly will translate into an economic monopoly only when consumers

perceive the information good protected by intellectual property right as being poorly substitutable for other information goods.[28] Only then does exclusive control over the non-substitutable resource confer market power to the owner, in inverse proportion to the substitutability of the good.

In some cases, when this market power is at a maximum, the legal monopoly becomes an economic monopoly; in other cases this is only partially achieved, but still allows the owner of the right to enjoy a certain profit margin. In any case, it is the prospect of securing supraprofits (and therefore market power) that constitutes the incentive to create, since a perfectly competitive market would deliver no extra profits and therefore zero incentive. The logic behind intellectual property rights is thus to reward successful ideas with market power: providing a monopoly, to a greater or lesser extent, as a private benefit in exchange for the creative effort/investment.

We have discussed how trademark increases the amount of information associated with a given product, thereby benefiting consumers, but that in so doing it also serves to differentiate the product from competing ones, that is, attributes market power, thereby benefiting the company that owns the right. A similar dynamic can be observed in other markets where the reward mechanism likewise hinges on the erection of *de jure* entry barriers that effectively confer market power.

Summarizing these arguments, therefore: a systematically competitive scenario implies a high level of substitutability between products; in this scenario, though, a complex apparatus such as intellectual property rights is not economically efficient, because it would be enough to directly finance the production of just a few ideas (as they are perfectly interchangeable) thereby sidestepping the social costs of intellectual property and the duplication of expense incurred in the production of equivalent ideas.[29] Now, according to the economic rationality hypothesis it can reasonably be expected that those in possession of market power will not stop at its temporary enjoyment but, as rational actors, will strive to maintain it over time. This is borne out by the comparison of intellectual property industries affected by pockets of monopoly (Shapiro, 2000).

Then the result is an intractable economic dilemma, which has thus far received very little attention: on the one hand, intellectual property rights can have the beneficial effect of stimulating the production of new ideas and competition, through the promise of temporary supraprofits; but on the other hand, because they introduce a monopolistic slant to the markets – to a greater or lesser extent depending on the conditions – they also foster the emergence of rent-seeking behaviours which gradually skew the competitive scenario, degrading its overall efficiency. To date, this drift has only been acknowledged in a few sporadic cases, where it was found that intellectual property rights can sometimes be manipulated for anticompetitive purposes.

The first precedent-setting case was probably *US* v *United Shoe Machinery*[30] in the United States, where both the antitrust authority and academics determined that an accumulation of intellectual property rights, albeit in respect of the law, could be used to restrict competition (Anderson, 1998).[31] Since then, similar cases have been described in the scientific and legal literature without, however, leading to a definitive and comprehensive theoretical paradigm. Some examples are the 'brand proliferation' strategy described by Schmalensee (1978), in which the ownership – and exploitation – of multiple trademarks within a given product category in reality conceals a strategy of foreclosing the market to potential competitors. In like manner, it has been shown that ownership of a large number of unexploited patents (commonly termed 'sleeping patents') has the real purpose of restricting the scope of competition open to newcomers (Gilbert and Newbery, 1982). By the same token, an incumbent might work around a principal patent – a practice known as 'inventing around' – with the sole purpose of producing sleeping patents and thereby locking out the competition.

Many sectors that produce intellectual property-protected goods are characterized, even today, by strategies such as these, which are open to conflicting interpretations. Now, the accumulation and the joint ownership of rights can be viewed not just as the signs of a market-foreclosure strategy, but also as practices aimed at minimizing risk – and which are therefore genuinely competitive – in markets characterized by great uncertainty (Ramello, 2003).

This ambiguity of interpretation crops up frequently in the literature and in practical cases, as a consequence of the fact that intellectual property is *ab origine* a monopoly space (albeit only potentially, and not always in practice) granted to creators as a reward for their activity. Hence, the anticompetitive behaviours of owners are to some degree consistent with the legal incentives provided to them, that is, rational and competitive within the altered economic framework.

Recent European and US legislation has repeatedly brought to light the tensions that exist on this matter. The European *NDC Health/IMS Health*[32] (National Data Corporation Health Information Services and Intercontinental Marketing Services Health Inc.) case, for example, clearly showed how exploitation of the legal monopoly granted by the right treads a fine line between legitimacy (under the intellectual property right laws) and illegitimacy (under the antitrust regulations). But even in the controversial US *Microsoft* case, where an abundance of disparate elements revealed a well-developed anticompetitive strategy, intellectual property rights were invoked to support the legitimacy of these behaviours.[33]

In general, therefore, the existence of market imperfections (Williamson, 1977), of complex economic strategies relating, for example, to interconnected markets and/or multiproduct firms (see De Vany and Walls, 1999 for

the extended video sector), of rent-seeking inertia arising from pockets of monopoly power, all have the combined effect of gradually altering the structure of the competitive scenario, causing it to drift away from the Schumpeterian model, progressively eroding the competitiveness and efficiency of the markets.

This dynamic effect therefore alters the overall balance of welfare, and so must be taken into account in the economic analysis of intellectual property.

Recent developments

The contributions discussed thus far all, in any case, refer to a scenario that could be described as obsolete from the technological, scientific and economic standpoints. Recent innovations have in fact already extensively called into question a regulatory system that is anchored to an out-of-date technological substrate, ignoring the effective rapid evolution and mutations of the technology landscape.

The lively contemporary debate, covered extensively in the media, reveals widely diverging views as to the function and effectiveness of intellectual property rights in the various current and future socioeconomic contexts. At the root of this disagreement are two specific, and in a sense opposing, phenomena. On the one hand, technological change has *de facto* weakened the right, by providing new tools for the duplication and dissemination of information, thereby endangering the consolidated interests of intellectual property stakeholders. This is what happened, for example, in the celebrated *Napster*[34] case, where it was not just the violation of copyright that was put on trial, but the new peer-to-peer file-sharing systems themselves.[35] So the advent of new behaviours made possible by novel technologies, coupled with reliance on traditional legal paradigms that no longer fully fit the altered context, poses a thorny dilemma that can be solved only by looking beyond vested interests. Without a doubt, blindly coming down on the side of intellectual property is not an optimal solution, because there is just as strong a need to preserve the innovative dynamic which, even if it compromises pre-existing interests as a side-effect, does not for this reason lose importance (Lessig, 2001).[36]

This idea finds solid support in the very same literature on innovation, pioneered by Schumpeter (1943), which in its time upheld the importance of intellectual property. Some recent authors have in fact gone back to this work as a starting point to rediscover Arrow's solution (1962) of an upstream remuneration system that does not disturb the competitive nature of the downstream market (Boldrin and Levine, 2002). On the other hand, the scientific and technological dynamic itself creates new opportunities for profit under the existing regulatory framework, thereby stimulating behaviours that have little to do with innovative investment, but are aimed instead at securing

new sources of income, if necessary by diverting resources away from other creative and productive activities (David, 2000). This state of affairs has been judged to exist, for example, in the case of the patenting of genetic material, where the current system is often manipulated for purely rent-seeking aims, with the not-inconsiderable side-effect of damaging scientific activity which is of necessity incremental, and relies upon appropriation of preceding knowledge (Barton, 2000; Bobrow and Thomas, 2001).[37] This threat is propagating across the entire sphere of open science and technology, which are increasingly encroached upon by the indiscriminate extension of proprietary systems and the resultant progressive erosion of the public domain, that is so necessary and essential for creative activities (David, 2000).

Generally speaking, today we are witnessing an unprecedented extension of traditional rights into new spheres, in a manner not always clearly justifiable by the logic of incentive. Some cases worth citing, among those currently being debated, are the patentability of business concepts (Merges, 1999) and software.

This relentless expansion of the current legal apparatus is generating a highly fragmented and complex system of rights, whose management incurs high transaction costs, with the effect of discouraging those types of creative activities that cannot afford these new costs. In other words, the markets of ideas – though rescued from the risk of failure due to free-riding – are now in danger of another type of failure arising from the proliferation of rights and the attendant heavy operating costs. This eventuality has been described in the literature as the 'tragedy of the anticommons' (Heller, 1998).

Finally, a number of quite recent contributions, driven also by the unprecedented creative dynamics of the Open Source software movement, which has even generated its own body of literature (see Lerner and Tirole, 2001; vol. 32, issue 7 of *Research Policy*, 2003), consider the possibility of formulating incentive paradigms for creative activities that are either alternative or complementary to intellectual property, and able to contain the costs of an overly burdensome legal system. In particular, these authors strive to better understand the processes involved in creativity and knowledge production by referring to specific studies from the other human sciences, invoking the 'gift' paradigm used in anthropology to at least partially explain the creative dynamic (Gordon, 2002, 2003; Zeitlyn, 2003), or the shared character of social processes that constitute the environment in which knowledge is produced (Benkler, 2003; Ramello, 2005).

Concluding remarks

This contribution attempts to systematize the law and economics theory as it relates to intellectual property rights, while at the same time suggesting new perspectives for analysis. In fact, as is discussed in the second and third

sections, the standard literature relies essentially on the thesis of an incentive to create and/or disclose new ideas. However, although this argument doubtless remains valid in the general case, it fails to satisfactorily take into account various consequences arising from the new legal institutions and the specific technological context. One important such consequence is the dynamic effect of intellectual property rights on the market structure of the sectors involved, which can at times interfere with the original competitive processes, or even drastically alter them.

An economic analysis based on these premises – though as yet fragmented and non-systematic – might reveal a different overall balance of welfare for the individual rights and therefore lead to different regulatory and policy indications. In particular, it would appear necessary to consider the peculiar social and productive attributes of the various markets of ideas, which require an ad hoc analysis, whereas reliance on a universal theory seems increasingly inadequate and conducive to inefficient economic results.

Notes

1. I am grateful to Hal Varian for his kind hospitality at the School of Information Management and Systems, University of California, Berkeley, where this chapter was drafted. I am equally greatly indebted for comments, suggestions and/or intellectual stimuli to Anna Maffioletti, Ahmed Silem, Francesco Silva, Vittorio Valli, several PhD students in law and economics at the universities of Siena, Trento and Carlo Cattaneo, and the members of the Ministero dell'Instruzione, dell'Università e della Ricerca, Italy, research groups on trademark and on intellectual property rights. The usual disclaimer applies. This work was supported by the MIUR research grant on 'The governance of intellectual property'.
2. We set aside, for the time being, the questions raised by certain authors as to the effective exclusivity of trade secrets, which will be discussed subsequently.
3. For a general overview, see Hardin (1968), Besen and Raskind (1991), David (1993), Besen (1998), Menell (1999) and Gordon (2003). On the possibility of alternative solutions to the private ownership of knowledge, see Arrow (1962).
4. For example, on intellectual property in Chinese culture, see Alford (1995). Again, on the problem of extending the Western doctrine to other cultures, with particular reference to Australian aboriginal society, see Blakeney (1998).
5. See *supra* n. 3.
6. And necessarily so, because it was only with Smith that economics emerged as a separate discipline in its own right.
7. On the theory of public goods and the dilemma of appropriability, a good review is given in Jha (1998, ch. 4).
8. For a historical (and dramatic) exposition of the modern theory of intellectual property rights, see the 'tragedy of the commons' by Hardin (1968). See *supra* n. 3.
9. For a more modern treatment connected with growth models, see Romer (1990).
10. See, again, Hardin (1968) and recently Heller (1998).
11. For a comparative presentation, see, for example, Metaxas-Maranghidis (1995) and Fawcett and Torremans (2001).
12. On this topic, see the Special Issue TRIPS (1998) of the *Journal of International Economic Law* and, in critical vein, Ryan (1998).
13. We avoid here a detailed legal explanation of the rights, referring the reader to the abundant literature available (see, for example, Bently and Sherman, 2001. See also Besen (1998) and Menell (1999) for a presentation focusing on economic analysis, but from a US perspective. Somewhat dated, but still of interest, is the contribution by David (1993).

14. For a concise but excellent introduction to the economic role of patents, see Oz (1995, pp. 244–6). A more extensive discussion is instead provided by Kitch (1977).
15. For a more in-depth analysis, see Gordon and Bone (1999) and Watt (2000).
16. For example, in the case of databases, protected in Europe subsequent to Directive 96/9/EC, the originality lies exclusively in the organizational format of the data.
17. In general such rights are of an economic character (so-called 'pecuniary rights'), that is, right of reproduction, right of distribution, right of public performance, right of public display, right of preparing derivative works; and of a moral character (moral rights), that is, right of paternity, right of publication and right of integrity (see, for instance, Bently and Sherman, 2001). On the (highly uncertain) role of such rights from a law and economics perspective, see Rushton (1998) and Hansmann and Santilli (1997).
18. See Ramello (2001).
19. As a rule, economic analysis deals with significantly shorter time horizons, which in no case extend beyond the lifetime of the individuals. One could of course argue that a perfectly informed copyright owner would be able to anticipate the expected profits, and recover them within his/her lifetime by selling the right in exchange for a remuneration. However, perfect information is by no means a feature of these markets where, on the contrary, the utmost uncertainty and asymmetry of information appear to reign (see, for example, Silva and Ramello, 2000; Lessig, 2001; Ramello, 2003).
20. For a more in-depth analysis, see Friedman et al. (1991) and Friedman (1998).
21. See Samuelson (2000) in particular n. 146 for a detailed discussion of the various positions taken in the debate on this point.
22. The different owners of the same trade secret must by definition be few in number and unaware of the fact; otherwise the protected information would effectively be disclosed, and hence not a valid subject matter for a trade secret.
23. In fact, in the case of multiple paternity the patent is attributed in the United States to the first who invents and in the EU to the first who registers. See Bently and Sherman (2001).
24. The differentiation strategies are in fact pursued by economic actors in order to secure a certain amount of market power. For an introduction to the topic, see, for example, Oz (1995, ch. 7). For an application to the case of trademarks, see Sappington and Wernerfelt (1985).
25. With a few exceptions that cast doubt only on the efficiency of the market, but not on its underlying structure, as in the case of excess innovation (see the historic contribution by Hirshleifer, 1971).
26. For the copyright industries, see, for example, Marvasti (2000) for the film industry, Silva and Ramello (2000) for the recording industry, and Armstrong (1999) for pay-TV. In general, see OECD (1998).
27. The equating of intellectual property with market power is certainly not automatic, as has been affirmed by the European Court of Justice (*Deutsche Gramophon GmbH* v. *Metro-SB-Grossmrkte GmbH*, 78/80, 8 June 1971, ECR 487) and by the US US FTC and DOJ (1995), *Antitrust Guidelines for the Licensing of Intellectual Property*.
28. The causal relation with interchangeability is still accepted in the US Guidelines (sect. 2.2) and in the European legislation, see *supra* n. 27.
29. See again Ramello (2003, p. 123).
30. 110 F. Supp. 295, D. Mass, 1953, aff'd *per curiam*, 347 US 251, 1954.
31. Although the first case in which a conflict between exclusive intellectual property rights arose was *United States* v. *Paramount Pictures Inc.*, 334 US 131 (1948).
32. COMP D3/38.044. This debate was launched by the 'Magil' case (*Radio Telefis Eireann (RTE)* v. *Commission of the European Communities* (C-241/91 P e C-242/91P, 6 April 1995).
33. 'Microsoft argues that the licence restrictions are legally justified because, in imposing them, Microsoft is simply "exercising its rights as the holder of valid copyrights" Appellant's Opening Br. at 102 … The company claims an absolute and unfettered right to use its intellectual property as it wishes: "[I]f intellectual property rights have been lawfully acquired,", it says, then "their subsequent exercise cannot give rise to antitrust liability."

Appellant's Opening Br. at 105', *USA* v. *Microsoft Corp*, US DC Court of Appeals, N. 00-5212 consolidated with 00-5213.

34. *A&M Records, Inc.* v. *Napster, Inc.*, 114 F. Supp. 2d 896 (N.D. Cal. 2000); *A&M Records, Inc.* v. *Napster, Inc.*, US Court of Appeals (9th Circ., 2001).
35. See, for example, Ramello (2001).
36. Lessig (2001, p. xvi) explicitly upholds the thesis of 'limiting the control that legal structures such as copyright give to the industries of yesterday to ensure that they can't use the law to constrain the creation of tomorrow'.
37. An OECD report (1998, p. 1997) states that 'IPR protection in some sectors (notably biotechnology) and countries may be so broad that it actually inhibits innovation'.

References

Alford, W.P. (1995), *To Steal a Book is an Elegant Offense. Intellectual Property Law in Chinese Civilization*, Stanford, CA: Stanford University Press.

Anderson, R.D. (1998), 'The interface between competition policy and intellectual property in the context of the international trading system', *Journal of International Economic Law*, **1**, 655–78.

Antill, J. and P. Coles (1996), 'Copyright duration: the European Community adopts "Three Score Years and Ten"', *European Intellectual Property Review*, **18**, 379–83.

Armstrong, M. (1999), 'Competition in the pay-TV market', *Journal of Japanese and International Economies*, **13**, 257–80.

Arora, A. and A. Gambardella (1997), 'Public policy towards science: picking stars or spreading the wealth?', *Revue d'Économie Industrielle*, **79**, 63–75.

Arrow, K.J. (1962), 'Economic welfare and the allocation of resources for invention', in R.R. Nelson (ed.), *The Rate and Direction of Inventive Activity*, Princeton, NJ: Princeton University Press, pp. 609–24.

Barton, J.H. (2000), 'Reforming the patent system', *Science*, **287**, 1933–4.

Benkler, Y. (2003), 'Coase's Penguin, or, Linux and *The Nature of the Firm*', *Yale Law Journal*, **112**, 369–446.

Bentham, Jeremy (1839), 'A manual of political economy', in Goldstein (1994), p. 174.

Bently, L. and B. Sherman (2001), *Intellectual Property Law*, Oxford: Oxford University Press.

Besen, S.M. (1998), 'Intellectual property', in P. Newman (ed.), *The New Palgrave Dictionary of Economics and the Law*, vol. 2, London: Macmillan, pp. 348–52.

Besen, S.M. and L.J. Raskind (1991), 'An introduction of the law and economics of intellectual property', *Journal of Economic Perspectives*, **5**, 3–27.

Blakeney, M. (1998), 'Communal intellectual property rights of indigenous peoples in cultural expressions', *Journal of World Intellectual Property*, **1**, 985–1002.

Bobrow, M. and S. Thomas (2001), 'Patents in a genetic age', *Nature*, **409**, 763–4.

Boldrin, M. and D. Levine (2002), 'The case against intellectual property', *American Economic Review*, **92**, 209–12.

Choi, J.P. (1998), 'Brand extension and informational leverage', *Review of Economic Studies*, **65**, 655–69.

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

Dasgupta, P. and P. David (1994), 'Towards a new economics of science', *Research Policy*, **23**, 487–532.

David, P. (1993), 'Intellectual property institutions and the panda's thumb. Patents, copyrights and trade secrets in economic theory and history', in M.B.Wallerstein, M.E. Mogee and R.A. Schoen (eds), *Global Dimensions of Intellectual Property Rights in Science and Technology*, Washington, DC: National Academy Press, pp. 19–62.

David, P.A. (2000), 'The digital technology boomerang: new intellectual property rights threaten "open science"', World Bank Conference.

De Vany, A.S. and W.D. Walls (1999), 'Uncertainty in the movie industry: does star power reduce the terror of the box office?', *Journal of Cultural Economics*, **23**, 285–318.

Drahos, P. (1996), *A Philosophy of Intellectual Property*, Aldershot: Dartmouth.

Evans, D.S. and R. Schmalensee (2001), 'Some economic aspects of antitrust analysis in dynamically competitive industries', NBER working paper no. 8268.

Fawcett, J.J. and P. Torremans (2001), *Intellectual Property and Private International Law*, Oxford: Oxford University Press.
Friedman, D.D. (1998), 'Trade secrets', in P. Newman (ed.), *The New Palgrave Dictionary of Economics and the Law*, vol. 3, London: Macmillan, pp. 604–607.
Friedman, D.D., W.M. Landes and R.A. Posner (1991), 'Some economics of trade secret law', *Journal of Economic Perspectives*, **5**, 61–72.
Gilbert, R.J. and D. Newbery (1982), 'Preemptive patenting and the persistence of monopoly', *American Economic Review*, **72**, 514–26.
Goldstein, P. (1994), *Copyright's Highway*, New York: Hill and Wang.
Gordon, W.J. (2002), 'Excuse and justification in the law of fair use: commodification and market perspectives', in N. Elkin-Koren and N.W. Netanel (eds), *The Commodification of Information*, The Hague: Kluwer, pp. 149–92.
Gordon, W.J. (2003), 'Intellectual property', in P. Can and M. Tushnet (eds), *The Oxford Handbook of Legal Studies*, Oxford: Oxford University Press, pp. 617–46.
Gordon, W.J. and R.G. Bone (1999), 'Copyright', in B. Bouckaert and G. de Geest (eds), *Encyclopedia of Law and Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 182–215.
Hansmann, H. and M. Santilli (1997), 'Authors' and artists' moral rights: a comparative legal and economic analysis', *Journal of Legal Studies*, **26**, 95–143.
Hardin, G. (1968), 'The tragedy of commons', *Science*, **162**, 1243–8.
Heller, M. (1998), 'The tragedy of anticommons: property in the transition from Marx to market', *Harvard Law Review*, **111**, 621–88.
Hirshleifer, J. (1971), 'The private and social value of information and the reward to inventive activity', *American Economic Review*, **61**, 561–74.
Jha, R. (1998), *Modern Public Economics*, London: Routledge.
Kitch, E.W. (1977), 'The nature and function of the patent system', *Journal of Law and Economics*, **20**, 265–90.
Landes, W.M. and R.A. Posner (1987), 'Trademark law: an economic perspective', *Journal of Law and Economics*, **30**, 265–309.
Lerner, J. and J. Tirole (2001), 'The open source movement: key research questions', *European Economic Review*, 45, 819–26.
Lessig, L. (2001), *The Future of Ideas*, NewYork: Random House.
Locke, J. (1690), *Second Treatise on Government*, accessed 15 July, 2003 at www.ilt.columbia.edu/academic/digitexts/locke/second/locke2nd.txt
Marvasti, A. (2000), 'Motion picture industry: economies of scale and trade', *International Journal of the Economics of Business*, **7**, 99–114.
Menell, P. (1999), 'Intellectual property: general theories', in B. Bouckaert and G. de Geest (eds), *Encyclopedia of Law and Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 129–88.
Merges, R. (1999) 'As many as six impossible patents before breakfast: property rights for business concepts and patent system reform', *Berkeley Law Technology Journal*, **14**, 577–615.
Metaxas-Maranghidis, G. (ed.) (1995), *Intellectual Property Laws of Europe*, Chichester: Wiley.
Organization for Economic Cooperation and Development (OECD) (1998), *Competition Policy and Intellectual Property Rights*, OECD, DAFFE/CLP(98)18, Paris: OECD.
Oz, S. (1995), *Industrial Organization*, Cambridge, MA: MIT Press.
Pepall, L.M. and D.J. Richards (2002), 'The simple economics of brand stretching', *Journal of Business*, **75**, 535–52.
Ramello, G.B. (2001), 'Napster et la musique en ligne. Le mythe du vase de Pandore se répèterait-il ?' ['Napster and online music. Is the Pandora box myth recurring again?'], *Réseaux*, **19**, 131–54.
Ramello, G.B. (2003), 'Copyright and antitrust issues', in W.J. Gordon and R. Watt (eds), *The Economics of Copyright: Developments in Research and Analysis*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 118–47.
Ramello, G.B. (2005), 'Private appropriability and sharing of knowledge: convergence or contradiction? The opposite tragedy of the creative commons', in L. Takeyama, W. Gordon

and R. Towse (eds), *Developments in the Economics of Copyright: Research and Analysis*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Romer, P. (1990), 'Endogenous technological change', *Journal of Political Economy*, **98**, 71–102.
Rushton, M. (1998), 'The moral rights of artists: droit moral ou droit pécuniaire', *Journal of Cultural Economics*, **22** (1), 15–32.
Ryan, M.P. (1998), *Knowledge Diplomacy: Global Competition and the Politics of Intellectual Property*, Washington, DC: Brookings Institution.
Samuelson, P. (2000), 'Privacy as intellectual property', *Stanford Law Review*, **52**, 1125–69.
Sappington, D.E.M. and B. Wernerfelt (1985), 'To brand or not to brand? A theoretical empirical question', *Journal of Business*, **58**, 279–83.
Schmalensee, R. (1978), 'Entry deterrence in the ready-to-eat breakfast cereal industry', *Bell Journal of Economics*, **9**, 305–27.
Schumpeter, J. (1943), *Capitalism, Socialism and Democracy*, London: Unwin.
Scotchmer, S. (1991), 'Standing on the shoulders of giants: cumulative research and the patent law', *Journal of Economic Perspectives*, **5**, 29–41.
Scotchmer, S. (1998), 'Incentives to innovate', in P. Newman (ed.), *The New Palgrave Dictionary of Economics and the Law*, vol. 2, London: Macmillan, pp. 273–7.
Shapiro, C. (2000), 'Competition policy in the information age', *Foundations of Competition Policy*, London: Routledge.
Silva, F. and G. Ramello (2000), 'Sound recording market: the ambiguous case of copyright and piracy', *Industrial and Corporate Change*, **9**, 415–42.
Smith, A. (1762), 'Lectures on jurisprudence', in P. Goldstein (1994), p. 173.
Special Issue TRIPS (1998), *Journal of International Economic Law*, **1** (4).
Strowel, A. (1993), *Droit d'auteur et copyright* [Author's Rights and Copyright], Brussels: Bruylant, and Paris: L.G.D.J.
Tadelis, S. (1999), 'What's in a name? Reputation as a tradable asset', *American Economic Review*, **89**, 548–63.
Watt, R. (2000), *Copyright and Economic Theory: Friends or Foes?*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Williamson, O.E. (1977), 'Predatory pricing. A strategic and welfare analysis', *Yale Law Journal*, **87**, 284–340.
Zeitlyn, D. (2003), 'Gift economies in the development of open source software: anthropological reflections', *Research Policy*, **32**, 1287–91.

10 Incomplete contracts and institutions

Antonio Nicita and Ugo Pagano

Introduction

The notion of incomplete contract has received increasing attention in law and economics literature as it permeates all economic problems that entail a legal solution. The incomplete contract framework is now applied to a wide range of research fields in law and economics, which cover, among other issues, the theory of the firm and economic organizations, the study of corporate governance and finance, the design of liberalization and privatization policies, the governance of intellectual property and international trade.

In perfect competitive markets the relevant action to be made by economic agents is that of a Pareto-relevant economic exchange (Bowles, 2003): given initial endowments, the mere existence of potential gains from trade generates that trade automatically. Under that framework, there is no need for legal rules or institutions to facilitate economic exchange. As the Coase theorem reveals, this is the tautological[1] outcome of perfect competitive markets: when transaction costs are negligible and property rights are well-defined, economic resources will automatically end up in the hands of those agents who value them the most, independently of any initial assignment of property rights on those resources. However, as long as transaction costs grow, potential Pareto-relevant exchanges could be inhibited, leading to an inefficient outcome. Since incompleteness raises relevant transaction costs, incomplete contracts are a potential source of market failure. In order to minimize transaction costs, an appropriate institutional environment has to be designed so as to enforce parties' obligations against opportunism and renegotiation. The law and economics of incomplete contracts thus refers to the compared analysis of transaction costs associated with the legal rules and private solutions designed to guarantee parties' performance in an incomplete contract framework.

Definition

According to the standard literature (Hart and Holmström, 1987; Milgrom and Roberts, 1990; Hart, 1995; Tirole, 1999), an incomplete contract is defined as an agreement whose contractual obligations are observable to contractual parties but not verifiable *ex post* by third parties, typically a judge or an arbitrator[2] to whom parties might eventually refer when controversies arise. The emergence of (a degree of) unverifiability on contractual terms

might be generated by several circumstances, such as: parties' bounded rationality and uncertainty concerning future events; high transaction costs incurred in writing and accurately describing any contractual feature (also called 'ink costs'), or going to court; and so on.

Starting from this standard definition of incomplete contracts, at least two approaches have been developed (Brousseau and Glachant, 2002): first, concerning the compared analysis of institutional arrangements designed to mitigate the effects of incompleteness on parties' incentives to perform; and second, a more recent approach focused on the analysis of the foundations of contractual incompleteness.[3]

Whatever is believed to be the origin of unverifiability, the economic reason why contractual incompleteness matters is that it may constitute a source of inefficiency when it inhibits Pareto-relevant exchanges. According to Williamson (1985), this inefficient outcome might emerge only when two other conditions are jointly satisfied: (i) the incomplete contract has to perform investments in specific assets; and (ii) at least one agent in the contract is opportunistic.[4]

The degree of asset specificity has been defined as 'the degree to which an asset cannot be redeployed to alternative uses and by alternative users without sacrifice of productive value' (Williamson, 1996: 59). Asset specificity thus creates a sort of *lock-in effect* which exposes the owners of specific assets to counterparts' economic dependency. Thus the sacrifice of economic value due to redeployment in alternative uses represents a measure of the parties' exit costs and the opportunity cost to dissipate contractual quasi-rents.

When specific assets are involved in incomplete contracts, the owner of the assets is locked in by the fact that the degree of asset specificity acts as a 'fundamental transformation' which reduces *ex post* the value of employing the assets in alternative uses (Williamson, 1985). This lock-in effect in turn generates the risk of opportunistic behaviour by contractual counterparts who may want to renegotiate terms in order to extract additional rents with respect to those contracted *ex ante* (the so-called 'hold-up problem'). The main consequence of linking together incomplete contracts, opportunism and asset specificity is that, under this framework, contractual parties maintain strong incentives to underinvest in asset specificity. As a consequence, the risk of a counterpart's opportunistic behaviour implies an unwillingness to generate potential quasi-rents.

The relevance of the analysis of incomplete contracts in law and economics stems from the selection of the legal and economic rules and institutions which might reduce the risk of post-contractual opportunism by optimally aligning parties' incentives to generate the highest level of specific investment (second-best outcome). In this respect, 'transaction cost economics' (Williamson, 1985), in particular, outlines a theory of private orderings (such

as contracts and economic organizations) as institutions performing alternative transaction costs to ensure the enforcement of incomplete contracts.

The tradeoff between opportunism and adaptation

One way of clarifying some essential features of incomplete contracts is to distinguish between exogenous and endogenous incompleteness. Roughly speaking, exogenous incomplete contracts refer to a dimension of contractual unverifiability which is independent of parties' actions. On the contrary, endogenous incomplete contracts refer to the idea that the degree of (un)verifiability in a contract could also be determined explicitly by contractual parties who may deliberately decide to leave unspecified some essential contractual term in the presence of uncertainty. This distinction is important since it outlines two opposite, and somehow contradictory effects of contractual incompleteness regarding parties' attitude to perform contractual obligations:

1. when the degree of unverifiability is exogenous, it weakens the probability that parties will achieve a contractual agreement in the first instance, given that at least one party could be exposed to a counterpart's post-contractual opportunism at the renegotiation stage (opportunism); and
2. when the degree of unverifiability is explicitly agreed upon by parties it may support contract formation and encourage parties' performance (adaptation).

As Coase (1937) outlines, there are economic contexts in which a contract, far from being detailed in every part, 'should only state the limits to the power' of one party on the other. This might be due to the fact that:

> [If] one contract is made for a longer period instead of several shorter ones, then certain costs of making each contract will be avoided … [especially when] owing to the difficulty of forecasting, the longer the period of the contract … the less desirable it is for the person purchasing to specify what the other contracting party is expected to do … All that is stated in the contract is the limits to what the person supplying the commodity or service is expected to do. The details of what the supplier is expected to do are not stated in the contract but are decided later by purchaser. (Coase, 1988, pp. 39–40)

The apparent contradiction between the relative efficiency of having a very detailed contract rather than a broader one relies upon the tradeoff between opportunism and adaptation in incomplete contracts. A very detailed contract, when feasible, might reduce the uncertainty on the degree of *ex post* verifiability, also inhibiting, however, any future efficiency-enhancing renegotiation of contractual terms. On the other hand, a broader and general

contract might favour *ex post* efficient adaptation of contractual terms after an uncertain contingency is realized, while at the same time increasing the risk of opportunistic behaviour.

The literature on incomplete contracts has mainly been concerned with the problem of opportunism, while only recently have the virtues of adaptation been pointed out, with particular reference to the compared efficiency of economic organization, such as firms, to deal with incomplete contract relationships (Williamson, 1996: pp. 228–9).

The following discussion outlines the economic context that characterizes, respectively, opportunism and adaptation, and it also illustrates some possible institutional solutions for mitigating the inefficiency of contractual incompleteness.

The hold-up problem

To illustrate the hold-up problem, consider the case of a company A producing automobile bodies and a company B producing automobiles. Imagine that A actually produces a standard automobile body which might equally be sold to B and to B's competitors C and D. The market value of the automobile crafted with the standard body is €40 000, while the price at which a body is sold by A to automobile companies is €20 000 per unit. Assume now that A and B meet to sign a contract for a new sophisticated automobile body to be produced by A specifically for B. The contract is typically incomplete, given that for the parties involved it would be very costly to specify detailed contractual terms according to all possible contingencies which may arise from a contract signed (at $t = 0$) according to A's performance (at $t = 1$). On the other hand, automobile bodies are specific investments given that, once produced with a specific design tailored for B, they preserve their economic value only if acquired by B. Assume that this specific investment amounts to a value s, such that $0 < s <$ €30 000 for each body unit and that the market price for the new model produced by *B* is equal to €100 000. At $t = 0$, A and B agree upon a surplus sharing rule which gives to *A* and *B*, respectively, half of the new model market price (€50 000 each). However, since the contract is not verifiable by third parties, A's production decisions, after $t = 0$, are exclusively determined by A's expectations of B's attitude either to cooperate or to renegotiate on contractual terms. Figure 10.1 illustrates how A will make his/her decisions.

Assume first that A decides to make a specific investment. What about B? B may choose to fulfil contractual obligations (*commitment*) and pay, say, €50 000 to A; or B may hold up A and threaten to exit the contract unless a price equal to A's best outside option (€20 000) is obtained at the renegotiation stage. In this last case (hold-up), B will extract a rent equal to €30 000 from A. Let us assume that at $t = 1$ B is likely to adopt an opportunistic

		Agent B	
		Commitment	Hold-up
Agent A	Standard investment	20 000, 20 000	20 000, 20 000
	Specific investment	50 000 – *s*, 50 000	20 000 – *s*, 80 000

Figure 10.1 The unilateral hold-up problem

behaviour, refusing to transfer the contractually agreed-upon price at $t = 0$. Will A select the specific investment? Of course not, given that by producing a standard body, he/she can always obtain €20 000, without incurring, in the case of the counterpart's hold-up, the monetary loss *s* associated with specific investments. This is the underinvestment outcome generated by the risk of hold-up by counterparts. Underinvestment leads to an inefficient outcome (standard investment) given that a potential Pareto-relevant exchange is truncated and a potential social surplus (equal to €60 000 – *s*) is entirely dissipated. One could ask whether the case of bilateral specific investments differs somehow from the above case. Assume then that the automobile producer B has also made an investment *b* which is specific to the body produced by A (for instance, the engine's dimension). The intuition is that with bilateral specific investments, parties may have a strong incentive both to reciprocally commit to contractual obligations and to share the maximum social surplus. However, as Figure 10.2 clearly shows, unless parties are able to implement some reciprocal commitment device, both may maintain strong incentives to delay investment choice until the counterpart has been committed to the contract. However, as long as one party commits to fulfilling the contract, the counterpart maintains strong incentives to hold up. As a consequence nobody will be induced to invest and investment decisions might be delayed indefinitely. The resulting equilibrium will be the inefficient one, characterized by bilateral underinvestments with a complete dissipation of the potential social surplus, equal to [60 000 – (*s* + *b*)], which would have been generated by specific investments.

		Agent B	
		Delayed investment	Specific investment
Agent A	Delayed investment	20 000, 20 000	80 000 – *s*, (20 000 – *b*)
	Specific investment	(20 000 – *s*), 80 000 – *b*	50 000 – *s*, 50 000 – *b*

Figure 10.2 The bilateral hold-up problem

The problem of adaptation

The example above illustrates how the risk of post-contractual opportunistic behaviour affects *ex ante* incentives to invest in welfare-enhancing specific assets. It is easy to show how, in a world of complete contracts, a penalty like $p° = €30\,000$ incurred by the renegotiating party, may align parties' incentives to invest efficiently after the contract has been signed. This may suggest to rational parties that they should determine *ex ante*, even in an incomplete contracts framework in which investments are not contractible, a very detailed contract at least in terms of pricing, exclusive clauses and breach penalties. However, writing a contract with verifiable fixed prices and/or breach penalties may generate a 'double hazard problem' since the potential victim of hold-up might be induced to underinvest after he/she has received a powerful safeguard. A fixed contract price may thus prevent renegotiation even in those cases in which it might be efficient for parties to renegotiate in order to adapt to new unforeseen circumstances. Two well-known examples of the inefficiency associated with fix-price contracts are given by *Fisher Body v. General Motors* (Klein et al., 1978; Klein, 1996) and *Alcoa v. Essex* (Speidel, 1981; Goldberg, 1985; Klein, 1996):

1. *Fisher Body v. General Motors* Fisher Body was an automobile body supplier. In order to produce the automobile bodies, Fisher Body had to make specific investments in stamping machines and dies. Due to the high degree of unverifiability on the nature of specific investments, any contract signed by Fisher Body was potentially vulnerable to a counterpart's hold-up. In 1919, Fisher Body signed a long-term contract with General Motors (GM) for the supply of closed-metal autobodies, containing several provisions aimed at protecting Fisher Body against a possible GM hold-up. A first safeguard was given by an exclusivity clause which obliged GM to buy all of its closed-metal autobodies from Fisher. Moreover, the contract defined a pricing formula for autobodies based on a cost-plus rule for which the final price was determined by labour and transportation costs plus a mark-up to cover capital costs. Two other contractual clauses ('most-favoured nation' and 'meeting competition'[5]) were aimed at preventing Fisher Body from exploiting its contractual power against GM. Between 1919 and 1924, however, the market registered a huge and unforeseen change in demand: wooden bodies were rapidly replaced by metal autobodies. This exogenous change in automobile bodies contrasted with the original pricing provisions agreed upon by the parties, which revealed that prices for metal bodies were too high. According to the traditional interpretation of this case, Fisher Body refused to renegotiate the pricing formula or to satisfy GM's request to locate Fisher plants next to GM ones, so as to at least reduce

transportation costs. The refusal to renegotiate by the *ex ante* most vulnerable party corresponds to a hold-up behaviour induced by rigid contractual clauses rather than by contractual incompleteness. In this case, contractual rigidity impeded the efficient *ex post* renegotiation of contractual terms which would have led to adaptation to new unforeseen contingencies.

2. *Alcoa v. Essex*[6] Essex was an aluminium cable manufacturer while Alcoa was an aluminium facility producer. In order to allow shipments of processed aluminium from Alcoa to Essex in molten form and thereby reduce production costs, Essex located its plant next to Alcoa's plants. This *site specificity* increased Essex's dependency on Alcoa's *ex post* hold-up. In order to protect Essex from Alcoa's post-contractual opportunism, Essex entered into a long-term contract with Alcoa, specifying predetermined price formula and output rates for Alcoa's processing of alumina into aluminium for Essex. The pricing formula linked the price for Essex to a wholesale price index for industrial commodities. However, after the 1973 crude oil supply crisis, electricity costs (which represented the most important non-labour cost in aluminium production) began to rise much more rapidly than the wholesale price index. As a consequence, the price at which Essex was receiving aluminium from Alcoa was set at a net cost of less than one-half of the contemporary market price of aluminium. This unforeseeable change in electricity costs resulted in turn in an unexpected gain for Essex at the expense of Alcoa. As Klein (1996) points out 'the enforcement by Essex of the literal terms of this imperfect contract can be considered a hold-up since it can be assumed to be contrary to the original intent of the contractual understanding'.

In both cases, *Fisher Body v. General Motors* and *Alcoa/Essex*, the *ex ante* potential victim of hold-up, after obtaining exclusivity clauses and fix-price contracts as safeguards against renegotiation, became itself the opportunistic party as some unforeseen exogenous change in market dynamics transferred to that party all the *ex post* bargaining power at the renegotiation stage.

The two examples above illustrate the tradeoff between opportunism and adaptation and outline how the pursuit of efficiency in an incomplete contracts framework always requires a complex governance structure which assigns to the most vulnerable party *ex ante* appropriate safeguards against opportunism without incurring the opposite risk of shifting contractual dependency on to the counterpart, when unanticipated changes in market conditions affect parties' *ex post* incentives to fulfil contractual obligations.

Institutional and contractual solutions for incomplete contracts

Several solutions have been investigated in order to minimize the risk of hold-up in incomplete contracts. Some may be jointly and autonomously implemented by economic agents; others rely on the emergence of an institutional setting which induces parties to actively cooperate by means of sanctions and enforcement devices based on reputation, social norms and private orderings.

Investments partitioning

A first solution for implementing specific investments in an incomplete contract framework is to split the expected specific investments into small and verifiable subinvestments (Pitchford and Snyder, 2004). In this case, the original contract is replaced by a series of smaller contracts, each governing the exchange of a portion of the original investment. At each stage the decision to continue to invest in the future depends on the previously observed behaviour of the counterpart. If both parties fulfil their contractual obligations at each stage then the contract is endogenously enforced. As long as the relationship continues, each party becomes more and more specific to the other, thus also raising the opportunity costs of breaching the contract. However, investment partitioning might be implemented as a solution only when the nature of investments makes it possible. Unfortunately, most specific investments often imply one-shot large-scale investments and thus some alternative enforcement device is required to optimally align parties' incentives to invest.

Vertical integration

Vertical integration is another solution for incomplete contracts. This refers to the idea of generating appropriate incentives to invest in specific assets by assigning to the investing party the right to be a *residual claimant* on the surplus generated. This could be done by transferring to the investing party the property rights on assets involved in the contract. With reference to the above example of Fisher Body and General Motors, Klein et al. (1978) have concluded that GM's decision to acquire the total amount of Fisher Body's shares in 1926 represented a way of overcoming – by vertical integration – Fisher Body's hold-up.[7] Vertical integration thus induces optimal incentives to invest by assigning to the investing party the right to the residual income generated, once other factors of production have been accounted for. This explanation has also provided an incomplete contract-based theory of the firm along the original intuition of Coase (1937): when the *ex post* transaction costs involved in incomplete contracts make it convenient to internalize market transactions into an integrated governance structure, a second-best solution to incomplete contracts is provided by vertical integration within a single firm (Williamson, 1985). Firms thus emerge as institutions of private

orderings governing incomplete contracts characterized by specific investments.

Authority and residual control rights

Vertical integration is based on the idea that assigning residual income to the investing party is a viable way to induce efficient levels of investments. However, as Hart (1995) pointed out, in many contexts 'residual income is not well defined'. For instance, in profit-sharing contracts each party is a residual claimant, but this does not imply that parties maintain appropriate incentives to invest. Moreover, the notion of residual income outlines only one of the features characterizing a property right. As a consequence, in order to understand the role played by vertical integration in enforcing incomplete contracts, it is necessary to investigate the functions performed by a property right. As Furubotn and Richter (1997) outlined, a property right embodies the right to use the asset (*usus*), the right to appropriate return from the asset (*usus fructus*) and the right to change its form, substance and location (*abusus*).[8] The reason why property rights assignment matters in a world of incomplete contracts is thus provided by the fact the property right gives the assets' owner the residual control rights over that asset, that is, with 'the right to decide all usages of the assets in any way' (Hart, 1995). This is why, in an incomplete contract world, 'ownership is a source of power' (p. 23).

Having residual control rights confers on the owner the power to take care of unspecified contingencies and to organize the production process involving own assets. This power is also defined as *authority*. The authority relationship from one side reduces the degree of contractual incompleteness by assigning to him/her the power to decide what to do when unforeseen contingencies arise; from the other, it induces efficient levels of investment from the owner's side, by assigning to the owner all the bargaining power in the *ex post* renegotiation stage. Residual control rights thus represent a powerful way of tackling the tradeoff between opportunism and adaptation in incomplete contracts. As the Grossman–Hart–Moore (GHM) theories provided by Grossman and Hart (1986) and Hart and Moore (1990) pointed out, the efficient assignment of residual control rights depends strictly on the nature of investments and on the degree of substitutability or complementarity among the agents involved in a given transaction. Given that property rights induce owners' optimal incentives to invest, while minimizing non-owners' incentives to invest, the (second-best) efficient assignment of property rights should be decided according to agents' ability to maximize social surplus.[9] Under this setting, when assets are independent, vertical integration will not help to increase parties' incentives to invest in specific assets, whereas when assets are strictly complementary they should be owned jointly.

Simple contracts on ex post *bargaining*

The assignment of property right is only one way – possibly the most convenient in terms of transaction cost minimization – to attribute authority to one party in a contract. Since authority is associated with the attribution of bargaining power at the renegotiation stage, it is also possible to imagine that parties may assign authority by contract. A recent field of research on incomplete contracts focuses on the emergence of first-best investment choices in 'simple contracts' characterized by some exogenous bargaining rule at the renegotiation stage (Aghion et al., 1994; Nöldeke and Schmidt, 1995; Edlin and Reichelstein, 1996). The basic intuition is that of defining a 'procedural verifiability' in a renegotiation game according to which the *ex post* bargaining can be designed *ex ante* (for a survey, see Schmitz, 2001). In this case it is sufficient to give one party with all the *ex post* bargaining power (which corresponds to the previous case of assigning residual rights to control). When there are positive gains from trade, the party which has all the *ex post* bargaining power will make a take-it-or-leave-it offer to the counterpart, and the first best will be achieved (Schmitz, 2001). Aghion et al. (1994) imagine the case in which parties may write a contract which specifies the nature of the investments in the presence of uncertainty on future contingencies that may affect the number of widgets exchanged within parties. They assume that parties may write specific performance contracts which determine a new *ex post* default point in the renegotiation stage. Parties defining the *ex post* default point so as to induce one party to efficiently invest, thus giving the other counterpart all the bargaining power at the renegotiation stage, will induce efficient bilateral investments. The Aghion et al. model implements a first-best contract, thanks to some restrictive assumptions (Hart, 1995) such as that parties sustain no transaction costs in writing and enforcing simple contracts.

Corporate culture, reputation and trust in implicit contracts

As Bowles and Gintis (1993) pointed out: 'the Walrasian general equilibrium model is based on an artificially truncated concept of self-interested behavior, depicting a charming but utopian world in which conflicts abound but a promise is a promise'. The idea that moral commitment may act as an enforcement device in incomplete contract has been studied, among others, by Kreps (1990), Crémer (1993), Lazear (1995) and Hodgson (1996), who emphasized the role played by reputation in favouring implicit and self-enforcing cooperation in repeated interactions. In many situations, the endogenous enforcement devices represented by trigger strategies and optimal penal codes in infinitely repeated prisoner dilemmas might even be cheaper than devising detailed and complicated contracts. Given that with trigger strategies players' deviations from cooperating are punished by other

players refusing to cooperate in the future, a 'good reputation' represents a powerful tool to minimize transaction costs in incomplete contracts (Hermalin, 1999) and to enhance trust of other counterparts. In this respect, firms should be interpreted as repeated players who may develop an internal system of contractual enforcement based on reputation and 'corporate culture', that is, on a dominant set of norms which guides the way in which work is accomplished within the organization.

Incomplete contracts and market dynamics

Fisher Body v. General Motors and *Essex v. Alcoa* have shown the relevance of market dynamics in affecting parties' incentives to either fulfil or renegotiate incomplete contracts. Notwithstanding the relevance of market dynamics in incomplete contracts, most of the theories outlined above are generally based on the assumption that agents' outside options are exogenous, focusing mainly on bilateral relationships.[10] The analysis of incomplete contracts characterized by specific investments has thus been confined to the Williamsonian 'fundamental transformation' (Williamson, 1985), for which an *ex ante* competitive transaction is *ex post* transformed into a bilateral monopoly. According to this perspective, the level of *ex ante* parties' outside options acts as a default point in the *ex ante* contracting game and as a threat point in the *ex post* bargaining over the joint surplus. The 'market' is implicitly assumed to be an *equilibrium market* and hence for contractual parties it is not possible to affect (and to be affected by) competitors' strategies.

One way of analysing how market dynamics affects incomplete contract theories is to shift from a bilateral contract to a complex transaction in which contract and market dynamics are interdependent, according to the original intuition outlined by Commons (1924, 1934, 1970; see also Nicita, 2001). As Commons (1970) has emphasized:

> [W]hen we reduce all prospective buyers and sellers upon a given market to those who participate in one bargaining transaction as our smallest unit of investigation, then they are the 'best' two buyers and the 'best' two sellers, meaning the two buyers who offer the highest prices and the two sellers who offer to accept the lowest prices, in consideration of transfers of ownership[11] … The best two sellers are those able to sell at the lowest price. They compete for choice of alternatives offered by the best two buyers, those able to buy at the highest prices, while, in turn, the best two buyers are competing for choice of alternatives offered by the best two sellers.

As a consequence, under a transaction, 'instead of the "exchange" of physical things between two parties, as contemplated in the former physical economics, there are five parties, all of whom are "potential" and then they are successively "actual" participants in the lawful alienation and acquisition of

ownership'. These five parties are four competitors (two buyers and two sellers) and the 'enforcer' or judge who is 'ready to issue commands to any of the buyers and sellers in the name of sovereignty, if any dispute arises'.

The above framework is very useful for analysing incomplete contracts since it shows that when contracts are incomplete a 'five parties transaction' does not necessarily collapse into a bilateral relationship, as in Williamson (1985), but it still involves four agents (contractual parties and their respective competitors). In other words, the notion of transaction outlined by Commons always maintains a market dimension within the framework of incomplete contracts. In this setting, parties' decisions are made considering not only the impact of investment decisions on contractual counterparts but also the impact exerted on the outside market. This has several important consequences. For instance, with endogenous outside options, parties may have strong incentives to overinvest in asset specificity when investments increase own outside options and reduce the counterparts' outside options (Nicita, 2001). Moreover, endogenous outside options may reverse some of the main conclusions of the GHM models. De Meza and Lockwood (1998) show that in some cases an agent with an important investment decision should not own the assets he/she works with; in other cases independent assets should be owned jointly whereas strictly complementary assets should be owned separately. The analysis of market-contract dynamics thus constitutes an interesting field for future research on incomplete contracts.

Incomplete contracts and institutional complementarity

Another interesting, and quite unexplored, field of research for incomplete contracts is given by the analysis of institutional complementarities in an incomplete contract framework characterized by multiple equilibria in which initial conditions are destined to reinforce over time. Standard theories on incomplete contracts seem to be very fruitful in explaining the efficiency of given governance structures but they fail to provide an explanation for their inefficiency, that is, for the selection of an inefficient equilibrium where multiple alternative equilibria are potentially available for economic agents. Under the framework already analysed, there is virtually no explanation for the persistence of an inefficient allocation of property rights over time. As Aoki (2001) pointed out, the notion of institutional complementarity may provide a useful tool to analyse how equilibria are selected in an incomplete contract framework characterized by multiple equilibria. The notion of institutional complementarity relies on the idea that, in a given institutional framework characterized by incomplete contracts, economic agents face different domains and do not strategically coordinate their choices across domain games. As a consequence, the institutional choices in one domain act as exogenous parameters in other domains and constitute the 'institutional envi-

ronment' under which choices are being made. In this setting 'one type of institution rather than another becomes viable in one domain, when a fitting institution is present in another domain and vice-versa' (Aoki 2001). Assume two domains of choices X and Y and, respectively, two sets of choices $\{X_1, X_2\}$ and $\{Y_1, Y_2\}$, with agents i choosing in X and agents j choosing in Y, according to their utilities (respectively, u for i and v for j). Standard conditions of institutional complementarity are defined by the two following circumstances (see Pagano, 2003):

1. for agent i, the additional benefit of having institution X_1 instead of institution X_2 in domain X is greater when institution Y_1 (instead of institution Y_2) is chosen in the domain Y: $u(X_1; Y_1) - u(X_2; Y_1) \geq u(X_1; Y_2) - u(X_2; Y_2)$;
2. for agent j, the additional benefit of having institution Y_2 instead of institution Y_1 in some domain Y is greater when institution X_2 (instead of institution X_1) is chosen in the domain X: $v(Y_2; X_2) - v(Y_1; X_2) \geq v(Y_2; X_1) - v(Y_1; X_1)$.

The above conditions, considered by Aoki (2001), restate in terms of institutional choices the super-modularity conditions among strategies considered by Milgrom and Roberts (1990) and are concerned with the property of incremental pay-offs with respect to a change in parameter value. They do not exclude the possibility that the level of the pay-offs of one rule is strictly higher than that of the other for the agents of one domain or of both domains, regardless of the choice of rule in the other domain. In other words, there is the possibility of a unique equilibrium. However, under the super-modularity condition, there can be two pure Nash equilibria (*institutional arrangements*) for the system that comprises X and Y, that is (X_1, Y_1) and (X_2, Y_2). When such multiple equilibria are possible, we say that domains X and Y are institutional complements of each other and that: X_1 and Y_1 are *institutional complements*; and X_2 and Y_2 are *institutional complements*. Moreover, as Aoki (2001) points out, when multiple equilibria exist, it is possible that the overall institutional arrangement could result in a Pareto-inferior outcome. For instance, suppose that (X_2, Y_2) is such that $u(X_2; Y_2) - u(X_1; Y_1) > 0$ and $v(X_2; Y_2) - v(X_1; Y_1) > 0$. However if initial conditions are such that X_1 or Y_1 is selected in one of the two domains so as to act as a parameter in the other one, this selection will induce the choice of the institutional complement, respectively, Y_1 or X_1. As a consequence, the equilibrium selected will be (X_1, Y_1) rather than (X_2, Y_2), that is, the Pareto-inferior outcome.

The analysis of institutional complementarities in an incomplete contract framework has the following implications: (i) the interdependence among domains may generate multiple institutional arrangements; (ii) according to

initial conditions affecting the available choices in one domain, some Pareto-inferior institutional arrangement may emerge; (iii) since institutional arrangements are pure Nash equilibria, they are self-enforcing in nature and destined to perpetuate over time (by path dependency and cumulative causation) unless some exogenous change affects one domain or the other so as to shift the choice to another institutional arrangement.

This approach has been employed to analyse, in an incomplete contract framework, the emergence of self-enforcing path-dependent equilibria in corporate governance between technological and financial domains, the variety of capitalistic systems, and the structuring of alternative capitalistic systems in the United States, Germany and Japan (for instance, see Aoki, 2001 and Nicita and Pagano, 2003, 2005).

Conclusions

This chapter has briefly outlined the main features of incomplete contracts and their relevance for law and economics research. While the standard literature on incomplete contracts is mainly concerned with the hold-up problem, the analysis of *Fisher Body v. General Motors* and *Alcoa v. Essex* has highlighted the centrality of the problem of adaptation in incomplete contracts and the tradeoff between adaptation and opportunism which pervades the law and economics of incomplete contracts and the design of appropriate institutions for their governance. Several solutions have been indicated in the literature to address the above tradeoff: vertical integration, authority relationship, simple contracts and reputation-based implicit contracts. All these solutions provide a powerful theoretical framework to explain the emergence of economic and legal institutions as transaction cost-minimizing devices in an incomplete contract framework. According to Coase (1988), Williamson (2003) and Hart (1995), proprietary integration of specific assets under a unified ownership within the firm should be explained in terms of firms' compared ability to reduce transaction costs in carrying out a market transaction, with respect to spot market exchanges. Some recent extensions have shown, however, how some of the main conclusions coming from standard literature on incomplete contracts could be reversed when considering endogenous outside options as well as the emergence of institutional complementarity in a multiple equilibria setting. These extensions may represent a powerful tool of analysis for future research in the law and economics of incomplete contracts.

Notes

1. On the tautology contained in the Coase theorem, see Usher (1998).
2. A complete contract is thus a contract whose terms are also observable to third parties. According to Shavell (2003): 'A contract will be said to be completely specified (or simply complete) if the list of conditions on which the actions are based is explicitly exhaustive, that is, if the contract provides literally for each and every possible condition

in some relevant universe of conditions. In a contract for a photographer to take wedding photographs, suppose that the universe of conditions is everything that could happen to the photographer (becoming ill, receiving an offer to take photographs at another wedding the same day) and everything that could happen to the wedding couple (becoming ill themselves, breaking off their engagement). A completely specified contract would then have to include an explicit provision for each of these possible conditions pertaining to the photographer and to the wedding couple'. Milgrom and Roberts (1992) follow a different approach, defining a complete contract as any agreement which might always be completed *ex post*.

3. This entry will focus on the first approach. For a survey of the debate concerning the foundation of incomplete contracts, see Hart and Moore (1999); Maskin and Moore (1999); Maskin and Tirole (1999); MacLeod (2002).
4. Williamson (1979, 234, n. 3) defines opportunism as 'self-interest seeking with guile. This includes but is scarcely limited to more blatant forms, such as lying, stealing and cheating. Opportunism more often involves subtle forms of deceit. … More generally, opportunism refers to the incomplete or distorted disclosure of information, especially to calculated efforts to mislead, distort, obfuscate, or otherwise confuse'. For a critique of this notion of opportunism, see Bowles and Gintis (1993).
5. The *most-favoured nation clause* meant that Fisher would be prevented from charging General Motors a higher price than that charged to other customers. The *meeting-competition clause* implied that Fisher was prevented from charging a price higher than the average market price for the autobodies.
6. *Aluminum Co. of America* v. *Essex Group, Inc.*, 499 F.Supp. 53 (W.D. Pa. 1980).
7. The original explanation for the merger between Fisher Body and General Motors due to Klein et al. (1978) has recently been debated by several scholars who pointed out that GM controlled 60 per cent of Fisher Body's shares several years before Fisher Body's refusal to comply with GM's requests. See Langlois and Robertson (1989), Helper et al. (1997), Bolton and Scharfstein (1998), Casadesus-Masanell and Spulber (2000), Coase (2000), Freeland (2000) and Pagano (2000).
8. For a survey on the economic notion of property rights, see also Libecap (2002).
9. This conclusion integrates the Coasean theory of the firm and the Coase theorem. According to Coase (1937), when transaction costs matter, market exchange is replaced by internalized transactions within firms, and firms are characterized by long-term contracts governed by authority relationships. According to Coase (1960), when transaction costs are relevant, the initial assignment of property rights matters for the purpose of efficiency.
10. Only some recent papers are explicitly concerned with investment decisions in a market environment, as Chatterjee and Chiu (1999) and de Meza and Lockwood (1998). The analysis of contractual enforcement in a market context is also studied by MacLeod and Malcomson (1993) and Bolton and Whinston (1993).
11. According to Commons, every economic exchange implies alienation and acquisition of property rights in a commodity, whereas the word 'commodity' identifies not only the merely possession of a physical thing but also a 'lawful ownership' (1970: 48–9).

References

Aghion, P., M. Dewatripont and P. Rey (1994), 'Renegotiation design with unverifiable information', *Econometrica*, **62**, 257—82.

Aoki, M. (2001), *Toward a Comparative Institutional Analysis*, Boston, MA: MIT Press.

Bolton, P. and D. Scharfstein (1998), 'Corporate finance, the theory of the firm, and organizations', *Journal of Economic Perspectives*, **12**, 95–114.

Bolton, P. and M. Whinston (1993), 'Incomplete contracts, vertical integration, and supply assurance', *Review of Economic Studies*, **60**, 121–48.

Bowles, S. (2003), *Microeconomics: Behavior, Institutions, and Evolution*, Princeton, NJ: Princeton University Press.

Bowles, S. and H. Gintis (1993), 'The revenge of homo economicus: post-Walrasian economics and the revival of political economy', *Journal of Economic Perspectives* **7** (1), 83–102.

Brousseau, E. and J.M. Glachant (eds) (2002), *The Economics of Contracts*, Cambridge: Cambridge University Press.
Casadesus-Masanell, R. and D.F. Spulber (2000), 'The fable of Fisher Body', *Journal of Law and Economics*, **43**, 67–104.
Chatterjee, K. and Y.S. Chiu (2000), 'When does competition lead to efficient investments?', paper presented at the Econometric Society World Conference, Seattle, WA.
Coase, R. (1937), 'The nature of the firm', *Economica*, November, 386–405.
Coase, R. (1960), 'The problem of social cost', *Journal of Law and Economics*, **3**, 1–44.
Coase, R. (1988), *The Firm, the Market and the Law*, Chicago: University of Chicago Press.
Coase, R.H. (2000), 'The acquisition of Fisher Body by General Motors ', *Journal of Law and Economics*, **43** (1) , April, 15–31.
Commons, J.R. (1924), *Legal Foundations of Capitalism*, London: Macmillan.
Commons, J.R. (1934), *Institutional Economics*, Madison, WI: University of Wisconsin Press.
Commons, J.R. (1970), *The Economics of Collective Action*, Madison, WI: University of Wisconsin Press.
Crémer J. (1993), 'Corporate culture and shared knowledge', *Industrial and Corporate Change*, **2** (3), 351–86.
de Meza, D. and B. Lockwood (1998), 'Does asset ownership always motivate managers? outside options and the property rights theory of the firm', *Quarterly Journal of Economics*, **2**.
Edlin, A.S. and S. Reichelstein (1996), 'Holdups, standard breach remedies and optimal investment', *American Economic Review*, **86**, 478–501.
Freeland, R.F. (2000), 'Creating holdup through vertical integration: Fisher Body revisited', *Journal of Law and Economics*, **43**, 32–65.
Furubotn, E. and R. Richter (1997), *Institutions and Economic Theory*, Ann Arbor, MI: University of Michigan Press.
Goldberg, V.P. (1985), 'Price adjustment in long-term contracts', *Wisconsin Law Review*, **3**, 527–43.
Grossman, S.J. and O.D. Hart (1986), 'The costs and benefits of ownership: a theory of vertical and lateral integration', *Journal of Political Economy*, **94** (4), August, 691–719.
Hart, O. (1995), *Firms, Contracts and Financial Structure*, Oxford: Oxford University Press.
Hart, O. and B. Holmström (1987), 'The theory of contracts', in T. Bewley (ed.), *Advances in Economic Theory*, Cambridge: Cambridge University Press, pp. 71–155.
Hart, O. and J. Moore (1990), 'Property rights and the nature of the firm', *Journal of Political Economy*, **98**, 1119–58.
Hart, O. and J. Moore (1999) 'Foundations of incomplete contracts', *Review of Economic Studies*, **66**, 115–38.
Helper, S., J.P. MacDuffie and C.F. Sabel (2000), 'Pragmatic collaborations: advancing knowledge while controlling opportunism', *Industrial and Corporate Change*, **9** (3), 443–83.
Hermalin, B.E. (1999), 'Economics and corporate culture', SSRN, mimeo.
Hodgson, G. (1996), 'Corporate culture and the nature of the firm', in J. Groenewegen (ed.), *Transaction Cost Economics and Beyond*, Boston, MA: Kluwer Academic Press.
Klein, B. (1996) 'Why hold-ups occur: the self-enforcing range of contractual relationships', *Economic Inquiry*, **34** (3).
Klein, Benjamin (2000), 'Fisher–General Motors and nature of the firm', *Journal of Law and Economics*, **43** (1), April, 105–41.
Klein, B., R.G. Crawford and A. Alchian (1978), 'Vertical integration, appropriable rents and the competitive contracting process', *Journal of Law and Economics*, **21** 297.
Kreps, D. (1990), 'Corporate culture and economic theory', in J. Alt and K. Shepsle (eds), *Perspectives on Positive Political Economy*, Cambridge: Cambridge University Press.
Langlois, R.N. and P.L. Robertson (1989), 'Explaining vertical integration: lessons from the American automobile industry', *Journal of Economic History*, **49** (2), 361–75.
Lazear, E.P. (1995), 'Corporate culture and the diffusion of values', in Horst Siebert (ed.), *Trends in Business Organization*, Tübingen: J.C.B. Mohr (Paul Siebeck).
Libecap, G.D. (2002), 'A transaction-costs approach to the analysis of property rights', in Brousseau and Glachant (eds), pp. 000.

MacLeod, B. (2002), 'Complexity and contract', in Brousseau and Glachant (eds), pp. 000.
MacLeod, B. and J. Malcomson, (1993), 'Investments, holdup, and the form of market contracts', *American Economic Review*, **83**, 811–37.
Maskin, E. and J. Moore (1999), 'Implementation with renegotiation', *Review of Economic Studies*, **66**, 39–56.
Maskin, E. and J. Tirole (1999), 'Unforeseen contingencies and incomplete contracts', *Review of Economic Studies*, **66**, 83–114.
Milgrom, P. and J. Roberts (1990), 'Rationalizability, learning and equilibrium games with strategic complementarities', *Econometrica*, **59**, 511–28.
Milgrom, P. and J. Roberts (1992), *Economics, Organization, and Management*, Englewood Cliffs, NJ: Prentice-Hall.
Nicita, A. (2001), 'The firm as an evolutionary enforcement device', in A. Nicita and U. Pagano, *The Evolution of Economic Diversity*, London: Routledge.
Nicita, A. and U. Pagano (2003), 'Corporate governance and institutional complementarities: the co-evolution of "finance" and "technology"', paper presented at the Thirteenth Annual Meeting of the American Law and Economics Association, 20–21 September.
Nicita, A. and U. Pagano (2005), 'Law and economics in retrospect', in E. Brousseau and J.M. Glachant, *New Institutional Economics: A Guidebook*, Cambridge: Cambridge University Press.
Nöldeke, G. and K.M. Schmidt (1995), 'Option contracts and renegotiation: a solution to the hold-up problem', *Rand Journal of Economics*, **26**, 163–79.
Pagano, U. (2000), 'Public markets, private orderings and corporate governance', *International Review of Law and Economics*, **20** (4), 453–77.
Pagano, U. (2005), 'Legal positions and institutional complementarities' in F. Cafaggi, A. Nicita and U. Pagano (eds), *Legal Orderings and Economic Institutions*, London: Routledge.
Pitchford, R. and C.M. Snyder (2004), 'A solution to the hold-up problem involving gradual investment', *Journal of Economic Theory*, **114** (1), 88–103.
Schmitz, P.W. (2001), 'The hold-up problem and incomplete contracts: a survey of recent topics in contract theory', *Bulletin of Economic Research*, **53** (1), 1–17.
Shavell, S. (2003), *Foundations of Economic Analysis of Law*, Cambridge, MA: Harvard University Press.
Speidel, R.E. (1981), 'Court-imposed price adjustments under long-term supply contracts' *Northwestern University Law Review*, **76** (3), 369–422.
Tirole, J. (1999), 'Incomplete contracts: where do we stand?', *Econometrica*, **67** (4), 741–81.
Topkis, D.M. (1998), *Supermodularity and Complementarity*, Princeton, NJ: Princeton University Press.
Usher, D. (1998), 'The Coase theorem is tautological, incoherent or wrong', *Economics Letters*, **61**.
Williamson, O. (1979), 'Transaction-cost economics: the governance of contractual relations', *Journal of Law and Economics*, **22**, 233–61.
Williamson, O.E. (1985), *The Economic Institutions of Capitalism*, New York: Free Press.
Williamson, O.E. (1996), *The Mechanisms of Governance*, Oxford: Oxford University Press.
Williamson, O.E. (2002) 'Contract and economic organization', in Brousseau and Glachant (eds), pp. 000.

PART III

PUBLIC LAW AND ECONOMICS

11 Central Bank

Željko Šević

Introduction

Central banks as social institutions have not attracted a great deal of attention from researchers in law and economics. The central bank is referred to as the bearer of banking supervision duties, so mainstream law and economics literature has tended to be concerned only with the central bank's regulatory function. However, this old social institution has a far broader role than that seen up to now, and it gives wide research opportunities in law and economics, and increasingly in law and finance (La Porta et al., 1996; 1997). The central bank as an institution has a particularly interesting history.

Central bank: history and its (*natural*) development

When the first central banks or, more correctly, governmentally-sponsored banks, were incorporated, such as the Swedish Riksbank and the Bank of England in 1668 and 1694, respectively, they were entrusted with a monopoly over money issuing in the metropolitan area or in a part of the country. In the majority of cases the central bank was, for a while, the only joint-stock bank in the country. Usually, this market advantage was 'paid' as a credit directly extended to the government. In some cases the establishment of a central bank had a nationalist impetus, as in the case of lately created national states, like Germany and Italy in the 1870s (Šević, 1996b), although some authors note that they were established to unify what was a quite chaotic system of note issue (Goodhart, 1988). In contrast to all the good points in the English and Prussian cases, the emergence of the central bank's monopoly had significant negative influences on the development of (commercial) banking in its early stage. Therefore, it is quite useful to compare the English and Scottish experience with 'central banking' and 'free banking', respectively (White, 1995). When creating central banks, the legislator did not think that central banking would go much beyond the simple issuing of national currency. However, modern central banks perform a number of duties besides issuing legal tender.

With the emergence of imperialism in the late nineteenth century, central banks started to be concerned about the metal reserves of the country, allowing them to facilitate the national payment system. History shows that the growth of central banking is closely connected with nineteenth century, especially its last quarter. The central bank in France was incorporated in 1800

(Boyer-Fraise, 1903), Norges Bank and Austrian National Bank in 1816, and in Denmark a year later. At the end of the nineteenth century, central banks were established in Portugal, Bulgaria, Romania, Serbia (Šević, 1996a) Egypt, Algeria, Turkey, Japan and Java (Kent, 1966). However, widely accepted opinion is that central banking started to be much more important in the twentieth century (Kent, 1966; Šević, 1996b). Besides, prior to 1900, economic theory was concerned with how note issue should be regulated, whether the metal reserves should be centralized and, consequently, how the central bank should establish control over them.

Also central banks are interesting institutions since they are, generally, in public hands and therefore entities regulated by public law. Even the banks which began to operate under private ownership (the vast majority of banks established before the First World War) were nationalized after the Second World War. The years immediately after the Second World War were shown to be a turning point of the central bank's engagement in commercial banking operations. From the 1950s, it was quite unusual to observe commercial banking operations being undertaken by a central bank, in a two-tier banking (financial) system. In so-called 'socialist countries', the mono-bank usually called a state bank resumed both commercial and central banking operations. Strictly speaking, the latter function was not really performed since the 'socialist society' was theoretically 'moneyless' (Šević, 1997a).

In our view central banking has had a quite natural development following the requirements of the system whose centre it is. A number of authors support this approach (Friedman, 1959; Goodhart, 1985, 1988, 1994, 1995; Šević, 1996b, 1999), in contrast to advocates of free banking theory who oppose it (Klein, 1974; Karekan and Wallace, 1984; White, 1984, 1995; Selgin and White, 1987; Selgin, 1988, 1994). It is generally agreed that free banking had its 'golden age' in the nineteenth century, but in a diversified, fast-growing economy with increased information flows, it is necessary to supervise the system and provide more harmonic development and growth. For instance, interest in the soundness of banking systems emerged in the 1880s–90s with capital concentration and the appearance of 'financial capital' (Hilferding, 1920). This is probably best recognized in German banking theory where governmentally-sponsored (privileged) banks are classified into three groups: (i) currency or central banks; (ii) note (issuing) banks and (iii) bankers' banks or reserve banks (Veit, 1969). This can also be followed through the historical development of German banking literature (Walcher, 1876; Helander, 1916; Steiner, 1924; Deckert, 1926, among others).

The central bank as, currently, an entity in the publicly regulated regime is quite strongly regulated by law (statute, act) and other legal acts derived from law. It is noted that both *de facto* and *de jure* independence of a central bank influences its performance as the centre of the financial sys-

tem. This is a possible avenue for the extension of research on the central bank in law and economics. Its path from a privately owned, but privileged (governmentally-sponsored) bank to the entity owned and closely monitored by the state is another reason to attract researchers' attention. Also, the widely accepted opinion is that a central bank's macroeconomic performance depends on its institutional features. So, what is a central bank today? What are its status, functions and social significance?

Central bank: position

The formal position of a central bank within a society is, as a rule, established by law enacted by the legislative body (Parliament), whilst in some developing countries the *de jure* position of a central bank may be regulated by the governmental. The former practice is regarded as normal, whilst the latter is usually encountered in non-democratic countries or those which practice some kind of command economy (Šević, 1996b). Although there are more than 160 central banks worldwide, it is almost impossible to find two whose status is the same. A common theoretical definition says that the central bank is the chief institution of a country's monetary system and executive body for the implementation of monetary policy. In most cases this definition is sustainable, but not in all. This is the situation where the central banks are *de jure* independent while, if they are dependent, they just execute government decisions on monetary, credit and financial policy. In theoretical models a central bank is a state-owned or quasi-state institution, regulated by public law, which is established in order to maintain the macro-economic liquidity of the financial system as a whole and prevent mass bankruptcy of (commercial) banks and a high rate of unemployment. A central bank is not incorporated to be profitable, although its note-issuing function can be a source of significant revenue for the government – seigniorage (Šević, 1996; 1997c). Such a definition begs the question of whether the central bank is a bank at all.

It seems so, at least on account of its two (contemporary) duties: government banker and lender-of-last-resort (bankers' bank). Besides exercising these two functions, the central bank can even be regarded as a specific entity established by law and regulated within the public law regime (Šević, 1996b). The respected British economist, Richard Sayers, argues the same: 'The central bank is the organ of government which undertakes the major financial operations of the Government and, which by its conduct of these operations and by other means, influences the behaviour of financial institutions so as to support the economic policy of the Government' (Sayers, 1960, p. 64). He also found that central banks are different from commercial banks in a number of ways. Central banks are usually governed by persons closely related to the government, and they do not seek to maximize profit, which is generally seen

as the main aim of the commercial banks. The central bank must establish and maintain the special, close relationship with the (commercial) banking sector, so as to be able to influence this sector in its implementation of agreed government economic policy (Sayers, 1960). However, we would not go as far as Sayers in concluding that the central bank is an organ of government which must be 'in some sense a part of the government machine' (ibid., p. 64). The point made here is that there is a significant difference in the way the 'state' is perceived in Continental European and Anglo-Saxon general legal culture. The Europeans developed the sense of 'state' as an eternal and continuous social institution of utmost social importance, while the Anglo-Americans developed the sense of 'government' as a current social power.

Central banks' *raisons d'être* lies in the fact that

i they provide the best solution with regard to the minimization of the risks inherent in modern (commercial) banking by providing an efficient lender-of-last-resort (LLR) function;
ii they provide an institutional mechanism to minimize or avoid the profligacy of governments in their fiscal roles, and
iii in a contemporary economy they have the means of effecting control over monetary variables so as to provide conditions appropriate for the improvement in real economic activity (Šević, 1996b).

A central bank can also have a development role, especially in a developing economy. In contrast, the actual existence of central banks has its contradictions. Central banks must be engaged in the resolution of conflicts which emerge from (long-term) financial intermediation (money market failures, for instance), in the government's power over money creation and the conflicts which might arise in the usual business cycle slowdowns. These must be effectively managed at both micro and macro levels. So all these perplexities determine, to some extent, the *de facto* position and prospective roles of the central bank. The *de jure* position, usually depends primarily on the vision of the legislator. Subsequent corrections of an initial view are not only normal but also highly recommended. In the field of central banking regulation, to stick to tradition cannot only be socially ineffective, but also very costly, because the central bank is prevented from responding appropriately and in timely fashion to innovations in the banking sector and money market.

Although the majority of countries have some kind of central monetary authority, the organizational form may vary. Thus, the central monetary authority can take form of: (i) a unique central bank; (ii) a complex system of the central bank(s); (iii) a unique central bank for several countries (so-called 'supranational' central bank); (iv) a separate government agency which operates as a central bank; (v) a number of commercial banks which are entrusted

with a note-issuing function and some other duties characteristic of a central bank; (vi) a special state body in charge of business finance, which assumes some central bank duties; (vii) a transitional central banking institution; (viii) a central bank in a formal monetary union, where the central bank of one country performs central bank duties on the (economic) territory of another country and (ix) a monetary board (Šević, 1996b). At the time of writing there are only two countries without a central bank or properly defined separate central monetary authorities: Andorra and the Federated States of Micronesia. The former is a small country, a former protectorate, tax and duty 'haven', which circulates both the French franc and Spanish peseta, while the latter is a newly liberated country. As has been noted (Šević, 1996a, 1996b), at the time of large social changes (revolutions, coups, upheavals, and so on), it is usually expected that the ministry of finance will assume some of the central bank duties and especially note issuing. The series of notes thus issued are referred to, in theory, as notes of state (government) issue, but, the duration for which the ministry is engaged in performing this duty differs. Usually, after a few months the central bank resumes all its functions, albeit closely monitored and supervised by the government.

The position and structure, as well as duties, of a central bank are fully legally stipulated. There is no possibility of assuming that the central bank has a duty if it is not clearly stipulated in the Central Bank Act. The structure of a central bank is usually designed in such a manner as to sustain the fulfilment of functions (roles) set out by law. So what are the functions of the central bank?

Central bank: functions

Regardless of its position and organizational form, a modern central bank has to carry out several specific tasks in order to achieve an optimum level of economic stability. The reasons for the existence of a central bank are: (i) protection of the internal value of the national currency; (ii) protection of the external value of the national currency; (iii) maintenance of 'health' of the national financial system; (iv) ensuring balanced economic growth and development and (v) the development of financial (money and foreign exchange) markets. Generally speaking, regulations do not impose a hierarchy of requirements and tasks on the central bank. Even in countries with a similar economic structure and level of development, the central bank has quite different priorities. For instance, in Guyana and Jamaica, economic development is given a priority, while the Bahamas accepts a more classical concept, preferring price stability (monetary stability). So, in practice, the legal framework can define either one target or a set of them for the central bank. (Šević and Šević, 1998). In the past a range of targets was preferred, but it seems now that one (principal) target is more common. It is argued that a precisely

defined target can eliminate all possible perplexities and facilitate the central bank's efforts in its achievement. This is especially true for transitional and developing economies with a poorly developed financial structure (Šević, 1996b). However, countries require the central bank to perform well and to influence the financial flows in the domestic economy, and to stress growth and development, the sustainable growth of real output, high employment and price stability. Even the Americans have not opted for a unique definition of the task, defining the objectives of their economic policy thus: '… the objectives of U.S. monetary policy are high employment, stable prices (no inflation) and growth in output on a sustainable basis. The ultimate goals, as they are often described, are not directly under control of the Federal Reserve' (FRB of San Francisco, 1987). However, it has been observed that 'when monetary policy aims at several objectives simultaneously, with the need for choice and balance between them, policy will be subject to greater political oversight and the CB [central bank] will be subservient' (Goodhart, 1994, p. 1427). Consequently, greater autonomy can be achieved when the central bank must meet one single objective. Also, in the situation where there is a set of objectives, the principal–agent problem appears. The senior management is in a position to define priorities, and in some cases may overlook the primary (ultimate) goal. Moral hazard and adverse selection situations are likely to appear.

However, even if the central bank has only one objective, this is usually very broadly defined. 'Price stability' may be the ultimate goal, but how can it be achieved? The legal position of the bank makes necessary presumptions, but does not guarantee the effectiveness of the central bank's actions. Central bankers must find an immediate target to allow them to achieve an acceptable final outcome. This is why central banks, especially in small, open economies aim at something more feasible to control, such as the exchange rate. Central banks in market-based economies applying monetary policy affect the yields on financial markets. Consequently, the exchange rate is an important asset price that can be targeted (Goodhart, 1995).

New Zealand's experience with the central bank contract has opened up another debate. The problem there is how the central bank can fight inflation (maintain price stability; that is no inflation at all), if it has been given an a priori inflation target, usually up to 4 per cent. The concept of price stability conflicts with previously agreed and usually politically sponsored inflation levels. This approach has shown itself quite efficient in the New Zealand case, although there are doubts over its long-run effects (Tietmeyer, 1994). With a contractual obligation to keep inflation below an agreed level, the central bank is forced, in practice, to abandon the immediate target strategy, and to adopt 'one-stage strategies with direct price target' (ibid.). Anyway, practice has shown that central bank officials, as a rule, prefer to have a final

price objective, rather than to define and follow an immediate target. This can be an argument in favour of a central bank contract. The practice of New Zealand, Canada and the United Kingdom has shown that the final objective can be achieved even when immediate targeting is abandoned. In New Zealand, even wage bargaining between trade unions and the government is always in 'the shadow' of the central bank contract. Trade unions' desire to have a pay rise of 8 per cent and more is, as a rule, defeated by the NZ Reserve Bank, which limits the rise of all inputs of production up to the desired rate of inflation, in order to fulfil the contract.

The conduct of monetary policy is currently the most important duty of the central bank and it takes a great amount of a central bank's time and efforts. However, the functions of the central banks are much broader than to meet the (primary) objective of monetary policy. Theoretically they are defined as (i) issuing and cancelling banknotes and coins; (ii) foreign exchange reserve management; (iii) banker of the state (government's banker); (iv) bankers' bank, when acting as an LLR, (v) supervision of banks and other financial institutions, and (vi) facilitating the payment system (Šević, 1996b). Also the central bank can be involved in foreign exchange control. Even in the defining of the other functions, there are large differences among countries. For instance, in Germany, Canada, Belgium and some Latin American countries, supervision of banks is carried out by structures not belonging to the central bank. In the United States, the responsibility for bank supervision is shared by the Federal Reserve System (the Fed), the Comptroller of the Currency, the Federal Deposit Insurance Corporation (FDIC), and the state bank supervision authorities (Spong, 1990). Functions within foreign exchange operations and control are very often shared with the state administration authority in charge of fiscal affairs (usually Ministry of Finance). A close connection between the central bank, even when it is independent, and the government is required when it is necessary to coordinate monetary and fiscal policies. In this situation the legal and institutional framework cannot help greatly, since it is always a question of mutual prestige.

The formal obligation of the central bank to deliver the target level of inflation can be a way to assess the achievements of the central banks management. The central bank's governor (president) may be reappointed if he or she succeeds in achieving the target(s) stipulated in the contract. However, there is also the case when the central bank exhibits the 'public' part of its nature: the central bank officials do not want to have their rewards linked to their results (Goodhart, 1994). They rather opt to have flat, stable civil servants' salaries. However, if the bank is to make price stability a primary and ultimate objective, control must be tighter, and an incentive system properly developed. This is another question where the law and economics literature can provide answers. Central bank independence is always legally defined,

but its application in practice attracts special attention, because it affects overall central bank performance.

Central Bank: Independence

Central bank independence, in particular, has attracted the attention of researchers in this field, who focus attention on *de facto* and *de jure* independence. Policy makers who are in senior central bank management sometimes question whether independence can really deliver anything. Paul Volcker, former Chairman of FED, stressed in the late 1980s that although the Banca d'Italia is a 'dependent' central bank, it performed well, even in a country with quite an unstable political situation, that is, government (Volcker, 1989). Thus the central bank, as a part of the public sector, obviously has close contacts with other governmental bodies. Consequently, it is subject to the influence of the government, and particularly its ministry of finance. However, the degree of (in)direct influence can vary significantly (Šević, 1996c).

Initially, the study of central bank independence was carried out through reading central bank statues (laws) rather than through a proper in-depth analysis of policy decision-making and enforcement of rules. North (1993) bluntly admitted that enforcement is a major issue in the process of ensuring one's credible commitment. Central bank independence, very narrowly understood as independence from the government of the day, became the 'must' of the 1990s. Certainly, even countries that did not experience any (significant) inflation, like Italy in 1992, Portugal in 1992, Belgium in 1993, France in 1993, Greece in 1993 and Spain in 1994, decided to delegate their monetary policy formulation to their central banks, making them independent of the executive power (Šević, 1996d). This trend was continued by Japan, which in the mid-1990s experienced deflation, but decided to opt for central bank independence. Consequently, it is very difficult to prove that high or hyper-inflation is the main motive for promoting central bank independence, as suggested by many neo-classical scholars (see, for instance, Blinder, 1999). Others, like Maxfield (1997), could not corroborate this conclusion, as it seems that making the central bank independent in developing countries, which can be easily extended to transitional economies, has not yielded any significant economic result. One of the intuitive explanations for this lack of a direct link could be that, in developing countries, discretion rather than rules dominates the political sphere. Informal links and networks are those that influence the actions of the central bank, even if the latter is *stricto lege* independent. Of course, there are questions of effective and efficient law enforcement and legal systems (as a set of rules) and legal order (as a social behaviour based on prescribed legal rules). In developed countries there is a minor, if not negligible difference between the legal system and legal order,

and this may be one of the reasons why many studies in central bank policy-making assumed that the policy-making process is predominantly, if not exclusively, directed by statute (statutory rules). Forder (1996; 2001) documents these trends in great depth. Still the research must begin somewhere, and considering the formal documents may be the first step to initiate further analysis and discussion.

Observing the relationship between the government (executive power) and the central bank seems logical, but quite narrow. In the best tradition of law and economics literature and legal research, the relationship between the central bank and all the state powers – legislative, executive and judiciary – must be considered. Legislative power can usually influence the central bank in three ways: (i) through personnel appointments; (ii) through changes in the legal framework (law, act, statute, and so on) and (iii) through a general political monitoring function. The influence on personnel can be seen, alternatively, through three types of relationships. First, the parliament is entrusted by law to appoint a governor, his deputy and/or vice-governor(s), and board of directors. Second, the parliament approves the appointment of the governor and senior management made by another government body; or third, some members of the parliament are appointed as directors on the central banks' board. In practice, the parliament usually appoints the governor and board of directors. In countries with a strong presidential influence, the president may be in charge of appointing the governor and directors. However, even in a pure presidential system, such as in the United States, the upper house of the US Congress, the Senate, must approve the appointments. In Cuba there is quite an interesting solution, since there is a shared authority between the National Assembly (parliament) and the State Council (head of state) to appoint the governor upon a mutual proposal made by the president of State Council and the prime minister (Art. 47. Decree-law 84, 1984). A more important influence that the parliament can impose is changing the law. Even in the United States when the Fed failed to show understanding and support for the general political course of the Congress, proposals emerged to change the Federal Reserve Act. Generally, when the central bank enforces a tight monetary policy for a long period of time, without some short-term relaxation, there are incentives to change the law and abolish or reduce the achieved level of (political) independence (Šević, 1996b). It is stressed by many authors (for example, Rogoff, 1985; Alesina and Tabelini, 1987; Cukierman, 1992; 1994) that the central bank is usually much more concerned with price stability (its ultimate task) than are the political authorities.

Finally, the legislative body, as by definition the highest democratic body in the country, is in charge of political monitoring of all institutions within the political system. In this way, the parliament can have particular 'moral suasion' influence over the central bank. Public criticism addressed towards

the central bank has no legal implications, but can urge the bank's senior management to reconsider some of their decisions. In a democratic country the parliament is the most democratically elected body and has particular social prestige and direct responsibility before the electorate. The central bank, as with all the other social institutions with appointed senior personnel, has only a derived, indirect responsibility. Again in the majority of democratic countries, tradition ensures that the central bank is a socially responsible and highly publicly appreciated institution. The early experience of some countries (Britain, Austria, Germany and Serbia; see Kent, 1966) showed that the 'personal charisma' of the governor is a very important element, which influences further development. Political supervision is performed through the central bank's obligation to submit yearly reports for approval to the parliament. Even if the report is to be submitted to the president, it will usually be finally assessed in the parliamentary session. This is the case even in the parliamentary system with a strong president (so-called 'quasi-presidential system') as in France. In our view the overall political, that is 'democratic', control over the central bank must stay in the hands of the parliament as the most widely elected political body, with overall responsibility for the well-being of the country.

The relationship between the central bank and the judiciary has not been widely considered in the theory. Economists have usually seen the central bank as an executive body for the implementation of monetary policy and, in that capacity, subject to the influence of the government. But, given the public nature of the central bank, it is to be expected that the courts may examine decisions of the central bank as they have all the necessary characteristics of an administrative act. There are countries such as Bahrain, where it is clearly stipulated in the law that the legality of central bank acts will be examined by the courts. Another possibility is not to define procedural subjectivity, but to define its special legal position (as in Germany). The Austrian law on central banks – NBG (*Nationalbankgesetz*, 1984) – has stipulated the establishment of an ad hoc arbitrage. In the Austrian case, this legal solution granted *de facto* indirect independence to the National Bank, although it is formally subordinated to the government. The NBG is very detailed in regard to the arbitrage, and regulates all the procedure before it, saying that the rules on civil procedure will be applied analogously. In countries with a continental legal tradition, it should be expected that every central bank legal act resulting from administrative procedure may be subject to examination of legality before the Administrative or High Court.

However, theory is usually most interested in the relationship between the central bank and the executive power (government). Comparative analysis of central bank laws has shown that the government can perform some personnel and general supervisory duties over the central bank in the United States.

Usually by law, governmental influence can be: (i) personnel and (ii) monitoring (Šević, 1996b). There is no country where the government has no influence at all over the bank's appointments at senior level. In some countries the executive power proposes the candidates, while in others it appoints directors of the central bank. It is quite rare for a particular minister to appoint leading officials at the bank, as is the case in Iceland. This is of particular interest since the central bank performs duties as the government's banker. Researchers try to measure central bank independence in many different ways, but there is a consensus that it is a very complex task since central bank independence is determined by a multitude of legal, institutional, cultural and personal factors which cannot be easily quantified (Cukierman, 1991, 1992, 1994). In the recent literature, attention has been paid to legal independence, the actual turnover of governors and answers by national policy makers to a questionnaire on the behaviour of the central bank in practice. This complex approach has notably been employed by Cukierman (Cukierman, 1992). Other more 'classical' authors paid attention to the institutional features and inflation volatility (Alesina and Summers, 1993; Wood, et al., 1993). In both approaches the authors consider legal proxies, such as the position of law towards issues such as the legally determined mandate in office for the governor and senior management, who appoints them, who can dismiss them and under what circumstances, who prevails in the case of policy conflicts between the central bank and the government, whether the government can issue orders to the central bank, how independent the central bank is from a fiscal point of view, what restrictions are imposed on central bank lending to the government and other entities in the public sector, and so on. Also considered is whether the law requires the central bank to achieve price stability (even at the cost of other real objectives). The empirically based studies have shown that the more independent central banks are better at achieving price stability and vice versa. At the same time, high and long-term inflation erodes central bank independence. Although researchers tend to stick to other, more quantifiable, variables rather than considering legal and institutional framework, it seems that in the end they have to go back to this. The institutional framework affects not only the structure of social entities, but also the way in which they interact (Šević, 1995).

Central Bank: Credibility, Accountability, and Social Responsibility

In order to discharge its statutory duties the central bank must be credible. If there is no credibility in the actions of the central bank, it is most unlikely that monetary policy goals will be achieved. It is widely believed that the post Second World War success of the Deutsche Bundesbank was largely due to the general social consensus that monetary stability is in the national interest (Goodman, 1992; Hall and Franzese, 1998; McNamara and Jones, 1996),

although there are also opinions which doubt that the German economic miracle was due to the Bundesbank prior to the 1970s, and especially in the early years of post-war modern Germany (Holtfrerich, 1999, as opposed to Goodman, 1992). As we have already seen, in order to strengthen credibility in the 1990s, a world-wide argument developed for the institutional independence of the central bank, fairly narrowly understood, as breaking mutual dependence with the national government. It has traditionally been perceived that politicians are short-sighted and interested only in election victory, while sound monetary policy requires a long view (due to implementation time lag), a high level of professional knowledge and a reason to exclude democratic (or populist) pressures of the day, and consistency. Other economic policies can suffer less damaging effects from the political factor of the day by being closely involved in decision-making. Certainly, credibility, or rather credible commitments, are very important for the good functioning of institutions (North, 1993). If there is no emphasis on the proper functioning of any social institutions, it is difficult to expect that any credibility will be built. As we have already seen, independence was largely perceived to be a prerequisite for central bank credibility in the 1990s with 'an excuse' of the need to fight inflation and to prevent the political business cycles motive to be prevalent in monetary policy decision-making. However, the credibility of the monetary authority has its public component that relies heavily on the public perception of the central bank and its commitment to the promulgated goal(s). Even before the very popular 'independent central bank movement' of the 1990s there was a claim that policy makers should make known their intention to maintain low inflation despite the benefits of surprises (Kydland and Prescott, 1977). So, the usual argument that governments are generally obsessed by short-term goals, especially on the eve of general elections, may stand to some extent; but it is very difficult to perceive that someone may claim political credibility by promulgating unsound policies of any kind. It is hard to believe that ill-conceived policy-making could land someone a landslide election victory. Also, there must be some level of general trust towards public institutions to see central bank policies endorsed and central bank projections supported by the public. Often in the dominant literature the concept of social trust is somewhat forgotten.

However, as we have pointed out, at the very beginning of the twenty-first century we believe that only an independent central bank delivers sound monetary policy, although practice has shown that central bank independence does not necessarily fight inflation as such. Consequently, the main focus shifted to the processes of formal 'professionalisation' of the central bank in order to strengthen the overall social institutional framework. The recent research suggests a much wider delegation in policy decision-making, and an even stronger formal independence of the central bank, although it is not

completely clear that there is a causal link between bank independence and effective monetary policy. It seems that it is, at least, a European trend to see the central bank as a supra-organization which acts as an independent agency. However, in the New Public Management concept, independent agencies are accountable to the government department or to the cabinet. If the supranational European Central Bank is brought into the picture, the situation becomes even more elusive. The European Central Bank was made independent of the national central banks that create the system of European central banks, of the relevant national governments, and finally of the European Commission.

So, how can the problem of accountability be resolved? In a classical model, the central bank is expected to submit reports once or, more usually, twice a year to the state authority. In some cases it was submitted to the government, in others to the President or the legislative power – the parliament. The independence the central banks gained in the 1990s led to a situation where monetary policy decision-making was equated with a 'holy grail' and a belief that only highly qualified professionals would be able to deliver the expected results. As some scholars proved in their analysis (see for instance: Berman and McNamara, 1999), there is no logical argument to claim that monetary policy is special compared to other areas of public policy making. So the arguments from one side calling for the 'delegation' of decision-making authority to independent agencies, one of them being the central bank, require that the problem of accountability be addressed. In order words, should those who delegated continue to control the 'delegees'? In general, principal-agent theory would argue in favour of control. However, advocates of central bank independence have a problem with that, claiming it may indirectly hamper the independence of the central bank and its ability to discharge its functions effectively (McNamara, 2002). This creates a problem, as central bank independence has three elements: personnel, financial and operational, as has been pointed out in the previous section. The first measures the degree of involvement of the bodies external to the central bank in the process of selection and appointment of the bank's officials. The second analyses the budgetary and financial operative independence of the central bank. Generally there is no problem if seigniorage revenue has to be transferred to the government, but the important issues are whether the bank can be forced to overshoot monetary targets or to buy government securities against its own decision. The third generally measures to what extent short- and long-term decisions have been influenced by factors outside the bank or by motives which are political in their very nature (Šević, 1996b).

In the end, the question remains as how to define central bank accountability. It seems that one way is to strengthen the link between the central bank and the legislative body, assuming that there is a difference between democratic oversight of the policy outcomes and policy implementation. Claims

that 'democracy can be dysfunctional for the economy' and positive outcomes achieved from delegation therefore 'can be argued to outweigh concerns about the loss of democratic accountability' (McNamara, 2002) can be supported for the process of decision-making, but the bank must be accountable to someone for the policy results within its mandate over a longer period of time. In order to build a proactive relationship between the parliament and the central bank, it is necessary to set clear statutory rules which require the central bank to report to parliament on its performance. The parliament has the supreme democratic legitimacy within the country and ultimately is responsible for the strategic direction of the country. It is more than logical to see parliamentary oversight of the central bank performance within its defined period of time. Certainly, claims that monetary policy suffers from longer policy lags can be overcome through definition of a periodic performance review, which can be set by law or through a performance contract between the 'nation' (represented by the parliament) and the central bank (as an independent public agency). The shifting focus from mere credibility to accountability may result in the central bank being more socially responsible, and by its nature better integrated into the society which it is supposed to serve. But, social responsibility experiences problems of its own, starting from its very definition.

To begin with, social responsibility is a relatively novel concept in modern economic literature. Lawyers and philosophers of law (from Hobbs and Hume to Kant and Hegel) tried to define the different concepts of responsibility and accountability, while economists focused more on the profit-maximizing (self-maximizing) behaviour of a rational economic agent. However, with the development of (corporate) social responsibility concepts (including social accounting) and (corporate) social performance models, organizations have become obliged to meet a set of social criteria. For instance, an entrepreneur who maximizes his or her well-being by behaving as a law-abiding member of society contributes to that society (through taxes, employment, and so forth). Consequently, business organizations have to maximize profit in meeting their economic responsibilities, obey the law in achieving legal responsibilities, act within the prevailing industry and societal norms in pursuing ethical responsibilities, and use their discretion to promote society's welfare in various ways, performing so-called discretionary responsibilities (Carroll, 1979; 1991; Aupperle, Carroll and Hatfield, 1985; Smith and Blackburn, 1988, etc.). Carroll (1979; 1991) conceded that these categories are not mutually exclusive and do not give different weightings to various social concerns.

Increasing concerns about social responsibility in the business sector have focused on the public or quasi-public sector. However one defines the central bank, it is an institution of public law in all countries in the world, and part of

the tradition of building a social or social-like economy following the human losses and destruction of the Second World War. The last central bank to be 'converted' into an institution of public law was the South African Reserve Bank (Šević, 1996b). So, if the central bank is really an institution of public law, should it not be subjected to public scrutiny?

As an institution of public law, the central bank can be treated as any institution and has to be subjected to public scrutiny. Also, as an institution of public law, especially if it is in the regime of an independent agency, it has to be subject to the principles of good governance, which entail – in addition to a well-defined accountability – a regard for the public and even responsibility for the welfare of individuals affected by its actions. Being a part of the state organism it has to align with the rest of the professional public services, while remaining independent in discharging its functions. In claiming a special status for the central bank, many (economists) forgot the issues of accountability and responsibility, which is reversal of rights. Maybe the modern central banks command an independent status but modern public services require a clear allocation of responsibilities and clearly defined roles. The central bank as a public entity has its promulgated mission and should focus on policy outcomes, supporting sustainable competition and being ultimately responsible for its actions. However, it seems that many, blinded by the need for independence, simply ignored the very fact of the status of the central bank: that is, as an entity of public law, it is ultimately accountable and socially responsible.

It seems that there is room for further research in this direction, focusing on the modern legal status of the central bank (as opposed to the long-lasting economists' call for (institutional) independence), especially within the modern framework of delegation of policy making to non-majoritarian institutions; its ultimate accountability to a democratically elected body (which it may be itself); and finally its inherent social responsibility as a body in the public regime, ostensibly serving the country and, in a democracy, the vast majority of the population. However, it seems that it will be some time before this line of research is pursued, as the theory is still dominated by those (mainly the neo-classical scholars) calling for an ever-increasing independence of the central bank (as if none has been achieved up to now), although recently even they admit that the link between independence and long-term economic growth (preferably sustainable) is not clear. But, it appears that this will come into the focus simply because accountability and social responsibility find their very roots in the basic principles of natural law. And, lawyers usually resort to this when other approaches fail to deliver.

Conclusion

The 1990s revived the old dilemma in the theory of what is the better, central or free banking. In our view this is a quite false, non-existent dilemma. The

central bank has proven to be an indispensable social institution, although some controversies on its status, duties and social importance remain. Central banking had its natural development. Nobody could have foreseen the range of central bank's duties when the first government-sponsored (privileged) banks were incorporated (Goodhart 1985; 1988). German banking theory probably depicted most accurately the central bank's development, firstly speaking about issuing banks, later central issuing banks, and finally, in the 1930s only about central banks. Although there is a similarity across central banks, it is impossible to find two equally regulated ones. As the result of traditional, cultural and other differences, there are huge variations in the institutional framework for different central banks. However, the late 1980s and 1990s brought new initiatives in central banking theory. At present the 'state-of-the-art' assumes central bank independence. This is usually seen as the ability of central banks to be fully separated from the government and its direct and/or indirect influence. Some empirical (econometric) studies conducted in the 1980s and early 1990s suggested that independence delivers higher monetary stability, but recent ones conducted in the very late 1990s failed to find a direct causal link between central bank independence and low inflation. Intuitively, it was assumed that *de jure* independence has to deliver, but it seems that there is a need for overall social 'institutional capacity building', as the central bank delivers in those countries where the entire institutional setting is rule-dominated.

The central bank is of interest for law and economics as the traditionally regulatory body, as well. From the end of the nineteenth century, the central bank has been in charge of supervising the banking system, as well as supporting individual banks in a liquidity trap. Banking regulation and its efficiency is closely related to quality of the institutional framework for the central bank. Banking crises and runs on banks can endanger the whole economy, and this is the reason for the central bank to resume the micro function of lender of last resort. In recent times the question of whether a compulsory deposit protection scheme should be organized has also been raised. The opponents of central banking usually argue that this can increase moral hazard. This could be true, but from which situation will society benefit more? Apart from the theoretical considerations, deposit protection is the current reality. However, the central bank itself also fails victim of 'political delegation' which strengthens its independence, as regulatory functions are increasingly taken away from the central bank and given to an independent regulatory body (as in the UK with the creation of the Financial Service Authority – FSA). The central bank will increasingly be focused exclusively on monetary policy and, as the trend shows, on financial system stability.

What are the perspectives for future research on the central bank? Probably, analysis of the central banks' efficiency from the property rights theory

perspective. Previously, this has been done for the individual banking system (Davies, 1981). Also, researchers on central bank independence need a comprehensive law and economics approach to facilitate research on the institutional framework and its efficiency, especially focusing on central bank accountability and social responsibility. These two latter areas could be a completely new avenue of research. Nevertheless, the time for research on the central bank in this branch of economics (and legal theory) is yet to come.

References

Alesina, A. and L.H. Summers (1993), 'Central bank independence and macroeconomic performance: some comparative evidence', *Journal of Money, Credit and Banking*, May, 152–62.

Alesina, A. and G. Tabelini (1987), 'Rules and discretion with non-coordinated monetary and fiscal policies', *Economic Inquiry*, **25**, 619–30.

Aupperle, K.E., A.B. Carroll and J.D. Hatfield (1985), 'An empirical examination of the relationship between corporate social responsibility and profitability', *Academy of Management Journal*, **28**, 446–64.

Berman, S. and K.R. McNamara (1999), 'Banks on democracy why central banks need public oversight', *Foreign Affairs*, **78**(2), 2–8.

Blinder, A. (1999), *Central Banking in Theory and Practice*, Cambridge: The MIT Press.

Boyer-Fraise, F. (1903), *La Banque de France et ses principales opérations (escompte, dépôts, avances)*, [The Bank of France and Its Principal Operations [Discount, Deposits, Advances)], Paris: Henri Jouvé.

Canzoneri, M.B. and P.R. Masson (1992), *Establishing a Central Bank: Issues in Europe and Lessons from the US*, Cambridge: Cambridge University Press.

Carroll, A.B. (1979), 'A three-dimensional conceptual model of corporate social performance', *Academy of Management Review*, **5**, 497–505.

Carroll, A.B. (1991), 'CSP measurement: a commentary on methods for evaluation of an illusive construct', in J. E. Prost (ed.), *Research in Corporate Social Performance and Policy*, **12**, 385–401.

Cukierman, A. (1991), 'Discretion, precommitments and the prospects for a European central bank – fixed parities versus a commonly managed currency', in Z. Eckstein (ed.), *Aspects of Central Bank Policy Making*, Berlin: Springer Verlag, pp. 147–98.

Cukierman, A. (1992), *Central Bank Strategy, Credibility, and Independence: Theory and Evidence*, Cambridge, MA: The MIT Press.

Cukierman, A. (1994), 'Central bank independence and monetary control', *Economic Journal*, **104**, 1437–48.

Davies, G.D. (1981), 'Property rights and economic behavior in private and government enterprises: the case of Australia's banking system', *Research in Law and Economics*, **3**, 111–42.

De Haan, J. (1997), 'The European Central Bank: independence, accountability and strategy: a review', *Public Choice*, **93**, 395–426.

Deckert, H. (1926), *Die Notendeckungsvorschriften der wichtigsten Zentralnotenbanken* [The Rules of Money Coverage in Most Important Central Banks], Berlin: Julius Springer.

Faust, J. and L.E.O. Svensson (2001), 'Transparency and credibility: monetary policy with unobservable goals', *International Economic Review*, **42**(2), 369–97.

Forder, J. (1996), 'On the assessment and implementation of "institutional" remedies', *Oxford Economic Papers*, **48**, 39–41.

Forder, J. (2001), 'Some methodical issues in statutory characterisation of central banks', *West European Politics*, **24**(1), 202–16.

Hall, P. and R. Franzese (1998), 'Mixed signals: Central bank independence, coordinated wage-bargaining, and European Monetary Union', *International Organization*, **52**(3), 505–36.

Federal Reserve Bank of San Francisco (1987), *Monetary Policy in the United States*, San Francisco: Federal Reserve Bank of San Francisco.

Friedman, M. (1959), *A Program for Monetary Stability*, New York: Fordham University.

Goodhart, C.A.E. (1985), *The Evolution of Central Banks: A Natural Development*, London: London School of Economics and Political Sciences Santory-Toyota International Centre for Economics and Related Disciplines.
Goodhart, C.A.E. (1988), *The Evolution of Central Banks*, Cambridge, MA: The MIT Press.
Goodhart, C.A.E. (1994), 'What should central banks do? What should be their macroeconomic objectives and operations', *Economic Journal*, **104**, 1424–36.
Goodhart, C.A.E. (1995), *The Central Bank and the Financial System*, Basingstoke: Macmillan.
Goodman, J. (1992), *Monetary Sovereignty*, Ithaca, NY: The Cornell University Press.
Helander, S. (1916), *Theorie und Poltik der Zentralnotenbanken in Ihrer Entwicklung* [Theory and Policy of Central Banks in their Course of Development], Jena: G. Fischer.
Hilferding, R. (1920), *Das Finanzkapital: eine Studie über due Jüngste Entwicklung des Kapitalismus*, [*Finance Capital: A Study of the Latest Phase of Capitalist Development*], Wien: Weiner Voksbuchhandlung, London: Routledge & Paul Kegan, reprinted 1981.
Holtfrerich in Bundesbank, ed. (1999), *Fifty Years of the Deutsche Mark*, Oxford: Oxford University Press.
Kareken, J.H. and N. Wallace (1978), 'Deposit insurance and bank regulation: a political equilibrium exposition', *Journal of Business*, **51**(3).
Kent, P. (1966), *Money and Banking*, New York: Holt Reinhard and Winston.
Klein, B. (1974), 'The competitive supply of money', *Journal of Money, Credit and Banking*, **6**(4), 423–53.
Kydland, F. and E. Prescott (1977), 'Rules rather than discretion: The inconsistency of optimal plans', *Journal of Political Economy*, **85**, 473–91.
La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R.W. Vishny (1996), 'Law and finance', NBER Working Paper no. 5661, Cambridge, MA: National Bureau of Economic Research.
La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R.W. Vishny (1997), 'Legal determinants of external finance', NBER Working Paper no. 5879, Cambridge, MA: National Bureau of Economic Research.
Maxfield, S. (1997), *Gatekeepers of Growth: The International Political Economy of Central Banking in Developing Countries*, Princeton: Princeton University Press.
McNamara, K.R. (2002), 'Rational fictions: Central bank independence and the social logic of delegation', *West European Politics*, **25**(1), 47–76.
McNamara, K. and E. Jones (1996), 'The class of institutions: Germany in European monetary affairs', *German Politics and Society*, **14**, 5–31.
North, D.C. (1993), 'Institutions and credible commitment', *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, **149**(1), 11–23.
Rogoff, K. (1985), 'The optimal degree of commitment of an intermediate monetary target', *Quarterly Journal of Economics*, **100**, 1169–90.
Sayers, R.S. (1960), *Modern Banking*, 5th edn, Oxford: Clarendon Press.
Selgin, G. (1988), 'Accommodating changes in the relative demand for currency: Free banking vs. central banking', *Cato Journal*, **7**(3), 621–41.
Selgin, G. (1994), 'Free banking and monetary control', *Economic Journal*, **104**, 1449–59.
Selgin, G. and L.H. White (1987), 'The evolution of a free banking system', *Economic Inquiry*, **25**, July, 439–57.
Šević, Ž. (1995), 'Concept of the financial system: An institutional financial economics approach', mimeo, Dundee: University of Dundee Department of Economics and Management.
Šević, Ž. (1996a), 'Centralna monetarna vlast u Srbiji, odnosno Jugoslaviji: Istorijska perspektiva' [Central Monetary Power in Serbia, i.e. Yugoslavia: An Historical Perspective], *Jugoslovensko bankarstvo*, **51**(5–6), 459–80.
Šević, Ž. (1996b), *Centralna banka: Položaj – organizacija – funkcije*, [Central Bank: Position – Organisation – Functions], Belgrade: Čigoja-štampa.
Šević, Ž. (1996c), *Financial Reform in a Transitional Economy: Some Conceptual Issues*, Tokyo and Belgrade: SYLFF-YASF.
Šević, Ž. (1996d), 'Odnos centralne banke i izvršne vlasti' ['The relationship between the central bank and the executive power'], *Jugoslovensko bankarstvo*, **26**(1–2), 29–36.
Šević, Ž. (1997a), 'The notion of "money" in the Soviet economic thought and its implication

in practice prior to 1924', mimeo, Dundee: University of Dundee Department of Economic Studies.
Šević, Ž. (1997b), 'Revenue sources' in J. Nemec and G. Wright (eds.), *Public Finance*, Bratislava: NISPAcee, pp. 250–94.
Šević, Ž. (1999), *Restructuring Banks in Central and Eastern European Countries as a Part of Macroeconomic Changes towards Market Oriented Economies*, Belgrade: Balkan Center for Public Policy and Related Studies and Čigoja-štampa.
Šević, Ž. and A. Šević (1998), 'Monetary policy in a transitional economy: some open questions and experience', *Economic Systems*, **22**(3), 299–304.
Smith W. and R. Blackburn (1988), 'Corporate social responsibility: a psychometric examination of a management instrument', *Proceedings of the Southern Management Association*, 293–5.
Spong, K. (1990), *Banking Regulation: Its Purposes, Implementation and Effects*, 3rd edn, Kansas City, MO: Federal Reserve Bank of Kansas City Division of Bank Supervision and Structure.
Steiner, F. (1924), *Notenbankpolitik und staatlich Anleihepolitik in den Österreich-ungarischen Nachfolgestaaten* [The Policy of Emission Banks and State Loan Policy in Newly Formed Countries of the Former Austro-Hungarian Empire], München and Leipzig: Dunker & Humbolt.
Tietmeyer, H. (1994), *Role and Instruments of Monetary Policy*, Kiel Vorträge Nr. 124, Kiel: Institut für Weltwirtschaft an der Universität Kiel.
Veit, O. (1969), *Grundriss der Währungspolitik* [The Outline of Money Policy], Frankfurt am Main: Knapp.
Volcker, P. A. (1990), 'The role of central banks', in Federal Reserve Bank of Kansas City (ed.), *Central Banking Issues in Emerging Market-Oriented Economies*, Kansas City: Federal Reserve Bank of Kansas City, pp. 1–8.
Walcher, K. (1876), *Die Notenbanken und die Währungsfrage* [Emission Banks and a Money Question], Berlin: Haude and Spener.
White, L.H. (1984), *Free Banking in Britain: Theory Experience and Debate, 1800–1845*, Cambridge: Cambridge University Press.
White, L.H. (1995), *Free Banking in Britain: Theory, Experience and Debate, 1800–1845*, London: The Institute of Economic Affairs.
Wood, G.E., T.C. Mills, and F.H. Capie (1993), *Central Bank Independence: What Is It and What Will It Do For Us?*, London: Institute of Economic Affairs.

12 Constitutional economics I

Francesco Farina

Introduction

While constitutional economics has to be traced back to the studies on politics by the Greek philosophers – first of all, Plato and Aristotle – the economic analysis of the constitutional themes has flourished since the seventeenth century. Thomas Hobbes argued that in the state of nature, individuals are compelled to agree on relinquishing their individual rights to a sovereign, so as to avoid the worst collective outcome of mutual defection from cooperation, due to the individual rationality of a free-riding strategy. Many political philosophers and economists – among them John Locke (1740 [1988]) and Adam Smith (1759 [1976]; 1776 [1976]) – contended his view by claiming that the state of nature is not always a war of all against all. Cooperation may arise inside a society without being imposed by an absolute power. Inside this line of thought, the names of Immanuel Kant and David Hume can be associated with the notions of empathy and reciprocity, respectively. In Kant, trust is founded on the coincidence of morality (the right of every individual to be treated fairly by everyone else) and rationality (the individual is rational when he/she impartially applies the universal law of fair treatment). In Hume, trust is founded solely on self-interested rationality, which drives individuals to find their advantage in reciprocity as a stable behaviour (Hume, 1740 [1978]). Especially in repeated social interactions, rational individuals discover that their goals are more efficiently pursued by a coordination strategy based either on the imposition of a moral constraint to their self-interested behaviour, or on the expectation that their self-interested behaviour endogenously changes through a learning process, whereby the ensuing reciprocal concessions are conducive to a cooperative equilibrium.

The analysis of institutions upon which individuals and communities rely for their interactions is founded on two pillars: (i) following methodological individualism, the individual behaviour is the basic unity of analysis; and (ii) 'the way the game is played' represents the collective pre-requisite for the players' strategies in any social interaction, so that social and market equilibria have a collective character (Arrow, 1994).

In being involved in social and market interactions, individuals face uncertainty due to the exogenous conditions given by the state of nature and to the possibility of opportunistic behaviour. A game-theoretic framework has therefore become frequent in social sciences. The pay-off matrices of social

interactions may take different forms – ranging from the well-known prisoner's dilemma (PD), to the 'hawk and dove', to the assurance games and so on – depending on the individuals interacting in society to respond to a variety of strategies, such as nastiness, free-riding, reciprocity and empathy. A 'self-enforcing' equilibrium outcome can be formalized in coordination games with conflict of interests, stemming from the individuals' consensus on the economic and social advantages of regularity in social behaviour (Schotter, 1981). Whenever individuals are able to create an institution, whereby their mutual advantage overcomes the opportunistic behaviour implied by self-interested rationality, their social interactions avoid coordination failures. In a self-enforcing equilibrium, the incentive to cooperate is endogenous to the game itself, and there is no need for an external *deus ex machina*. It is also worth stressing that many social interactions are characterized by multiple equilibria (Aoki, 2002), so that each equilibrium is associated with one rule of the game, in combination with each individual's strategy.

Institutions consist of *formal* (constitutions, laws and regulations), and *informal* (contracts, social norms and customs) rules of interaction. Self-enforcing equilibrium outcomes depend on the interplay between formal and informal institutions. The status of social equilibria is acquired by formal and informal rules of interaction through the explicit or implicit enforcement of a sanction, respectively (Marx, 1844 [1970]; North, 1990, 1991). To start with informal institutions, a well-known example is the 'social capital', a name for shared values, such as trust. Social capital facilitates individuals in their market interactions, as they drastically reduce uncertainty on the future spot market exchanges and improve the efficiency of long-term contractual relationships. A widespread presence of social capital inside the society magnifies the good functioning of formal institutions to a great extent.

Informal institutions arise in the markets to foster cooperative agreements by counteracting uncertainty and opportunistic behaviour. The theory of 'contractual incompleteness' claims that contracts are imperfect because their clauses do not comprehend all future possible contingencies; moreover, courts are not usually in a position to verify their implementation. The Coasean view (Coase, 1961) of social and market interactions maintains that the decentralized organization of exchange relationships may be inefficient because individuals are often burdened by high transaction (mainly, bargaining) costs. By reducing uncertainty on future spot market exchanges, informal rules sustain the adoption of cooperative strategies by individuals once the relevant information comes true. Otherwise, formal institutions have to intervene and organize exchanges by a centralized command.

In pre-modern Europe, intra-community agencies and enforcement mechanisms were decisive in building up and preserving the medieval merchants' reputation along with the development of impersonal exchanges (Greif, 2001).

Commercial law was born in the codes of conduct of the guild merchants and then improved by the *lex mercatoria*, the regulatory institution which ruled on commercial transactions in medieval Europe, protecting the security and the certainty of exchanges and preserving the markets' functioning from abuse of dominant positions (Milgrom et al., 1990).

Important factors determining the outcome of social interactions are *complementarity* (the equilibrium outcome in one game depends on the way in which the equilibrium outcome is taking shape in another game) and *path dependence* (the choice made at a certain fork of an extensive-form game heavily impinges on the final outcome). These features of the dynamics of social interactions mould the expectations of the individuals and drive them to select a certain equilibrium (Greif, 1994). The production of informal institutional rules also comes through the cultural beliefs, which are produced by individual expectations and rule on the selection of individual strategies for the subsequent social games in which individuals will be involved. In this perspective, markets and informal institutions are complementary in governing transactions. Overall, the framework of informal institutions in which market exchanges take place may be characterized by tight complementarities across social games and by path dependence from social norms and cultural beliefs of the community.

By using his socio-antropologic method, Polanyi has shown that the market is an artificial construction: it originates within the constraints created by the organization of the state previously agreed on by the society (Polanyi, 1957). The theory of 'market failures' explains the emergence in the last century of the 'mixed' economy, a structure composed by private firms operating in the markets under regulations provided by public agencies, and public institutions devoted to allocative, stabilization and redistribution policies. A huge amount of research work has been produced in the last decades on the size and the functioning of the public sector institutions in the advanced countries. The more social capital is rooted in a society, the better formal institutions function. As a matter of fact, social capital makes the rules sustaining the cooperative outcome endogenously respected. Due to the spontaneous commitment to laws and norms by the community at large, many suboptimal outcomes of the social games fade out (Coleman, 1990; Putnam,1993). In societies characterized by more collectivistic values, the legal and public institutions are less decisive than the informal ones. In many Asian countries, such as Japan (Morishima, 1982) and China (Weitzman and Xu, 1993), social capital mainly consists of the reputational effects associated with the respect of social norms.

Since informal and formal institutions happen to be closely interwoven, a flawed connection among them gives birth to inefficient rules of the game. Along with market failures, the political economy literature has analysed a

series of government failures. In fact, government relations between voters (as principals) and politicians (as agents), as well as the behaviour of elected representatives, are often plagued with conflicts of interest and opportunistic behaviour, which also affect the functioning of the market economy. The question is to what extent and through which instruments constitutional rules are capable of constraining human behaviour both in the market and in the government. Let us then turn to formal institutions.

The liberal appraisal of constitutional rules

Neoclassical economics considers social welfare as the mere aggregation of individual choices. The liberal view is that the evaluation of social welfare regards only rules and procedures, as outcomes cannot be compared and ranked. In the liberal tradition, the two main views on the constitutional organization of modern democracies were put forward by Friedrich von Hayek and James Buchanan, who trace back the origin of constitutions to spontaneous evolution and intentional construction, respectively.

The Hayek view

According to the Coase theorem, the parties could succeed in devising and implementing all the contract clauses, including the optimal definition and assignment of legal rights by voluntary actions. Yet, positive transaction costs may cause the bargaining performed by individuals in their social interactions to result in an inefficient outcome. The main source of transaction costs is strategic behaviour, which creates reciprocal externalities. Individuals interacting in the markets will then be unable to reach a Pareto-efficient equilibrium on the contract curve of the general equilibrium analysis. The Chicago school argues that the legal production may be instrumental in overcoming contractual failures, as transaction costs can be reduced by imposing the correct allocation of property rights. The complementary role played by institutions such as the judiciary and the legislation – which drive individuals to internalize the externalities and realize the 'core' of the efficient equilibrium exchanges – helps the free market to be the fundamental institution of every economic system.

The liberal vision of the society and the economy affirmed by the Chicago school is also endorsed by the constitutional theory put forward by Hayek. The Austrian economist and philosopher thinks that the evolutionary process of selection rules consists not only in the informal institutions created by market forces, but also in the constitutional architecture emerging from the common law produced by judges. In the atomistic market organization, the dispersed actions of individuals reaching unknown ends, guided by signals represented by market prices, make the aggregate outcome at the same time unintentional and efficient. Similarly, a competition process between juris-

dictions for the best performance in terms of social welfare amounts to the selection of their preferred rules and procedures according to the method of evolutionary learning (Vanberg, 1994). Any existent equilibrium is efficient by definition, because it derives from the evolutionary nature of the spontaneous market order framed inside the traditional set of rules and customs. However, differently from the evolutionary nature of the spontaneous order of the market resulting in the equilibrium prices, the intentional institutional 'competition as a discovery procedure' – that is, the competition for the best performance among different institutional structures – lacks the necessary conditions for comparison: first of all, complete information (Hayek, 1979). Hence, constitutional and, more generally, public law 'merely organises the apparatus for the better functioning of the more comprehensive spontaneous order' (Hayek, 1973; 1978: 79).

Following his conception of society as a spontaneous order, Hayek conceives institutions as 'the result of human action but not of human design'. He observes that sometimes constitutions no longer represent the coercion needed to limit the liberties of the individuals, but implement general principles of 'social justice' which might jeopardize the efficient outcome produced by the spontaneous order of the market. The tendency of the legislative bodies to deal with both the promulgation of laws devoted to regulate specific matters of interest to particular groups, and the general legislation concerning universal norms of behaviour, has led to excessive 'democratisation'. Therefore, governments should not manipulate the overall income distribution resulting from individuals' decentralized decisions (Hayek, 1976). Constitutional law should recognize that uncertainty and risk may negatively affect the efforts of the individuals operating in the market, so that a threshold in terms of 'minimum income' has to be provided against 'bad luck'. Yet, the dividing line must be precisely drawn between this universal 'safety net' as a general principle which is valid for everyone, and the pernicious legitimization of any general principle of redistributive justice which any interest group in society could advocate. In general, 'democratisation' should be limited, as the public intervention into the economy must only devise rules and procedures to cope with financing public goods, and their provision has to be organized by private companies. In the following, 'public goods' is a blanket name, including merit goods and social protection – such as education, unemployment and poverty subsides, health care, pensions – whose supply and/or demand suffer from moral hazard and adverse selection, or produce externalities.

The Buchanan view

The constructivistic approach endorsed by Buchanan alternatively views the constitution as the conscious effort by the individuals of a community to engage in the formalization of a 'social contract'. A constitution is that

particular kind of formal institution whereby the state is legitimated by the fundamental principles a community has unanimously agreed on. Individuals have to agree on those values and principles setting the constraints within which they could act and markets operate. The constitution recognizes the property rights to be negotiated and enforced, so that uncertainty is reduced and economic incentives strengthened. Economies of scale make the state – as the organization of the cooperative games between the government and the citizens – superior to private associations in the definition and protection of property rights.

In competitive markets for private goods, whatever the number of individuals, the bargaining remains bilateral and the 'invisible hand' is capable of setting the optimal price. In the political markets, which concern the production of public goods and the cooperative solution to PD situations, the more the number of participants in the social community increases, the more the 'political exchange' among individuals is burdened by opportunistic behaviour (Buchanan, 1975). The constructivistic approach aims at fighting the ubiquitous interpersonal conflicts of interest characterizing non-market interactions (for example, public goods provision). In a vision that founds the existence of society on the adhesion to the social contract by each individual, any constitutional rule must comply with everyone's interests. Since the social contract must be justified to all citizens, a constitutional rule should be approved by unanimity rule (Wicksell, 1896). The strategy towards limiting the opportunistic behaviour which jeopardizes social interactions relies on the distinction between constitutional and post-constitutional choices.

The conception of the state underlying the 'market incompleteness' approach assigns the constitution the task to cope with the impossibility of foreseeing all future contingent states of the world. A community agree on a constitution as a set of rules – both for the private markets and the political relations – to which they resort once relevant information is revealed. Similarly, constitutional rules should aim at constraining individual behaviour in interactions taking place in the post-constitutional stage. Individuals are sufficiently informed to be able to evaluate costs and benefits of alternative institutional organizations. Yet, they are constrained by a veil of uncertainty in assessing which constitutional rules will most aptly serve their own post-constitutional interests. Since the redistributive consequences of constitutional choices concerning political and economic questions will unveil only in the long run, it is individually rational to agree on a certain set of constitutional rules and procedures. For this reason, a unanimous agreement among citizens can be expected (Buchanan and Tullock, 1962).

Buchanan traces back the many kinds of impasse from which in the real world social contracts suffer to the 'non-market failures' of modern democracies. Although constitutional rules set the framework for the political process

enacted by a government, the devolution of individual rights to a political authority may result in the absolute command of a Leviathan. Individuals should point to the cooperative Pareto-optimal outcome yielding net benefits to all by sharing the tax costs towards the financing of public good provision. Instead, although benefits for all parties are potentially available in the PD game over the state provision of public goods, each group points to the free-riding pay-off. The main question about majority decision making is that social games among parties are plagued by rent-seeking behaviour. An example is the practice of 'log-rolling'. By reciprocating the endorsement of laws aimed at boosting their respective self-interest, political representatives end up approving laws which may undermine social welfare. More importantly, due to the rotation among majorities, any governing coalition of parties is in a position to pass by majority voting laws protecting its own interests, and burdening the other groups with the costs. By means of simple PD game examples, Buchanan shows that the succession of redistributive processes generates a 'present value' pay-off corresponding to the status quo of no collective action. In the long run, the legislation as a repeated game ends up with the worst aggregate pay-offs on average. Collective-action mechanisms are needed for individuals to protect themselves from 'the tyranny of the majority'. Unless a constitutional constraint prevents differential treatment (so that constitutional limits are posited on the off-diagonal pay-offs of the PD game matrix), politics under majority rule is bound to become increasingly distributional.

Buchanan is a subjectivist-contractarian who conceives the 'optimality' condition in the perfect competition market not as the consequence of objective production and exchange conditions, but as maximization criteria stemming from the 'consensus' among the participants (Buchanan, 1991a). He refuses the allegation of a unique efficient allocation of resources claimed by standard welfare economics. Also, market exchanges are ruled by strategic interaction, so that there is no objective criterion to determine the efficiency of a market outcome. Since this distributive compromise makes a given set of voluntary market exchanges a Pareto optimum, efficiency and distribution are the joint outcome of bargaining on gains from trade in the market interaction. On the basis of the existing institutional organization of property rights, each party bargains with another party to convince it to accept compensation in exchange for gains for a trade favouring the first party (Buchanan, 1991b).

Therefore, Buchanan thinks that self-interested individuals can reach cooperative agreements in their non-market relations (for instance, public goods provision), but this does not imply that a Pareto-optimal equilibrium represents the fulfilment of a collective optimum. According to the Buchanan contractualist view, there is no 'common good' to which a government could finalize the maximization of a social welfare function (Buchanan, 1991a).

Many principles of justice typically make some individuals better off and others worse off, thus falling short of satisfying the criterion of unanimity voting on constitutional rules by rational and self-interested individuals. Buchanan endorsed the implementation of horizontal equity, which in principle does not rely on a conception of justice, as everyone's right to equal treatment. The liberal approach to state intervention bans redistributive policies that could not be justified to all individuals participating in the social contract. However, redistribution as a byproduct might be legitimate. Provided that a flat tax rate is imposed upon a generality rule, and is accompanied by a set of demogrant (equal per head) transfers equally available to all, no allegation of infringement – either of individual liberties, or of the principle of 'negative freedom' – can be invoked (Brennan and Buchanan, 1985; Buchanan, 1993). Since these instruments of fiscal policy produce net gains for the poor and net losses for the rich, the resulting redistribution occurs on an 'equal treatment' basis.

The constraints imposed by constitutional rules on social interactions – first of all, the existing organization of property rights and the threat power of enforcement – are instrumental for the stability of social and market equilibria. However, any constitutional agreement is sustainable until individual revenues equal social revenues. Whenever the balance between benefits (and positive externalities) and costs in terms of taxes (and negative externalities) turns out to be negative, the constitutional rules are no longer unanimously agreed on and a reform is required. By comparing alternative institutional organizations, the individuals unanimously agree on substituting the existing set of rules, as the previous market outcome is considered suboptimal. The allocation of property rights resulting from new rules allows either the exploitation of some reciprocal interdependencies among individuals or the avoidance of some wasteful investment of resources in predation and defence (Buchanan, 1991b). The unanimous consensus accompanying market relations of production and exchange corresponding to the new property rights organization implies a new Pareto-optimal equilibrium (Buchanan, 1985). Therefore, the objective of any constitutional reform should consist in the design of new governmental institutions to provide citizens with a more efficient system of incentives, on which economic growth, technical change and the expansion of social welfare all depend.

Individual and social rights

A constitution is the most prominent formal institution, as it is the *grundnorm* which determines the set of rules governing political institutions and market interactions. The contractarian approach to constitutions builds upon the ideal covenant among individuals, who recognize their self-interest in entering a political institution fostering social cooperation. The first aim of any

constitution is to stop the insurgence of the Hobbesian state of nature, by allocating and enforcing individuals' property rights, and facilitate social and market exchanges (Backhaus, 1996).

The conception of liberty and rights which a constitution should put forward is one of the most critical and controversial questions of constitutional theory. The view that individuals have moral rights against any attempts by the government to impose moral values on them is also shared by Dworkin's view on liberalism. The state should not endorse any substantive ethical theory, as the realization of moral values is a prerogative of individuals only, and the judiciary must be ethically neutral to be compatible with every individual's equal right to individual liberties. However, Dworkin more ambitiously argues that in as much as income inequality is due to the ethically non-neutral distribution of natural talents and abilities across individuals, constitutions should allow for progressive taxation aimed at eliminating undeserved conditions of 'disadvantage' (Dworkin, 1978). In the same vein, Rawls advocates the equality of primary goods: the surplus from social cooperation should be devoted to improving the well-being of the worst off (Rawls, 1971) and international law should impose redistribution worldwide to improve the chances of progress of the most deprived peoples (Rawls, 1999).

Habermas (1988) points to a social dimension of individual identity by blurring the division between individual and social rights. While sharing the view that moral values are founded in the individual, he claims that rights establish a tight linkage between each individual and society: if one accepts the precondition that rights are rooted in the cultural underpinnings of society and are essential to define an individual, then the individuals' identity is moulded according to the values of the society. Habermas views citizenship as everyone's right to well-being, whereby the eligibility to benefit from equity-concerned public policies should not be conditional on subjective factors, such as the place of birth or the availability of employment. Individuals should be aware that everybody's equal worth depends on warranting to each individual an equal opportunity to a decent life. Since well-being is linked to access to health care, education, environmental protection and so on, state intervention is required to organize a system of public and merit goods (Marshall, 1981).

We are then faced with alternative conceptions of the individual which constitutional rules should address. On the one hand, the economic liberalism advocated by Hayek and Buchanan criticizes excessive state intervention in the sphere of autonomy of the individual, such as the equalizing policies pursued by fiscal authorities which diminish the propensity to risk, and result in a fall of the incentive to invest and to work. On the other hand, a variety of principles of justice – ranging from equality of opportunity to equality of

resources – interpret democracy as the political organization in which the individual is endowed with the rights defined by the cultural underpinnings of a society. The notion of equal right to equal liberty then establishes a link between the system of rights and democracy, whereby public policies need to cope with multidimensional inequality due to differential opportunities across individuals.

These two different approaches to a liberal and democratic society are reflected by the two principles of 'negative freedom' and 'positive freedom', respectively (de Ruggiero, 1925 [1927]; Berlin, 1958). The first principle entitles every individual to the right not to be exploited by others and by the government. The liberal appraisal of methodological individualism maintains that any voluntary participation in a 'social contract' entitles one to the right to not be exploited by a democratically elected government. Whenever some kind of market failure makes state intervention necessary, the government must allocate resources to public goods without affecting the distribution of resources across individuals. The aim of the constitution should consist in enabling individuals to fully dispose of the fruits of their own talents and abilities. The voting decision mechanism represents the democratic method by which – just as the consumer can exit from a seller or a contractual relationship – the citizen is empowered with the right to change the government by joining another party (or coalition of parties) in the polls.

The second principle entitles every individual to the right to enjoy all material and moral conditions for self-realization. The democratic and socialist lines of thought, which have in different ways developed the Marxian view of society and the economy, argue that a constitution should indicate which institutional setting is most capable of coping with the problem of inequality 'at the starting gate', which is typical of modern capitalist society. The multiple dimensions of inequality (family environment, health, education, one's community social capital and so on) could prevent equal individuals from being equally free, in the sense of the positive freedom to have access to an equal chance of a good life. A conception of rights based on 'negative freedom' interprets a right as the prohibition of interference with an individual's freedom. A conception of rights based on 'positive freedom' instead asks society to create the appropriate conditions for the right to be exerted. Some constitutions explicitly call them 'social rights'.

Due to scarcity of resources, and despite the tumultuous growth experienced in the last centuries, modern capitalist societies are still constrained by the necessity to choose among alternative objectives. Suppose that the right to have the state not interfere in the matter of inheritance has to be ranked against the right to study whose funding has to rely on an inheritance tax. The choice between the right one wishes to defend may depend on which of the two principles of negative or positive freedom is considered more important:

adherence to the former favours the right to have the state not interfere in the matter of inheritance; while adherence to the latter favours the right to study.

The need for 'active' policies to affirm and guarantee the exercise of 'positive' rights is justified by the existence of wide income inequalities causing large disparities across individuals regarding their chance of a good life. An important aspect of this problem is the tendency of inequality to reproduce itself across generations. A possible economic cause of the intergenerational path dependence of life conditions is the power asymmetry among individuals, all equal in the matter of democratic voting but often unequal in the marketplace. This hiatus between the market and the democratic institutions – whereby the market responds to the principle of *suffrage censitaire* and the democratic institutions respond to the principle of *suffrage universel* – establishes the legitimacy of democratic institutions to have a say in the regulation of market exchanges (Fitoussi, 2002). The constitution should aim to put individuals in a position whereby they have equal opportunities to exploit their own talents and abilities, which requires appropriate institutions to foster social cooperation. However, the recognition that the individual is entitled to rights – and what these rights consist of – is only spelled out in constitutions. The theoretical debate is still far from a thorough investigation into the question of devising institutions that foster equal opportunities and make viable the realization of individual rights.

The rules of the game

This section will examine the main rules of the game which constitutions make explicit: the horizontal and vertical separation of powers; voting mechanisms for constitutional rules, election rules, and forms of government.

The horizontal separation of powers

The division of powers among constitutional bodies can be organized both at the horizontal and the vertical levels. The horizontal separation of powers was first conceived by Locke (1640 [1988]) to protect the freedom of citizens by constraining rulers within the limits of the law, and then theorized by Montesquieu (1748 [1955]), as a system of 'checks and balances' whereby authority is horizontally divided among the three constitutional bodies: the executive, the legislative and the judiciary.

James Madison, one of the leading framers of the US Constitution, believed that institutional competition is the key instrument to prevent both the horizontal and the vertical division of powers from being jeopardized by a constitutional body acquiring excessive bargaining power. In the US, the efficiency-enhancing use of competition applies to competition both *between* and *within* constitutional bodies. In order to represent competing interests, horizontally the two legislative branches have different criteria of representa-

tion, and vertically the division between the federal and the state government rules is the self-enforcing mechanism protecting citizens from the central government. Unlike the English tradition of common law, where the judiciary's independence is guaranteed against the lawmakers' interference, the preservation of freedom by checks and balances implies that elected judges are confronted by elected representatives (La Porta et al., 2002). Therefore, winning an elective post entitles one to compete as the representative of a specific interest.

The economic analysis of political institutions maintains that the separation of powers between the executive and the legislative bodies creates the correct incentives, provided that the system of checks and balances respects the following two conditions: (i) a conflict of interest between the bodies must be created by putting the elected representatives in competition in order to impede collusion (for instance, by attributing the agenda-setting power over the size of the budget to one of them, and over its composition to the other); and (ii) policies must be jointly decided and implemented in order to improve their accountability (Persson and Tabellini, 2000). This institutional setting also allows voters to elicit private information about the activities of elected representatives and public officials, thus undermining their efforts to extract rents from asymmetric information (Persson et al., 1997). However, the modern 'incomplete contracts' interpretation underlines the uneasiness stemming from the divide between a constitution which is not a complete contract, and elected representatives delegated to make decisions while not necessarily sharing the same preferences as the social planner who wrote the constitution. Therefore, in all cases not foreseen in the constitution, discretionary choices by elected representatives and/or public officials are to be limited in order to minimize welfare losses. Appropriate and optimally interconnected incentive schemes must be envisaged for the principals of the various bodies (Laffont and Tirole, 1990).

Incomplete information and opportunistic behaviour may also generate the problem of moral hazard by self-interested politicians. The separation of powers has to cope with the moral hazard problem in the economic policy decisions. As for fiscal policy, moral hazard is manifested as the political pressure to expand public expenditures. The fear that the fiscal stance will drift towards a 'soft balance constraint' has inspired specific legislation aimed at limiting the expansionary tendencies of public expenditure by the fiscal authorities, such as the Graham–Rudman law in the US, which impedes further expenditures in the case of failure to meet the limit for the public budget deficit.

Similarly, as confirmed by the empirical evidence of a positive relationship between central bank independence and inflation (Grilli et al., 1991), in many countries the constitution protects the independence of the central bank, so as

to shield monetary authorities from pressure by elected representatives. The objective is to make monetary policy both more effective and more credible, thereby avoiding a situation in which improved output and employment levels would result in a permanently higher inflation rate. The theoretical underpinnings of this tenet consist in the presumption that monetary stability is a public good overriding any other objective assigned to monetary policy. However, a prolonged tight monetary stance may create a hysteresis problem in the unemployment rate. Therefore, a value judgement is always at the root of the macroeconomic governance of the unemployment–inflation tradeoff (Stiglitz, 1998). The conduct of monetary policy is influenced by lobbies in society, where industrialists and trade unions oppose inflation much less than financial institutions do (Posen, 1994). In fact, the tendency to use tight monetary policy as a rule serves the vested interest of financial markets' operators to increase profits by being shielded from uncertainty and price instability.

The independence of the central banker rests on less-objective underpinnings than suggested by mainstream economics. The notion of independence may be at odds with democratic legitimacy. Whenever a constitutional rule protects the central bank's independence, a democratic deficit occurs inside the institutional organization of the state. The lack of democratic legitimacy of the monetary authorities' power may be mitigated by their obligation to accountability. In the European Union (EU), the elected representatives of the European Parliament may comment on the European Central Bank's monetary policy, but they are not empowered with enforcement devices. However, in the US, the Federal Reserve is obliged to present an annual report to Congress, which has the right to express an opinion about the consequences of monetary decisions on the employment level, and could in principle enforce the sanction of passing an amendment changing the central bank's powers and prerogatives. Since this control power might influence the decisional process about the monetary stance, accountability may partially counterbalance independence in the behavioural function of monetary authorities.

The vertical separation of powers

The vertical separation of powers allows the devolution of powers from central government towards lower-level jurisdictions. A multi-level constitutional organization is created by attributing legislation and taxation prerogatives to their elected representatives. According to the 'market-preserving view', decentralization ensures that the central government has the task of policing the market, of ensuring mobility of goods and factors across lower-level jurisdictions, and of complying with a 'hard budget constraint' (Weingast, 1993, 1995; Rodden and Rose-Ackerman, 1997).

In a multi-layer constitutional organization, the central government experiences a tradeoff between policy coordination – the internalization of all the jurisdictions' reciprocal externalities – and accountability, as local communities can observe the welfare level of their jurisdiction. Yet, as verifiability is impossible to implement, the central government cannot be punished for inefficient decisions (Seabright, 1996). To cope with this problem, Inman and Rubenfield (1997) laid down a blueprint for 'democratic federalism' aimed at the middle way between a decentralized government, which conveys advantages in terms of close monitoring on politicians, and a centralized government, which conveys advantages in terms of efficient public goods provision and internalization of spillovers. The constitutional organization as a multi-tier system stems from the interpretation put forward by Brennan and Buchanan (1985) of fiscal decentralization as an instrument to keep to a small size the public sector to avoid burdening the fiscal stance with excessive deficits. The separation between the local fiscal instruments and the centralized power of money creation should guarantee that the hard budget constraint is met (McKinnon and Nechyba, 1997). Yet, by gearing the majority rule to the *log-rolling* method of reciprocation, the representatives of the lower-level jurisdictions may indulge in policies financed by pooling the common taxes but benefiting single jurisdictions. Therefore, responsibilities in the domains with large externalities among lower-level jurisdictions should be assigned to the central government in order to make decentralization inefficient.

Voting mechanisms for constitutional rules

The institution-building process of a federation implies a group of countries voting on a constitutional agreement. If constitutions were complete contracts, unanimity would be the best rule in the approval of constitutional rules, as it is the most demanding and should guarantee *ex post* compliance with obligations undertaken in the agreements. By the adoption of this voting mechanism, each country would credibly manifest its intention of abiding by the contract. In other words, countries would be signalling that any *ex post* deviation is unlikely, as it is perceived to be Pareto suboptimal from the *ex ante* viewpoint. The choice of the unanimity rule would then reflect a successful bargaining resulting in a perfect contract. Yet, the contractual incompleteness view demonstrates that all contracts – constitutions included – are incomplete, as all possible future contingencies cannot be specified and *ex post* verified by a court. Since *ex post* compliance is not credible, there is no point in searching for unanimity in constitutional voting. After its approval, a constitution is always exposed to the possibility of future demands for modifications, a right explicitly acknowledged by a section of many constitutions. In order to increase the probability of unanimity, a clause can be inserted that recognizes the right of all participants to revoke some articles of the agreement after a certain number of years.

One unanimity mechanism, aimed at increasing the chances of unanimous agreement in the light of the shortcomings of the Wicksellian procedure for unanimity voting (first of all, strategic behaviour in preference revelation), is 'veto voting', suggested by Mueller. This mechanism is composed of the following steps: (i) each participant puts forward a proposal (by trying to internalize the other participants' preferences, by anticipating their proposals in order to minimize the probability that its own proposal will be vetoed); (ii) a ranking among the other participants is then set up; and (iii) through a sequence of veto voting, a proposal that nobody has vetoed will eventually be accepted (Mueller, 2003). Provided that the number of participants is sufficiently large to avoid strategic ranking, veto voting has the virtue of forcing convergent proposals, by incorporating the incentive to skip those proposals that are too far removed from the other participants' expected preferences. Other voting mechanisms particularly relevant in institution-building processes are the various majority rules, which are the most used voting mechanisms. These rules range from a simple majority (50 per cent of the votes plus one), to a super-majority (for instance, the so-called 'min–max' rule, which is the closest to the simple majority rule without the Condorcet cycles), to a qualified majority (usually, 66 per cent of the votes).

Election rules and forms of government

A constitution frames the political institutions representing the 'rules of the game' played by voters and political parties. In the literature, two important types of political institutions capable of influencing the economic outcomes have been investigated: election rules and forms of government (Persson and Tabellini, 2003).

With regard to the electoral system, policies can be seen as equilibrium outcomes of delegation games. The agency problem formalized by this game is characterized by the double conflict between the principals and their elected representative and among the principals themselves. Empirical evidence seems to show that majority rule systems guarantee more accountability to the voters, and a more direct link between the representative's performance and his/her probability of being reappointed. As a consequence, the government has a stronger incentive under majority rule, than under a proportional rule system, to perform well – for instance, not to augment public expenditures in the period preceding a general election and so on. Representation is better preserved under proportional than under majority rule, because with more than two parties it is very common to have a single-party government empowered by less than 50 per cent of the voters. In addition to this flimsy democratic legitimacy, the sharp division of the electorate under majority voting makes the electoral rule less conducive to the needed provision of public goods.

Regarding the forms of government, a presidential regime may stimulate more competition among politicians and be more accountable to the voters than a parliamentary regime, where collusion is easier and larger rents are extracted by politicians (Persson and Tabellini, 2000). However, in the presidential regime the formation of vote-by-vote coalitions, caused by the weakness of the party system, puts the legislation in the hands of all sorts of lobbies. As a consequence, similar to majority rule, earmarked taxes finance a small provision of public goods targeted to small communities. The dispersion of power within a government seems to be a significant predictor of public deficits and eventually public debts. Sound public budgets are considered the most important advantage of single-party majority governments, whereas coalition governments present higher public deficit/GDP ratios. According to the 'weak government hypothesis', fragmented governments are less keen to keep the tendency to high public deficits under control than governments with one dominating party (Roubini and Sachs, 1989). The proposed reform against fragmentation consists of super-majority rules, that is, constrained procedures to prevent differential treatment and maximize consensus. Qualified majorities are particularly needed for constitutional changes, in order to prevent government decisions that exacerbate distributional conflicts (for instance, exempting heirs from inheritance taxes). The proportional system, which often comes with the parliamentary regime, generates coalition governments more oriented towards large programmes of social protection aimed at obtaining consensus among larger social groups. The empirical evidence shows that the average tax rate is higher and the distribution of income less unequal in countries where coalition governments have been formed following elections under a proportional representation system, than in countries using a majority system with two parties (Austen-Smith, 2000).

Therefore, there is a relationship between electoral rules and forms of government, in terms of both the tradeoff between accountability and representation, and the amount of social expenditures. The presidential regime more often has a single-party government elected by majority voting, while the parliamentary regime usually has a coalition government elected under the proportionality rule. Majority rule electoral systems and presidential regimes are associated with a variable degree of accountability (depending on the percentage of voters in general elections) and a small welfare state, while proportional rule electoral systems and parliamentary regimes have better representation and maintain a constant role of social expenditures.

Theories of multi-level constitutional organization

The organization of the state falls into two main categories: the unitary state, characterized by a mono-jurisdictional organization, where the devolution to lower-level administrative bodies does not entail the transfer of taxation

powers; and the federal state, characterized by a multi-jurisdictional organization, where lower-level legislative bodies are entitled to levy taxes.

In the literature on federalism, following the *Federalist Papers* (see under Madison, 1787–88), the state is conceived as a multi-layer institution aggregating different levels of government empowered with certain functions and prerogatives, whether on an exclusive or on a shared basis. The breakdown of the governmental functions and prerogatives across jurisdictions occurs in both the horizontal dimension, whereby each equal-level government – such as states, regions, provinces or municipalities – exercises power within the boundaries of a certain territory, and the vertical dimensions, whereby the federal government circumscribes the prerogatives of the lower-level governments. One of the main questions is whether individual utility is maximized when the packages of taxes and public goods are supplied by the central government or by lower-level governments. Economic theory has put forward the following criteria. Under two conditions centralization is relatively more efficient and two conditions leading to the superiority of decentralization: (i) public policies of a jurisdiction have a high number of spillover effects – such as externalities and interdependencies – on the individuals who are residents of the remaining jurisdictions and (ii) public policies enjoy economies of scale. Two other conditions instead favor a preference for decentralization: (iii) high heterogeneity of individuals' preferences and (iv) more and better information is available for the lower-level jurisdiction governments than at the central level, because of their proximity to residents.

There are three main approaches to the analysis of governments:

1. Standard welfare economics assumes a benevolent government, with politicians appointed by majority voting to public office pursuing the maximization of citizens' economic welfare. To avoid the shortcomings of centralization stemming from heterogeneous preferences, a centralized system governed by a benevolent social planner could allocate different levels of public goods, each of them responding to the different preferences of local communities. However, the conflict of interests about how to distribute costs and benefits across jurisdictions may cause excessive public expenditures due to strategic delegation in the elections (Besley and Coate, 2000).
2. The public choice view maintains that central governments are inefficient Leviathans. The centralized state facilitates politicians maximizing their own utility function (that is, the objective of re-election), instead of pursuing the common good defined by a social welfare function. The expansion of public deficits for electoral reasons, as formalized by the well-known 'theory of the political cycle' (Nordhaus, 1975), creates the negative externality of a soft budget constraint. The public choice ap-

proach argues that fiscal competition is a useful tool to keep the size of government small. Because of the proximity to the preferences of local communities, the competition principle is supposed to make economic efficiency a possible objective for lower-level jurisdictions.

3. The rent-seeking approach differentiates from the previous view because government failures are traced back to the competing interests of groups engaged in lobbying and bribing practices with the government, thus causing the diffusion of corruption. A decentralized provision of public goods is advocated to ensure that opportunities for rent seeking and the size of the central government do not mutually reinforce each other (Persson and Tabellini, 1994). Local governments are thought to be more efficient because the close monitoring of elected representatives by the local communities may discipline fiscal policy.

The following sections will examine, four categories of federalism: cooperative, fiscal, competitive and functional.

Cooperative federalism

The objective of cooperative federalism is to guarantee the right to uniform conditions of life across jurisdictions. The peculiarity of this institutional organization is the strong interconnection between the functions and prerogatives of the central government and those of the lower-level governments (the typical examples of which are Germany and, to a lesser extent, Canada). The objective of social cohesion stressed in the cooperative federalism literature can be analysed with reference to the two equalization principles: horizontal and vertical equity.

Under a uniform federal tax rate, disparities in per capita income across jurisdictions make the 'net fiscal residuum' (the difference between paid taxes and received benefits) different for individuals with an *equal* income, the reason being that taxes are a function of each individual income, while benefits depend on the different resource endowments of each jurisdiction. The adoption of the principle of horizontal equity on a federation-wide basis copes with this differential fiscal treatment given to equal-income individuals in lower-level jurisdictions (Buchanan, 1950). The cooperative federalism literature deals with the implementation of the objective of horizontal equalization through horizontal and vertical transfers, the latter going both downward (from the central government towards the 'poor' lower-level jurisdictions) and upward (from the 'rich' lower-level jurisdictions to the central government). Depending on how large are the disparities in the balance between rich and poor lower-level jurisdictions for both resources and needs of the population, equalizing inter-jurisdictional grants where transfers are decided on per capita income basis may even result in redistribution from the poor of

the wealthy jurisdiction to the rich of the backward jurisdiction. Jurisdictions producing public goods also enjoyed by other jurisdictions are entitled to 'conditional' matching grants (compensatory subsidies linked to the fulfilment of particular conditions). In the case of positive externalities, the principle of horizontal equity (equal net fiscal residuum for equal-income individuals) implies that the amount of transfers should correspond to the equalization between marginal cost and marginal benefit.

The cooperative federalism literature also deals with the objective of vertical equity. Income disparities across individuals, which are not provoked by their market decisions (their leisure/work or effort choices) but by the conditions of the jurisdiction they live in, should be compensated. Widespread interdependencies inside a federation make the Pareto improvements relevant from both the equity and efficiency points of view. Under a uniform federal tax rate, Pareto suboptimal equilibrium position on the utility possibility frontier may be due to the differential public and merit goods enjoyed by equal-talented individuals. The larger the disparities across lower-level jurisdictions in terms of resource endowments and public responsibilities stemming from the different needs of the population (infrastructures, education, health care and so on), the more fiscal equalization by matching grants is justified on the basis that with an equal federal tax rate the same level of public services should be provided federation-wide. An improvement in social cohesion and efficiency can be expected by the correction of resource imbalances across jurisdictions.

The equalization of fiscal capacity and public services is implemented by differentiating the receiving from the contributing jurisdictions. The computation is made according to the ratio between the index of fiscal capacity and the index of equalization for each jurisdiction (in order to annul the bias of different tax rates across jurisdictions). For the equalization objective pursued by the federation to prevail on the subsidiarity principle, the main state functions should be an exclusive prerogative of the central government, whereas the functions shared with the lower-level jurisdictions should be exerted within the limits determined by the central government's legislation. Adding the vertical dimension to inter-jurisdictional transfers required by horizontal equity may substantially change the interpersonal income distribution that was reached *within* each jurisdiction. In fact, two unwelcome consequences may result. First, equalization grants aimed at increasing the benefits for individuals living in poor jurisdictions might require much wider federal transfers than those aimed at achieving equal net fiscal residuals for equal-income individuals living in different jurisdictions. This may start an inter-jurisdictional conflict. Second, the implementation by the federal government of a principle of vertical equity through the compensation of disparities in fiscal treatment caused by per capita income divergences across jurisdictions jeopardizes the horizontal equity obtained through the transfers

aimed at annulling differences in the net fiscal residuum of equal-income individuals. This may cause a clash between the two principles. Moreover, the principle of vertical equity entails that the central government implements a principle of vertical equity evenly shared by all communities belonging to the federation. Hence, local governments are not allowed to implement their own preferred degree of vertical equity. The local communities, if net contributors, will complain that they should not be forced to assume financial responsibility with the other jurisdictions for the implementation of whatever degree of vertical equity at the federal level. In fact, the objective of social cohesion endorsed by cooperative federalism could be rejected, as 'freedom of choice' claims that each jurisdiction has the right to set up its own degree of vertical equity (Boadway and Flatters, 1982).

The crucial problem undermining the implementation of cooperative federalism concerns the system of incentives. To what extent the redistributive effort implied by horizontal and vertical equity is carried out heavily depends on the tax rate which a society is keen to afford. In a multi-layer constitutional organization, centralization entails the integration of the jurisdictional programmes of public policies and the aggregation of the two electorates. Suppose two jurisdictions of equal size and income distribution. In the case of social insurance programmes the kernel of income distribution is *right-skewed*, so that the federal majority voting will favour a lower federal tax rate than it was in the jurisdictions (Persson and Tabellini, 1994). In the case of unemployment insurance programmes, the kernel of income distribution is *left-skewed*, which reverses the result. In fact, a too high federal tax rate would be decided by constituencies formed by the electorates of two jurisdictions with a high and a low probability of being hit by a negative asymmetric shock, respectively. The wider is the gap between the two jurisdictions in the probability to be hit by a negative shock, the more the risk-averse poor in the 'low-risk-of-unemployment ' jurisdictions voting in favour of a larger unemployment insurance will exceed in number the risk-averse rich of the 'high-risk-of-unemployment' jurisdiction voting against (Persson and Tabellini, 1996a, 1996b).

Fiscal federalism

Theories of multi-layer institutional organization builds upon the Oates Decentralization theorem (Oates, 1972), according to which the upper government level comes in only when the lower government level proves inefficient for some reason. Heterogeneity in preferences across local communities is expressed by different demand curves (each one a function of the price of the public good) and the desired quantity of the public good is determined at the intersection with its price level (which is equal to a constant marginal cost). The centralized federal provision averaging out heterogeneous preferences

across communities entails a welfare loss for some local community, either because of excess provision (the marginal cost is above the marginal benefit) or because of an insufficient provision (the marginal cost is below the marginal benefit).

In the Oates model, spillover effects across jurisdictions are assumed to be non-existent. The implicit hypothesis is that every 'package' of taxes and public goods offered by a local government is perfectly oriented to its own community, because in each jurisdiction all individuals share the same 'jurisdictional' preference for taxes and public goods. Yet, the decentralization theorem is very heroic, due to the 'correspondence principle' which postulates a complete overlapping between the territory of the political jurisdiction and the area over which the economic effects of public policies are spread. On the other hand, the capacity of the packages to satisfy the particular preferences of each community does not rule out the possibility of externalities. In other words, the jurisdiction does not necessarily correspond to the 'economic' jurisdiction. The public goods produced in a jurisdiction often have positive spillovers, as they can be used by individuals belonging to other jurisdictions (for example, health care, transportation and so on). Benefits stemming from public goods may not be fully enjoyed by those individuals who paid for them, and some of these benefits may even accrue to individuals who did not contribute to their production. Therefore, the real world suggests that fiscal federalism has to take issue with two important questions: (i) individuals belonging to the same community have heterogeneous preferences; and (ii) the compatibility among the optimal dimensions of all the goods in the package is not guaranteed.

As a solution to these questions, the Tiebout model (Tiebout, 1956) has suggested that competition among lower-level jurisdictions in offering packages can be instrumental in fostering economic efficiency. Given similar jurisdictions with the same tax rate, the following conditions should hold for decentralized taxation and public goods provision to be efficient: (a) zero cost mobility of individuals (and financial capital); (b) perfect supply elasticity of competitive jurisdictions; (c) public goods are provided at the minimal average cost; (d) complete information (and common knowledge) about preferences and the offered packages of tax and public goods; and (e) inter-jurisdictional externalities are absent. The resulting efficiency condition is that the taxes levied to each individual equalize the cost of production of the public goods desired by residents. The production of public goods responds to the well-known Samuelson condition whereby the marginal benefit equalizes the marginal cost. If goods and services are exposed to congestion, for the competition among the lower-level jurisdictions to be efficient, the additional qualification has to be added that the public sector production abides by the condition of the Buchanan's 'theory of club goods': the average

cost of one more resident must equal the marginal cost of one more resident (Buchanan, 1965).

Tiebout's rationale for inter-jurisdictional competition is to mimic the functioning of the perfect competition market. Each package represents the preferences of a certain type of individual. In order to attract the highest number of heterogeneous residents–consumers, each jurisdiction supplies a variety of packages and each individual will become resident of the jurisdiction in which his/her utility is maximized. Since packages come in many combinations (for instance, a certain public transportation system, associated with a day-care system with given characteristics, associated with a garbage collection system fitting the metropolitan environment and so on), the number of packages needed for all 'joint preferences' to be satisfied is extremely large. The condition of the perfect competition market is met when many jurisdictions supply every package.

The Tiebout model is interesting because it allows the existence conditions for a competitive market for public goods to be neatly analysed. First, the supply of public goods is efficient when it creates benefits only for residents and taxes are levied only on them. The consideration of jurisdictions which are different for even a single factor produces interdependences among them, thus preventing the efficiency conditions from being met.

A typical differential factor is represented by technological conditions of production across jurisdictions. Inside large jurisdictional entities, such as the US and the EU, different technology levels are a main determinant of the divide between the centre and the periphery. Let us assume that economies of scale have fostered territorial concentration of research laboratories and firms operating in advanced sectors – such that vacancies are concentrated in a certain area of a federation. This raises a problem. On the one hand, for the number of jurisdictions compatible with perfect competition to exist, the communities should be uniformly spread over the territory. On the other hand, the high labour demand in the 'agglomeration jurisdiction', where most dynamic industrial sectors and services (and most job vacancies) are concentrated, attracts many residents from other jurisdictions.

Note that this example abides by the Tiebout hypotheses. In particular, it does not conflict with condition (e). The presence in the agglomeration area of firms operating a demand for labour which attracts workers from outside does not represent an externality but an interdependence, as it can be solved by the system of prices absorbing the excess demand. Were the supply of public transportation not equalized with the commuters' demand, the commuting private costs could be too high and the mobile individuals would be forced to move to the agglomeration area. This possible outcome fails to meet condition (b) of efficient supply of jurisdictions, as for the optimal dimension for all public goods to be obtained the infinite supply of jurisdic-

tions should be violated. The advantage of living in an advanced area may be counterbalanced by the price differentials (for instance, in housing) which opens *vis-à-vis* the other jurisdictions due to the industrial concentration, which encourages commuting. By defining B(1) the benefits and MC(1) the marginal costs of congestion in the agglomeration area and B(n – 1) the benefits and MC(n – 1) the marginal costs of congestion in the other (n – 1) jurisdictions, the inequality B(1) – MC(1) < B (n – 1) – MC(n – 1) follows. To cope with the augmented congestion costs, taxation is increased in the other jurisdictions (let us call them 'residence jurisdictions') and the supply of public goods will shrink. This public sector reaction to excess demand disturbs the conditions for perfect competition among jurisdictions. The failure to meet the residence choice of the commuters reveals that the conditions of perfect competition for the supply of 'residence jurisdictions' are no longer fulfilled.

In the opposite case, the increased demand for public transportation is 'accommodated' and an inefficient outcome ensues. Since the residents' choice conflicts with the economies of agglomeration, in the residence jurisdictions the excess demand for public transportation by the commuters working in the agglomeration area determines a much higher dimension of public transportation with respect to the one that would equalize the demand for the other public goods (assuming that they equalize with the federal average). The supply of jurisdictions, and thus of packages, is in the right dimension, but at the cost of causing capital and labour misallocation across jurisdictions. With the given quantity of capital and labour, the optimal dimension of the public good 'transportation' will be obtained in the residence jurisdictions by reducing to a suboptimal dimension the supply of other public goods. The transportation supply has to satisfy the commuters' demand, and at the same time comply with the balanced public budget constraint (which is implicit in the perfect competition conditions for the supply of packages). Hence, the excess expenditure for transportation is a cause of crowding-out. Commuters and non-commuters of the residence jurisdictions would suffer from the costs of a lower supply of the other public goods. Since earmarked taxes cannot exceed the amount related to the financing of the optimal dimension, the public sector reaction to higher taxes is offset by a public deficit, in order to cope with the optimal dimension in the supply of all public goods.

Let us now waive the condition of a given quantity of capital, which is unrealistic once capital market liberalization is considered. On the one hand, every excess of public expenditure to be financed gives to each residence jurisdiction the incentive to attract foreign investments looking for the best profit conditions in liberalized financial markets. In order that the optimal dimension be financed and the public budget be in equilibrium, the tax rate is increased. On the other hand, under the hypotheses that individuals will move

into those jurisdictions offering the packages that best suit their preferences, and the income level being the main determinant of preferences, each jurisdiction involved in the strategic game has the incentive to attract the individuals with the largest tax base by reducing the tax rate and targeting the public goods preferred by them. The ensuing tax differentials may undermine the efficiency level of the transfers and services ranging from the risk insurance to the purely redistributive institutions of social protection (health care, education, poverty subsidies and so on). Fiscal competition across jurisdictions produces a reinforcing of spillover effects, which may foster a 'race to the bottom' of tax rates and public goods across jurisdictions.

The outcome is that the type composition of jurisdictions changes: they will no longer comprise heterogeneous individuals, but will drift towards 'homogeneity', as some jurisdictions will essentially be composed of rich individuals, while poor individuals will concentrate in other jurisdictions. A possible explanation is that the goods offered by the public sector are merit and public goods, for which competitive markets fail. In these service sectors characterized by market failures, the efficiency-enhancing incentives corresponding to the profit motivation of competitive firms cannot operate. Introducing competition among local governments in the supply of packages amounts to reproducing the same failures of private markets of perfect competition (Sinn, 2003). The suggestion is that a devolution process endowing lower-level jurisdictions with substantial fiscal autonomy should not be devised along the lines of the Tiebout model.

As far as social theory is concerned, at the beginning of the last century Max Weber perceived the trend towards homogeneous communities and theorized that individuals as rational maximizers pursue the formation of 'social closures' because of their fall-out in terms of identity of interests (Weber, 1922 [1974]). Empirical research work on social segmentation in the US seems to confirm that the reduction in income redistribution depends on the increasing income homogeneity in local jurisdictions, as an effect of the community formation process according to the Tiebout model (Epple and Sieg, 1999). They even make recourse to a policy of 'exclusionary zoning', that is, regulations aimed at excluding from the area people with the same preferences as the community but a much lower income level (Cooter, 2000).

Competitive federalism

The Tiebout strategy to endow jurisdictions with the optimality conditions for perfect competition consists in getting rid of both the consumers' and the politicians' moral hazard behaviour. On the one hand, in this decentralization model the preference revelation problem, which is typical of public goods, is solved. Inside smaller communities, the politicians' information costs are drastically reduced, so that the asymmetric information problem between

residents and elected representatives is overcome. Buchanan's optimality condition for public goods subject to congestion, with exclusion by a positive price, is a more general revelation mechanism of individuals' preferences than the Samuelson condition. Once fully mobile individuals have autonomously chosen to settle in the most preferred lower-level jurisdiction, the correct preference is elicited from residents by a local government. On the other hand, similarly to 'price-taking' firms in competitive markets, the local jurisdictions of the Tiebout model are 'utility-taking', as the consumer – no longer the central government – is the principal of the local government agent. Since the decision to settle in a certain jurisdiction corresponds to a willingness to 'buy' a certain package, this preference expressed by mobile individuals replaces equilibrium prices as the signal of efficiency. The assumption is that local governments efficiently organize the provision of public goods by taxing mobile factors with benefit-related levies. Individuals are then informed of the cost of consuming different levels of public goods and pay the exact amount of taxes for the benefits they receive.

The so-called 'yardstick competition' approach takes a step in the direction of making the individual the principal of the local government agent. By eliciting preference revelation from consumers, an increase in the efficiency of public goods provision is expected. Granted that lower-level jurisdictions are sufficiently similar to be comparable, voters may use information on other jurisdictions as a yardstick to evaluate their government's achievements. The correspondence between market competition and inter-jurisdictional competition is achieved when individuals – in addition to the exit option both from the market and from the jurisdiction – also enjoy the condition of freedom of exit from a jurisdictional government. They get rid of public policies they dislike by voting against the party – or the coalition of parties – which has implemented them, instead of abandoning the jurisdiction (Salmon, 1987). In the Tiebout model, the 'voting with one's feet' mechanism fulfils the objective of no longer subsuming the jurisdictions' public good provision to voting majorities manipulated by political parties, but subjecting it directly to the scrutiny of the market by means of the exit option. On the contrary, in the yardstick competition approach, voting against the government corresponds to making recourse to 'voice' (Hirschman, 1970), which has a greater generality than exit and does not require the hypothesis of zero cost mobility.

Given the objective of convincing the non-sympathetic electorate to join its own political side, opposition political coalitions governing similar lower-level jurisdictions struggle to be recognized as the best performer. The reciprocal monitoring among the competing governments of lower-level jurisdictions not only solves the typical agency problem of asymmetric information between government and citizens, but is also alleged to have the same efficiency-enhancing effect of fiscal competition. Moreover, similarly

to the horizontal competition across governments of lower-level jurisdictions, vertical competition between federal and local elected representatives sharing the same area of responsibility may discipline lower-level politicians aiming at upper-level governmental positions.

The economic analysis of yardstick competition also relies upon the unexpected utility theory of rational choice. The transaction utility theory (Thaler, 1985) builds upon the so-called 'reference point effect'. In evaluating political issues, individuals may recognize a reference point – a standard – and are influenced by it. Whenever the goal of economic efficiency is associated with compliance with a certain standard (for instance, a balanced public budget or a certain employment rate), the individuals of one jurisdiction – by looking at the other jurisdictions' performance – might attach a negative (positive) utility in the event that their jurisdiction's performance is worse (better) than the others. The conclusions drawn from the comparison will dictate their behaviour in the polls. The main difference between the yardstick competition and the transaction utility theory approaches is that in the latter the behaviour of both the politicians and the voters is biased by the reference point effect, while in the former there is an information asymmetry favouring the politicians, as voters are unaware of all circumstances determining the respective performance of governments of similar jurisdictions.

Since the degree of similarity among lower-level jurisdictions is often low, the moral hazard problem raised by asymmetric information is a major obstacle. The specialization in legal services pursued by the state of Delaware is an interesting application of the yardstick competition approach to subsidiarity as an efficiency-enhancing mechanism. The Delaware politicians chose to boost tax-raising by giving the incentive to private companies looking for a reduction in negotiation and litigation costs to establish their legal location in their state through the strengthening of the supply of legal services (Romano, 1987). It has been alleged in the literature that excessive deregulation has been produced by courts handing down sentences that are aimed at pursuing efficiency in company law by heeding company interests. Since juridical positions in the United States are elective, and the lawyers hired by the managers of the companies often become judges and vice versa, the thesis has been put forward that this outcome depends upon the personal relationships between managers and judges. The indication is that the moral hazard problem stems from the perverse functioning of the agency relationship between mobile companies and public officials, whatever the level – central or local – of government.

Functional federalism

In contrast to the benevolent government of the traditional welfare economics analysis, in Tiebout's fiscal federalism model or in the yardstick competition

approach institutional competition is considered an effective instrument to tame the Leviathan. Fiscal and competitive federalism aim to extend the incentive system implied by competition to the organization of jurisdictions. The decentralized organization is alleged to restrict the political domain and prevent elected representatives from pursuing their own personal interests. This should eradicate the problem of excessive and/or inefficient public intervention and boost efficiency in public goods production. However, the extension of the conditions for perfect market competition to jurisdictions is hampered in the Tiebout model by the heroic condition of zero cost mobility, and in the yardstick competition approach by the assumption of zero transaction costs of exiting from the government.

The functional federalism programme aims to show that both these shortcomings disappear once the conceptual tension existing between the two institutions of the market and the government is solved. It has been observed that consumers can enter and exit from markets, but with positive mobility costs they are 'locked in' their political relations with the state (for example, the taxes and transfers determined by the government) (Buchanan, 1992). While a market that ends up with no customers just disappears, the political entity ruling in a certain territory can survive a certain amount of community aversion to the public goods provided. To establish the parallel between exit from markets and exit from governments, the positive costs of both 'mobility by feet' and 'mobility by voting against the government' are to be annulled. Once individuals are endowed with the right to opt out from a jurisdiction in the matter of consumption of public goods, institutional mobility comes at no cost.

To this end, functional federalism advocates the strategy of separating public goods provision from the political jurisdictions. Individuals who are not satisfied with the provision of public goods in their jurisdiction should not have to physically abandon the jurisdiction or to exit from their government by voting against it. Residents of different areas, but sharing the same preferences (for instance, because of the similarity of their income levels), may reach an agreement for a collective action aimed at setting up an agency for the provision of public goods (or choose among already established mono-functional providers) (Casella and Frey, 1992). The homogeneity of preferences that characterizes the markets for private goods is supposed to be replicable in the collective action for the production of public goods.

The strategy is the following. Since individuals belonging to different jurisdictions join a 'club' and each individual can participate in various clubs, the organization of the production of a public good overlaps various jurisdictions. Each individual or each community can freely enter or exit from one or more agencies, each offering a different package of tax and public goods. The self-selection of participants makes endogenous the number of agencies as a

function of desirable packages. The only limit to participation is the fulfilment of the congestion constraint according to Buchanan's theory of clubs. The optimal dimension of each club is endogenously determined through the aggregation of the demands for a specific public good and controlled by exclusion clauses for members who do not pay contributions, so that the congestion constraint is met. Therefore, this model of organizing public goods production builds upon two main characteristics: (i) the clubs are mono-functional, whereby homogeneous groups of individuals vote on mono-purpose taxes, each one earmarked for the financing of a specific public good; and (ii) the clubs are inter-jurisdictional, as they do not derive their legitimacy from a state settled in a territory, but solely from the free choice of individuals to join one or more 'extra-territorial jurisdictions'.

The functional federalism way to extend market competition to jurisdictions might require not only either Tiebout's exit or the 'voice' mechanism devised by competitive federalism, but also the 'loyalty' to mono-functional and inter-jurisdictional agencies detached from the political jurisdictions settled in a territory. In the trans-jurisdictional agencies all three options put forward by Hirschman (1970) might be present: loyalty, which could also rely on voice, which in turn might be accompanied by the threat to exit.

However, the extension of market organization to public goods provision could result in a failure to fulfil both the efficiency and the equity objectives. As for the first, the stability across time of the sense of collective identity of a community group depends on the efficient governance of clubs and the influence on the functioning of the private sector. Due to interdependences and complementarities, the individuals participating in the various agencies are bound to suffer from strategic behaviour due to uncertainty and *ex post* opportunism in the agencies as well as in the markets. As for the second, the income-homogeneity character of the trans-jurisdictional agencies, by magnifying the demise of the solidarity principle, is disruptive for social cohesion.

It can be observed that the organization of the mono-functional agencies envisaged by functional federalism closely resembles the relationship with clients entertained by private companies operating in the insurance markets. Due to the well-known moral hazard and adverse selection problems of insurance markets, these latter agencies tailor contracts to suit just low-risk individuals. The high-risk individuals, to whom contracts are offered at the very high price corresponding to their 'actuarially fair' value, cannot afford insurance. They might also be low-income, due to spillovers across dimensions of well-being. The 'cream skimming' practised by private insurance companies (that is, the selection of best clients by implementing separating equilibria), represents the institutional design for the agencies operating public goods provision across jurisdictions. Moreover, those in the high-income

and low-risk category have an incentive to change residence and settle in low tax and social expenditure jurisdictions. Income homogeneity across jurisdictions is the possible final outcome of fiscal competition among jurisdictions of the Tiebout model, although this process is slowed down in many countries by the still low degree of labour mobility. In the functional federalism programme, income homogeneity represents both a condition for optimality in public goods provision and the founding character of inter-jurisdictional agencies, as similar individuals sharing the same income-related preferences are expected to aggregate in mono-functional agencies.

The tendency to income homogeneity is not opposed by economic liberalism. On the contrary, the liberal appraisal of the social contract questions the redistributive outcome implied by the 'pooling' of high-income and low-risk individuals in the tax funding of public social protection institutions (Sugden, 1993). However, globalization has severely challenged the capacity of the modern democratic states to implement the social rights implied by the principle of 'positive freedom'. As a consequence, the recognition that a person is an individual with rights is now an important factor that many international organizations for world cooperation incorporate explicitly in their statutes.

Constitutional change

Constitutional change – the building or the disintegration of a state – is often a consequence of a new foundation of economic structures. The determining causes typically are a war or a severe economic crisis. We shall hereafter deal with the nexus among markets, institutions and the state, with a focus on the linkage between the evolution of economic structures and of political structures, the adhesion to the process of construction of a federation by a group of founding countries or entry at a subsequent stage, and the secession of a jurisdiction from the federation of belonging.

Economic integration and political institutions

Since there is sometimes a substitution nexus and at other times a complementarity nexus between economic structures and political organizations, their relationships are very complex and difficult to assess.

Let us start from the substitution nexus. Globalization is interpreted as the emergence of a market economy freed from the regulation constraints imposed by national states. In contrast to the traditional link between firms and national economic systems, multinational companies are supposed to waive exchange rate policies that protect competitiveness, monetary policy for stabilization purposes, and fiscal policies that guarantee public transfers to economic sectors and areas in need. The establishment of an international economic system centred on private companies operating in globalized mar-

kets is alleged to make the traditional state function as an obstacle to static and dynamic efficiency. In order to overcome all the inefficiencies attributed to a 'big government', the demise of fiscal policy centralization is also advocated. Overall, economic integration should be followed by political disintegration in small countries with a 'light' public sector (Alesina et al., 2000).

This view has been formalized in a model of federalism – denominated 'flexible union' – midway between centralization and decentralization, inspired by the Inman and Rubenfeld (1997) proposal, referred to previously. A flexible union is defined as the institutional organization where the supply of a public good by the federal government is inferior to the level that would prevail in a fiscal union. Following the subsidiarity principle, the federal supply is determined to internalize externalities across countries. Countries with a lower-than-median preference for the public good do not add a supplementary supply, and countries with a higher preference (among them, possibly also the median country) supply more (Alesina et al., 2001a). Public goods provision is Pareto superior in a flexible union compared to a centralized one. The demonstration draws on the median country's special position. If its preference is low, the supply by the central government is considered sufficient; however, if the median country is in the high-preference group and the supply level close to the centralized one is preferred, then the median country government increases its supply to its favoured level.

As for the complementarity nexus between economic integration and political integration, the historical evidence suggests that political unification should precede economic integration. In recent times, this 'rule' has been violated on only two occasions. In 1834, a customs union (*Zollverein*) promoted economic integration among some German states, and preceded the political union, which did not occur until 1871, in the Second Reich. The process of European economic integration, where the progressive release of powers to supra-national institutions – culminating in the 1990s with the single market, the common currency and the submission of national fiscal policies to the European Commission's verification of compliance with the Stability and Growth Pact – took place without a previous political unification. The persistence of the lack of European Union political representation favours the reductionist interpretation presenting European integration as a typical case of globalization.

As a matter of fact, the rapid transformation of the European markets *vis-à-vis* the lengthy period of evolution for political structures seems to legitimate this interpretation. Among the main clues of globalization are the concentration in the central EU regions of transnational clusters of economic activities in advanced sectors and public and private research institutions, the outsourcing in the East European countries of parts of the productive processes of Euro-

pean companies, and the financial mergers in the manufacturing and banking sectors across European countries. Yet, the improvement of communitarian legislation in various fields of economic regulation is a signal of the need for institutional underpinnings to sustain these economic integration processes. Although the traditional causality nexus, from the construction of a fully-fledged federation to market integration, functions the other way round in the European Union, the evolution of regulatory institutions is coping with the problems stemming from the market and monetary integration that has preceded the political integration.

The role of institutions in economic integration

Institutions play a decisive role in orienting the organization of markets. An example is an integration process whereby founding members and accession countries have to reach a decision on standardization (for example, a common poverty line, common environmental regulations and so on). The increasing heterogeneity across member states created by successive accession agreements threatens the efficiency of the integration process.

A common policy may consist of the member states' convergence to a common standard. This strategic enterprise can be sketched as a 'battle of the sexes' (see Figure 12.1a), a coordination game with a conflict of interests whereby each country prefers convergence to its own preferred outcome, with three Nash equilibria possible. Although the pay-off structure shows that both players are discontented if the other one prefers a different standard, both are better off with a consensual solution. Convergence by a player to the other player's preferred standard is then Pareto superior to the defection strategy.

It will be shown that in an institution-building process, the threat power capable of fostering the coordination outcome can be determined by the juridical institutions, such as the articles of a treaty stipulating the principles

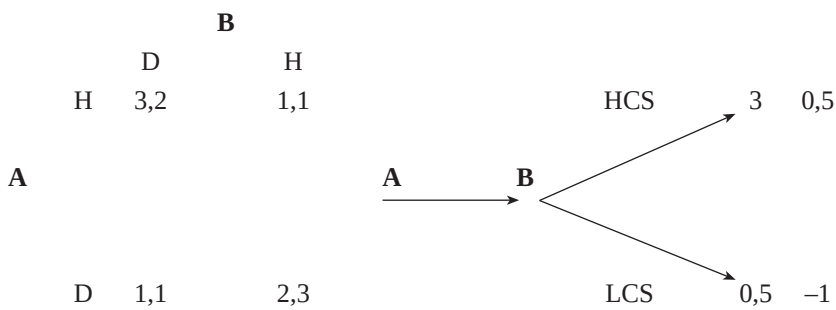


Figure 12.1 Role of institutions in economic integration

to be upheld by the member states as well as by the accession countries (in the given examples, the right to social inclusion and to a healthy environment, respectively).

In the game in Figure 12.1a, each player's choice to comply (dove: D) with the other player's strategy leads to the (3,2) or (2,3) outcomes (then excluding the one in the mixed strategy), which are Pareto-superior with respect to the aggressive strategy (hawk: H) to stick to his own preference (coordination failure: 1,1). Figure 12.1b illustrates that the strategic environment modifies when a supra-national institution affects the players' ranking between strategies. Suppose that a group of countries (A) prefers a high cost standard (HCS), and another group of countries (B) prefers a low cost standard (LCS), e.g. for anti-poverty or anti-pollution policies. Hence, group A is doomed to suffer from an externality (e.g., generous subsidies attract poor immigrants from B, or high pollution damages, cost to A a loss = 0,5). The supra-national institution could consider the anti-poverty or anti-pollution policies a common interest, so that some articles are included in a treaty aimed at playing the role of the *deus ex machina* of the game. The articles state that group B must take an action (e.g. investments in anti-poverty policies, in anti-pollution policies, etc.) to comply with the HCS standard (investment costs = 1,5) within a deadline, after which group A has a right to retaliation (e.g. a tariff on export from B to A, costing to B a loss = 2) in the case of non-compliance. Due to externality and retaliation, in Figure 12.1b the pay-offs of the cell of coordination failure (1,1) are reduced to 0,5 and –1, respectively. The supra-national institution has then changed the structure of the game, as the articles represent the information, for both groups A and B, that A has been assigned the advantage of committing to its preferred common standard HCS and just wait for the B's coordination. In game-theoretic language, an extended form of the game could be envisaged in which the treaty dictates a sub-game perfect equilibrium. In our simplified appraisal of this strategic interaction, group B knows for certain that in pursuing its own preferred common standard (LCS), by continuing the game on the bottom half of the tree, the worst outcome for both A and B (0,5, –1) would result. Group B is then forced to comply with the common standard HCS, which amounts to choosing the top half of the game tree. Hence, the (3, 0,5) outcome obtains, with the division of the surplus favouring group A because of the anti-poverty or anti-pollution policies implemented by group B. Therefore, for each group the balance between benefits and costs stemming from the common policy – and then the success or failure of the coordination – also depends on the capacity of the juridical architecture of the collective action to constrain the players' strategies.

Institution-building

The strategic game among countries engaged in an institution-building process is a coordination game with a conflict of interest. The cooperative agreement reached by majority rule manifests a tradeoff between the benefit of reducing the compensation for the losing interest groups and the cost of possibly paving the way to the 'tyranny of the majority'. The lower the size of the required majority, the higher the probability of a successful bargaining on the compensation for the losers, but the more likely the expropriation of the minority by excessive *ex post* redistribution. In a fully-fledged federation, the negative effects of this tradeoff shrink, compared to a mono-level constitutional organization. For historical reasons, federations show a strong heterogeneity in the distinctive features and preferences of their constituent communities. By virtue of the heterogeneous composition both within and between the jurisdictions of the federation, the probabilistic value of the risk of belonging to an exploited minority declines. In the polls, a minority supporting a certain policy in a state is likely to be backed by a majority in another state endorsing the same policy. The federal voting then provides a sort of pooling of the risk of being in the exploited minority (Aghion and Bolton, 2003).

Let us now examine the institution-building of a federation by a group of countries. The choice between a decentralized and a centralized fiscal policy depends on the tradeoff between the factors described earlier. On the one hand, the internalization of spillovers across jurisdictions and economies of scale are two important sources of cost reduction, favouring centralization. On the other hand, local governments are expected to be more efficient, being more informed about the local communities' preferences, and a high heterogeneity across jurisdictions as to preferences for public goods makes centralization costly for countries with a per capita income very different from the median country. This favours decentralization.

The dispersion of per capita income across jurisdictions is a crucial determinant of the choice between centralization in a fully-fledged fiscal union and decentralization. The greater the expected benefits from spillovers' internalization and economies of scale stemming from common policies that do not modify the per capita income dispersion across jurisdictions (transportation infrastructures, ICT networks and so on), the larger the number of countries that would incur a loss by not participating in common policies (Alesina et al., 2001b). Yet, the higher the per capita income dispersion across member states, the wider the dispersion of preferences regarding social insurance policies which are expected to generate spillovers across member states (unemployment transfers, health care, social security, education and so on), the higher the number of countries expecting to lose after centralization. Given the divergent preferences held on desired taxation and public goods by

the decisive median voters in each jurisdiction, it is unlikely that an agreement on centralization would be reached. In addition to the expected dispersion of per capita income *between* jurisdictions, it is also possible that the expected federal median-to-mean income ratio *vis-à-vis* the median-to-mean income ratio *within* each jurisdiction will be decisive in the choice between centralization in a fully-fledged fiscal union and decentralization. The federal voting must be carefully analysed: if the inter-jurisdictional income distribution of the federation is more dispersed than most intra-jurisdictional income distributions, in a would-be federal election the decisive median voter is likely to be poorer than the mean-income-voter and would vote for a higher tax rate than most jurisdictional median voters are currently choosing. The higher the fear of an excessive inter-jurisdictional redistribution, the higher the probability that a country where a 'high' per capita income is associated with a high median/mean income opts out from the institution-building process (Croci Angelini et al., 2001).

Moreover, apart from the expected inter-jurisdictional redistribution, some countries may fear that economies of scale or positive spillovers stemming from the accumulation of common policies – due to complementarities and path dependency – could pave the way to a federation, which they do not support. Thus, the larger the number of countries participating in an institution-building process, the higher the probability that some countries will refrain from furthering the institution-building process (or from participating in some common policies, if opting out is allowed).

The public choice school has analysed another decision mechanism governing public policies' centralization. Political pressure to adopt a common policy may be exerted on national governments by interest groups, such as lobbies, possibly paying contributions to their governments. These interest groups operating in the same economic sector but belonging to different member states may make transnational alliances on a functional basis. The aim is to exert political pressure on the national governments for the implementation of a common policy from which they could receive high profits due to economies of scale, internalization of negative externalities and so on. A transnational coalition of interest groups may even acquire the bargaining power needed to force governments to pass a common policy at the supranational decision-making level in every country, regardless of the preferences of the national median voters.

The strong incentive to log-rolling could increase the number of centralized policies. Due to the political influence of transnational interest groups, a common policy proposed by countries gaining from a common policy will also be endorsed by countries penalized by it, to which they will reciprocate by endorsing another common policy favouring these latter countries. Moreover, the larger the profits collected by a transnational lobby, the greater

also the possibility of compensating some losers and increasing the number of member states participating in the launch of a common policy. Given this positive correlation between the number of transnational interest groups and the number of common policies, a positive correlation may also be found between the size and the scope of a political integration process.

Secession

The determinants of secession are similar to those of accession, but the costs arising from the separation process also have to be taken into account. The seceding jurisdiction bears both sunk costs and losses of economies of scale related to infrastructures, networks of public utilities and so on. The transaction costs have to be considered too: there might be a bargaining among jurisdictions over public debt division and a share of it to be repaid is assigned to the seceding jurisdiction.

In a fully-fledged federation, the fiscal system determines the federal interpersonal distribution through a risk-sharing scheme to deal with the macroeconomic risk of unemployment. The jurisdictions with the highest per capita income inequality are also those with a lower probability of being hit by a negative shock. These jurisdictions might complain about being drawn in by those with a 'high risk' of negative shock and be tempted by secession (Bolton and Roland, 1997). Like the rationale discussed above, in addition to the per capita income dispersion across jurisdictions, the median-to-mean income ratio also affects the propensity to secede. If the federal interpersonal distribution is greater than the inter-jurisdictional distribution of the various member states, then the federal median voter is likely to have a lower income than the federal mean – compared to the median voters of the various member states with respect to their national mean income. The voter will then be in favour of a higher tax rate. The more an inter-jurisdictional redistribution is pursued by the central government, the less the per capita income will vary among jurisdictions, and the higher the incentive for 'rich' jurisdictions to secede. In the opposite case of the federal median voter favouring a lower tax rate, a jurisdiction plagued with very low levels of both per capita income and median-to-mean income ratio (with respect to the other jurisdictions) would complain about the ineffective redistributive effort taking place both at the inter-jurisdictional and the interpersonal levels inside the federation. If the jurisdiction's influence on the central government is negligible, the jurisdiction may seek secession.

An incentive strengthening the decision in favour of secession is an 'outside option', that is, the opportunity to join another political entity (the secession of Slovakia from Czechoslovakia can be explained by the expectation of a higher risk insurance to be obtained inside the European Union than through the Prague government). Secessions, devolution to lower-level gov-

ernment (Spain, the United Kingdom, Italy, Belgium) and the de-structuring of large political entities (the Soviet Union, Yugoslavia) have been interpreted as a historical tendency to the formation of smaller political entities after markets' liberalization and globalization (Alesina et al., 2000). Yet, the need for insurance after the increased uncertainty brought about by globalization is fostering the search for supra-national institutions and increasing the attraction power exerted by political entities under construction. In Europe, the fragmentation tendency has been counteracted by the economic integration process that created the single market and monetary union, and is now generating expectations for a political union. The EU's eastern enlargement demonstrates that political fragmentation does not only represent a successful episode of market globalization, but may reflect the dawn of a historical phase of regeneration of political institutions.

References

Aghion, P. and P. Bolton (2003), 'Incomplete social contracts', *Journal of the European Economic Association*, **1**: 38–67.

Alesina, A., I. Angeloni and F. Etro (2001a), 'Institutional rules for federations', mimeo.

Alesina, A., I. Angeloni and F. Etro (2001b), 'The political economy of international unions', NBER Working Paper 8645, National Bureau of Economic Research, Cambridge, MA.

Alesina, A., A. Angeloni and L. Schuknecht (2001c), 'What does the European Union do?', NBER Working Paper 8647, National Bureau of Economic Research, Cambridge, MA.

Alesina, A., E. Spolaore and R.Wacziarg (2000), 'Economic integration and political disintegration', *American Economic Review*, **90**: 1276–96.

Aoki, M. (2002), *Towards a Comparative Institutional Analysis*, Cambridge, MA: MIT Press.

Arrow, K.J. (1994), 'Methodological individualism and social knowledge', *American Economic Review*, **84**: 1–9.

Austen-Smith, D. (2000), 'Redistribution of income under proportional representation', *Journal of Political Economy*, **108**: 1235–69.

Backhaus, J. (1996), 'An economic analysis of constitutional law', in E. Buscaglia, W. Ratliff and R. Cooter (eds), *Law and Economics of the Development*, Greenwich, CT: JAI Press.

Berlin, I. (1958), *Two Concepts of Liberty*, Oxford: Clarendon.

Besley, T. and S. Coate (2000), 'Centralized versus decentralized provision of local public goods: a political economy analysis', NBER Working Paper 2495, National Bureau of Economic Research, Cambridge, MA.

Boadway, R. and F.R. Flatters (1982), 'Equalization in a federal state: an economic analysis', Economic Council of Canada, Ottawa.

Bolton, P. and G. Roland (1997), 'The breakup of nations: a political economic analysis', *Quarterly Journal of Economics*, **112**: 1056–90.

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

Buchanan, J.M. (1950), 'Federalism and fiscal equity', *American Economic Review*, **40**: 586–88.

Buchanan, J.M. (1965), 'An economic theory of clubs', *Econometrica*, 32: 1–14.

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

Buchanan, J.M. (1985), 'Rights, efficiency, and exchange: the irrelevance of transaction costs', in *Liberty, Market, and the State*, Brighton: Wheatsheaf, pp. 38–59.

Buchanan, J.M. (1991a), *The Economics and the Ethics of Constitutional Order*, Ann Arbor, MI: Michigan University Press.

Buchanan, J.M. (1991b), 'Economic origins of ethical constraints', in Buchanan (1991a), pp. 179–93.
Buchanan, J.M. (1992), 'The individual as participant in political exchange', in S. Spilerman and A. Sorenson (eds), *Essays in Honor of James D. Coleman*, New York: Praeger, pp. 121–42.
Buchanan, J.M. (1993), 'How can constitutions be designed so that politicians who seek to serve "public interest" can survive and prosper', *Constitutional Political Economy*, **4**: 1–6.
Buchanan, J.M. and G. Tullock (1962), *The Calculus of Consent*, Ann Arbor, MI: Michigan University Press.
Casella, A. and J.S. Feinstein (2000), 'Public goods in trade: on the formation of markets and jurisdictions', *International Economic Review*, **39**: 24–76.
Casella, A. and B. Frey (1992), 'Federalism and clubs: towards an economic theory of overlapping political jurisdictions', *European Economic Review*, **36**: 639–46.
Coase, R. (1961), 'The problem of social cost', *Journal of Law and Economics*, **3**: 1–44.
Coleman, J. (1990), *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.
Cooter, R.D. (2000), *The Strategic Constitution*, Princeton, NJ: Princeton University Press.
Croci, Angelini E., C. D'Ambrosio and F. Farina (2001), 'Do preferences in EU member-states support fiscal federalism?', Quaderni del Dipartimento di Economia Politica, Working Paper 346, Department of Political Economics, University of Siena.
de Ruggiero, G. (1925 [1927]), *The History of European Liberalism*, Boston, MA: Beacon Press.
Dworkin, R. (1978), 'Liberalism', in S. Hampshire (ed.), *Public and Private Morality*, Cambridge, New York: Cambridge University Press.
Epple, D. and H. Sieg (1999), 'Estimating equilibrium models of local jurisdictions', *Journal of Political Economy*, **107**: 645–81.
Fitoussi, J.P. (2002), 'Democratie et mondialisation', *Revue de l'OFCE*, **75** : 7–19.
Greif, A. (1994), 'Cultural beliefs and the organisation of society: a historical and theoretical reflection on collectivist and individualist societies', *Journal of Political Economy*, **5**: 912–50.
Greif, A. (2001), 'Impersonal exchange and the origin of markets: from the community responsibility system to individual legal responsibility in pre-modern Europe', in M. Aoki and Y. Hayami (eds), *Communities and Markets in Economic Development*, Oxford: Oxford University Press, pp. 33–56.
Grilli, V., D. Masciandaro and G. Tabellini (1991), 'Political and monetary institutions, and public policies in industrial countries', *Economic Policy*, **22**: 37–68.
Habermas, J. (1988), 'Individuierung durch Vergesellschaftung' ['Identity through socialization'], in *Nachmertaphysisches Denken*, Frankfurt a/M: Suhrkamp, pp. 184–236.
Hayek, F.A. (1973), *Law, Legislation, and Liberty, Vol. 1, Rules and Order*, London: Routledge & Kegan Paul.
Hayek, F.A. (1976), *Law, Legislation, and Liberty, Vol. 2, The Mirage of Social Justice*, London: Routledge & Kegan Paul.
Hayek, F.A. (1978), 'The confusion of language in political thought', in *New Studies in Philosophy, Politics, Economics and the History of Ideas*, London: Routledge & Kegan Paul, pp. 71–97.
Hayek, F.A. (1979), *Law, Legislation, and Liberty, Vol. 3, The Political Order of a Free People*, London: Routledge & Kegan Paul.
Hirschman, A.O. (1970), *Exit, Voice, and Loyalty*, Cambridge, MA: Harvard University Press.
Hume, D. (1740 [1978]), *A Treatise of Human Nature*, Oxford: Oxford University Press.
Inman R. and D. Rubenfield (1997), 'The political economy of federalism', in D. Mueller (ed.), *Perspectives on Public Choice: A Handbook*, Cambridge: Cambridge University Press.
La Porta, R., F. Lòpes-de-Silanes, C. Pop-Eleches and A. Shleifer (2002), 'The guarantees of freedom', NBER Working Paper 8759, National Bureau of Economic Research, Cambridge, MA.
Laffont, J.J. and J. Tirole (1990), 'The politics of government decision making: regulatory institutions', *Journal of Law, Economics, and Organization*, **6**: 1–32.

Locke, J. (1640 [1988]), *Two Treatises of Government*, Cambridge: Cambridge University Press.
Madison, J. (1787–88), 'Federalist 10', in Alexander Hamilton, James Madison and John Jay, *The Federalist Papers*, New York: Bantam Books.
Marshall, T.H. (1981), 'The right to welfare', in *The Right to Welfare and Other Essays*, London: Heinemann Educational Books.
Marx, K. (1844 [1970]), *Critique of Hegel's Philosophy of Right* (*Kritik des Hegelschen Straatrechts*), Cambridge: Cambridge University Press.
McKinnon, R.I. and T. Nechyba (1997), 'Competition in federal systems: the role of political and financial constraints', in J. Ferejohn and B. Weingast (eds), *The New Federalism: Can the States be Trusted?*, Stanford, CA: Hoover Institution Press, pp. 68–94.
Milgrom, P., D.C. North and B.R. Weingast (1990), 'The role of institutions in the revival of trade: the law merchant, private judges, and the champagne fairs', *Economics and Politics*, **1**: 1–23.
Montesquieu, Charles-Louis de Secondat (1748 [1955]), *The Spirit of Law* (*L'Esprit des Lois*), Chicago: Encyclopedia Britannica.
Morishima, M. (1982), *Why Has Japan Succeeded?*, Cambridge: Cambridge University Press.
Mueller, D.C. (2003), *Public Choice III*, Cambridge: Cambridge University Press.
Nordhaus, W. (1975), 'The political business cycle', *Review of Economic Studies*, **42**: 169–90.
North, D.C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
North, D.C. (1991), 'Institutions', *Journal of Economic Perspectives*, **5**: 97–112.
North, D.C. and R.P. Thomas (1973), *The Rise of the Western World: A New Economic History*, Cambridge: Cambridge University Press.
Oates, W.E. (1972), *Fiscal Federalism*, New York: Harcourt-Brace & Jovanovich.
Oates, W. (1999), 'An essay on fiscal federalism', *Journal of Economic Literature*, **3**: 1120–49.
Persson, T., G. Roland and G. Tabellini (1997), 'Separation of powers and political accountability', *Quarterly Journal of Economics*, **112**: 1163–202.
Persson, T. and G. Tabellini (1994), 'Does centralisation increase the size of government?', *European Economic Review*, **38**: 765–73.
Persson, T. and G. Tabellini (1996a), 'Federal fiscal constitutions: risk sharing and moral hazard', *Econometrica*, **64**: 623–46.
Persson, T. and G. Tabellini (1996b), 'Federal fiscal constitutions: risk sharing and redistribution', *Journal of Political Economy*, **104**: 979–1009.
Persson, T. and G. Tabellini (2000), *Political Economics. Explaining Economic Policy*, Cambridge, MA: MIT Press.
Persson, T. and G. Tabellini (2003), *The Economic Effects of Constitutions*, Cambridge, MA: MIT Press.
Polanyi, K. (1957), 'Economy as instituted process', in K. Polanyi (ed.), *Trade and Market in the Early Empires. Economies in History and Theory*, New York: Free Press, pp. 297–331.
Posen, A. (1995), 'Central bank independence and disinflationary credibility: a missing link?', staff reports 1, Federal Reserve Bank of New York.
Putnam, R. (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ: Princeton University Press.
Rawls, J. (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.
Rawls, J. (1999), *The Law of Peoples*, Cambridge, MA: Harvard University Press.
Rodden, J. and Rose-Ackermann S. (1997), 'Does federalism preserve markets?', *Virginia Law Review*, **83**: 1521–72.
Romano, R. (1987), 'The political economy of takeover statutes', *Virginia Law Review*, **73**: 97–111.
Roubini, N. and J.D. Sachs (1989), 'Political and economic determinants of budget deficits in the industrialised democracies', *European Economic Review*, **33**: 903–38.
Salmon, P. (1987), 'Decentralisation as an incentive scheme', in *Oxford Review of Economic Policy*, **3**: 24–43.
Schotter, A. (1981), *The Economic Theory of Social Institutions*, Cambridge: Cambridge University Press.

Seabright, P. (1996), 'Accountability and decentralization in government: an incomplete contracts model', *European Economic Review*, **40**: 61–89.
Sinn, H.-W. (2003), *The New System Competition*, Oxford: Blackwell.
Smith, A. (1759 [1976]), *The Theory of Moral Sentiments*, Oxford: Oxford University Press.
Smith, A. (1776 [1976]), *An Inquiry into the Nature and the Causes of the Wealth of Nations*, Oxford: Oxford University Press.
Stiglitz, J. (1998), 'Central banking in a democratic society', *De Economist*, **146**: 199–226.
Sugden, R. (1993), 'Rights: why do they matter, and to whom?', *Constitutional Political Economy*, **4**: 127–52.
Thaler, R. (1985), 'Mental accounting and consumer choice', *Marketing Science*, **4**: 199–214.
Tiebout, C.M. (1956), 'A pure theory of local expenditures', *Journal of Political Economy*, **60**: 416–24.
Vanberg, V.J. (1994), *Rules and Choice in Economics*, London: Routledge.
Weber, M. (1922 [1974]), *Economics and Society*, San Diego: University of California Press.
Weingast, B.R. (1993), 'Constitutions as governance structures: the political foundations of secure markets', *Journal of Institutional and Theoretical Economics*, **149**: 286–311.
Weingast, B.R. (1995), 'The economic role of political institutions: market-preserving federalism and economic development', *Journal of Law, Economics, and Organization*, **11**: 1–31.
Weitzman, M. and C. Xu (1993), 'Chinese township and village enterprises as vaguely defined cooperatives', *Journal of Comparative Economics*, **18**: 121–45.
Wicksell, K. (1896), *Finanztheoretische Untersuchungen* [Studies in the Theory of Public Finance], Jena: Gustav Fischer.

13 Constitutional economics II

Ludwig Van den Hauwe

Introduction

The term 'constitutional economics' or 'constitutional political economy' was introduced in the 1970s to designate a distinct strand of research that emerged from the somewhat older public choice branch of economics.[1] In the 1990s, constitutional economics developed into a major research programme. At a time of massive worldwide constitutional change, it came as no surprise that the focus of public choice discussion was shifted away from ordinary political choices to the institutional–constitutional structure within which politics takes place.

However, the subject matter is not new. Broadly conceived, constitutional economics is an important component of a more general revival of the classical approach. It draws substantial inspiration from the encompassing theoretical perspective and the reformist attitude that were characteristic of Adam Smith's vision. Buchanan's constitutional political economy can be considered the modern-day counterpart to what Smith called 'the science of legislation', an academic enterprise that seeks to bring closer together again the economic, social, political, philosophical and legal perspectives that were once part of the study of 'moral philosophy'.

One might be tempted to characterize constitutional political economy simply – and somewhat narrowly – as 'the economic analysis of constitutional law'. It cannot be denied that the examination of real-world constitutions using the perspective of modern constitutional political economy is an interesting exercise and may provide a kind of test for the usefulness of this approach. Reference can be made to several interesting case studies.[2] However, such a definitional strategy may tend to be somewhat misleading. The use of the term 'constitutional' in the self-description of the subdiscipline is largely metaphorical. Constitutional economics as a research field comprises but is at the same time broader than, 'the economic analysis of constitutional law'.

Constitutional economics as a scientific subdiscipline is characterized by a particular kind of orientation in social analysis. Whereas orthodox economic analysis attempts to explain the choices of economic agents, their interactions with one another and the results of these interactions, within the existing legal–institutional–constitutional structure of the polity, constitutional economic analysis attempts to explain the working properties of alternative sets

of legal–institutional–constitutional rules that constrain the choices and activities of economic and political agents. The emphasis is on the rules that define the framework within which the ordinary choices of economic and political agents are made. Thus constitutional economics analysis involves a 'higher' level of inquiry than orthodox economics. It examines the choice of constraints as opposed to the choice within constraints. Thus the constitutional economist has nothing to offer by way of policy advice to political agents who act within defined rules. On the other hand, the whole exercise is aimed at offering guidance to those who participate in the discussion of constitutional change. Constitutional economics offers a potential for normative advice in constitutional matters, whereas orthodox economics offers a potential for advice to the practising politician.

A preliminary illustration may be drawn from the economics of monetary policy. Events in the European Monetary System, on the one hand, and monetary disintegration in the former Soviet Union, on the other, have revived interest in the question of how to design and choose a monetary regime for both parts of Europe that ensures monetary stability. The constitutional economist is not directly concerned with determining whether monetary ease or monetary restrictiveness is required for furthering stabilization objectives in a particular setting. However, he/she is directly concerned with evaluating the properties of alternative monetary regimes (such as complete monetary union versus currency competition).[3]

Of course, there exists a whole set of subdisciplines that all draw some attention to the legal–political constraints within which economic and political agents choose. Differences can be identified, however. Thus public choice, in its non-constitutional aspects of inquiry, concentrates attention on analyses of alternative political choice structures and on behaviour within those structures. Its focus is on predictive models of political interactions, and is a preliminary stage in the more general constitutional inquiry. Law and economics remains somewhat closer to orthodox economic theory than constitutional economics or public choice. The standard efficiency norm remains central, both as an explanatory benchmark and as a normative ideal.

One of the leading journals of the subdiscipline is *Constitutional Political Economy* (CPE). Some intuitive understanding of what constitutional political economy is all about can be gained from explaining the logic behind the logo of this journal, which is drawn from Greek mythology. The logo is a representation of the familiar Homeric account of how Ulysses heard the sirens singing, and survived (Kliemt and Brennan, 1990). Ulysses wanted to hear the exquisite voices of the sirens. He was passing close by and, in principle, there was nothing to prevent him from listening to them while continuing his journey. However, he recognized that the power of these voices was such that he would steer the ship ever closer to the rocks where

the sirens were located. The ship would be wrecked and he would be unable to continue his journey.

Formally, Ulysses faced a problem of time inconsistency in his optimal plan. His optimal plan was to listen to the sirens and then continue his journey. But this was time-inconsistent because, once he had embarked on the plan by listening to the sirens, he would not have been able to implement the later part of the plan, the rest of his journey. By contrast, a time-consistent optimal plan is one that specifies a sequence of actions (A_t, A_{t+1}, A_{t+2} and so on), one for each moment in time (T, T + 1, T + 2 and so on), which enjoys the property that the individual will actually choose in each time period the action specified by the plan. Thus, when T + 1 occurs, having undertaken A_t in T, the individual will still choose A_{t+1} as the best action rather than some other, and so on.[4]

The time inconsistency arises because the sirens affect Ulysses' preferences. His perception of the best action changes in the middle of the plan and this leads him to deviate from the original version. Ulysses implemented his optimal plan by denying himself freedom at the later stage of the plan. Having instructed his men to tie him to the mast and to ignore any orders to do anything other than sail past the rocks, he told them to plug their ears and row. Thus, Ulysses established for himself a private constitution, a set of more or less binding rules that constrain his future choices. By exploiting elements of his natural and social environment, Ulysses was able to subvert certain inclinations of his future self, inclinations that he knew would be destructive of his overall interests but which would nevertheless prove irresistible when they arose.

Though the theory of private constitution is a (small) part of the domain of constitutional political economy (Buchanan, 1990, p. 3), the principal issue for constitutional political economy is that of forming a mutually agreeable constitution for social arrangements among a community of persons. Ulysses is therefore to be seen not merely as a single actor but more particularly as representing society as a whole, and the mast and rope are to be identified as the rules by which ordered society is governed.

As Kliemt and Brennan (1990, p. 125) point out, some care must be taken in interpreting any such image. Following the individualist methodology, 'social action' must be decomposed into the actions of the individuals of whom society is made up; the exercise of social binding, specifically, must be seen as an intrinsically multilateral activity. Each agrees to a set of rules and procedures because this is the price each must pay to restrict the conduct of others. 'Weakness of the social will' will arise precisely because it is opportunistically rational for any individual to depart from the collectively agreed rules and procedures.

Moreover, in the setting with which constitutional economics is concerned, there is no external technology available that is totally effective or that is not

excessively costly. The tools of enforcement and maintenance must themselves be socially constructed. Human beings are not bound by nature to pursue rules: they are endowed with the capacity to deviate from rules if it is profitable to do so. Accordingly, we must search out rules which so order individuals' behaviour that it is individually profitable for most people to keep and enforce those rules most of the time. The gains from violation should not be too great. The analysis of the kind of rules and the associated institutional apparatus that exhibit these properties represents a centrepiece of constitutional political economy as an area of inquiry.

Theoretical foundations and intellectual origins: the Wicksellian ancestry

Constitutional economics is informed by an explicit methodological individualism (Buchanan, 1990, p. 13). Only individuals choose and act. Whatever phenomena at the social aggregate level we seek to explain, we ought to show how they result from the actions and interactions of individual human beings who, separately and jointly, pursue their interests as they see them, based on their own understanding of the world around them (Vanberg, 1994, p. 1).

Beyond the logical – and largely tautological – presuppositions of individualism, orthodox public choice models usually obtain operational content through the postulate of *homo economicus*. Individuals are assumed to be utility-maximizing and to seek their own interests. It is increasingly recognized, however, that at least a part of the traditional public choice emphasis had been wrongly placed. Thus the emphasis is shifted away from the motivational postulates for political actors to the incentive structures of politics. In Buchanan (1993a, p. 69) it is argued that the seminal Alchian (1950) analysis of the market's analogue to evolutionary selection can be extended to politics in a relatively straightforward fashion, the difference between the two evolutionary models lying in the compatibility with overall efficiency. The structure of the politics in which politicians act requires them to act contrary to the public interest if they are to survive at all. For the constitutional economist the relevant question then becomes: 'How can constitutions be designed so that politicians who seek to serve "public interest" can survive?' (Buchanan, 1993b).

The germs of the recent re-emergence of the research programme of constitutional political economy were contained in *The Calculus of Consent* (Buchanan and Tullock, 1962; also Wagner, 1988 and Tullock, 1987). The distinguishing feature of the Buchanan and Tullock approach to the study of political institutions from a normative viewpoint was to treat the political process by which individuals advance their interests as one of exchange. In adding this second element – 'politics as exchange' – to the utility-maximizing models for individual choice behaviour in politics, they were directly influenced by the great work of Knut Wicksell.

Constitutional political economy could be characterized as 'Wicksellian' political economy. In his basic work on fiscal theory, Wicksell (1896) called attention to the significance of the rules within which choices are made by political agents, and he recognized that efforts at reform must be directed towards changes in the rules for making decisions rather than towards modifying expected results through influence on the behaviour of the actors. In order to take these steps, Wicksell needed some criterion by which the possible efficacy of a proposed change in rules could be judged. He introduced the now familiar (near to) unanimity or consensus test. Thus, for Wicksell, 'the consent of the governed' was the point of departure for the evaluation of government activities. As he concluded:

> It is a necessary condition that expenditures and the means of financing them be voted upon simultaneously. … If this procedure should become general practice, a very important practical step would have been taken in the direction of the system proposed in this essay. The requirement of the veto right of the minorities would follow sooner or later as a logical and necessary consequence. … It stands to reason that a combination which satisfies everyone … must be imbued with more justice than any other which might appeal more to an accidentally greater half of those interested, but which would be at the expense of the others. Once this is conceded, the right of minority veto is already recognized in principle. (Wicksell, 1896 [1962], p. 116)

This 'Wicksellian' idea has had considerable influence on Buchanan's approach. Buchanan maintains that politics must be understood according to the model of market exchange. Thus the political process is conceptualized as one of mutually beneficial exchange. It is for this reason that he is drawn to unanimity as a collective decision rule. Since the choice among rules is more a social choice than an exchange, the form of voluntary exchange is political consent. At the most fundamental level of constitutional choice, consent serves as the basis of justification. It provides the ultimate criterion of efficiency. Unlike other economists who have emphasized either the efficiency or rationality of rules, Buchanan is concerned exclusively with whether or not people consent to them. Through the emphasis on 'consent' or 'agreement' as a normative yardstick, the research programme of constitutional political economy became closely related to the contractarian tradition in political philosophy (Buchanan, 1975). In contrast with Paretian 'optimum resource allocation', a situation of 'Wicksellian efficiency' will be characterized by the fact that citizens are satisfied that the extant system of rules, institutions and policies of their society is free from improper coercion (Wiseman, 1990, p. 110).

Thus Buchanan and traditional economic analysts develop the relationship between autonomy and efficiency in exactly opposite ways (Coleman, 1990, p. 141). Traditional economists believe that efficiency can be defined as a

property of social states independent of the process of voluntary exchange. For example, the perfectly competitive market is efficient, but the outcome of the prisoner's dilemma is not. And given the logic of the relevant concepts – especially Pareto superiority – it follows logically that people would consent to efficient rules. Consent follows from efficiency. Buchanan puts the matter exactly the opposite way. What people consent to is efficient. Efficiency follows from consent.

As Buchanan sees it, contractarian political institutions typically exhibit three attributes. First, the place of the individual is central to the contractarian vision of the political process. Individuals' own – and necessarily subjective – evaluations, their interests and values constitute the relevant benchmark or criterion against which the efficiency or desirability of alternative sets of rules are to be judged. Contractarianism complies with this criterion by according each individual equal treatment at the constitutional stage. This normative individualism should be distinguished from the methodological individualism discussed above.

Second, there is the fundamental distinction between actions taken within the constitutional rules, and changes in the rules themselves. The latter are to occur only at the constitutional stage and ideally are made using the unanimity rule. Whereas Wicksell did not move beyond the development of criteria for evaluating policy alternatives one at a time, Buchanan and Tullock (1962) operationalized Wicksell's (1896) insights and extended the applicability of the unanimity or consensus criterion from the level of particular proposals to the level of rules – to constitutional rather than post-constitutional or in-period choices. The image of political activity as a two-stage process, first developed in *The Calculus of Consent*, recurred in many of Buchanan's later writings as a sort of normative benchmark or yardstick by which to measure the quality of a community's political institutions.

Third, actions taken in the second stage of the political process should be effectively constrained by the rules written in the first, constitutional stage, and this is true, not only for the individual citizen, but also for the elected representatives, and the bureaucrats and jurists who administer the system.

The shift of the Wicksellian criterion to the constitutional stage of choice has some remarkable consequences. It becomes conceivable to allow for the possibility that preferred and agreed decision rules might embody sizeable departures from the unanimity limit, including simple majority voting in some cases and even less than majority voting in others (Buchanan, 1987, p. 135). The constitutional calculus suggests that both the costs of reaching decisions under different rules and the importance of the decisions are relevant. Since both of these elements vary, the preferred rule will not be uniform over all ranges of potential political action. The in-period Wicksellian criterion may remain valid as a measure of the particularized efficiency of the

single decision examined. But the in-period violation of the criterion does not imply the inefficiency of the rule so long as the latter is itself selected by a constitutional rule of unanimity.

For Buchanan and Tullock (1962, ch. 6) constitutional design was a matter of comparing the interdependence costs of public and private decisions over a range of activities to determine which activities would be assigned by the constitution to the state and which voting rule or choice mechanism would be specified by the constitution for each state activity. The best public decision rule for each activity was the one that minimized interdependence costs. It was specified that the representative individual perceived interdependence costs for an activity as the sum of the anticipated external costs levied on that individual if not part of the decision set, and the anticipated decision cost experienced by the individual if part of the decision set. The sum of both external and decision costs was shown to have a unique minimum somewhere between the extremes of individual rule and unanimity rule, the exact position depending on relative external and decision costs.

Thus, while it was recognized that unanimity and not majority rule is the pivot of constitutional democracy, it was equally demonstrated that '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' (ibid., p. 96).

The general problem of efficient constitution formation and maintenance

The choice situation at the constitutional as well as the post-constitutional stage is ususally modelled as a classic prisoner's dilemma (Figure 13.1), at least in so far as it involves potential conflict of interests between rational persons (Gwartney and Wagner, 1988a, p. 32; Buchanan, 1993b, p. 2). In 'generalized prisoner's dilemma situations', that is, social constellations under which individuals, in separate and rational pursuit of their own interests, unintentionally but system-

		B	
		C	D
A	C	3,3	1,4
	D	4,1	2,2

Figure 13.1 Classic prisoner's dilemma

atically contribute to an overall outcome that is undesirable for all of them (or in any case less desirable than some alternative outcome that could be realized by concerted, organized action) there may exist a potential for mutual gains by collective action (collective organization).

Thus the constitution is essentially a contract intended to secure the mutual gains from social cooperation and to avoid the dominant defective strategy in the prisoner's dilemma game which leads to a socially inefficient Nash equilibrium solution. Since the mutual gains from social cooperation constitute a public good, the maintenance of the constitutional contract gives rise to a problem that will not resolve itself naturally.

Even when it is supposed that agreement on appropriate rules can be achieved at the stage of constitutional contract formation, it should be recognized that individuals and interest groups inevitably will attempt to engage in post-contractual opportunism (the problem of constitutional maintenance). Thus the agreement, once achieved, must be enforceable. This opportunism takes several forms. First, each individual may have an incentive subsequently to defect from the cooperative agreement (the compliance or unilateral defection problem). Whether or not it is rational for persons to comply with rules that they constitutionally may agree on is a matter of contingent, factual circumstances. It depends on whether or not the constraints that persons face after the agreement, that is post-constitutionally, make it rational for them to comply with previously agreed rules.

A second form of post-contractual opportunism consists of rent seeking and special-interest plundering which ultimately reduce the value of post-contractual cooperation and undermine the constitution itself. Groups of individuals have an incentive to seek to capture the instruments of state power and to use them as vehicles to enrich themselves in ways that are not possible for private citizens. 'Rent seeking' is a term used by economists to describe actions taken by individuals and groups to alter public policy in order to gain personal advantage at the expense of others. The social costs entailed by this process are called 'rent-seeking costs' or, by some, 'Tullock costs', after Tullock (1967).[5] Tullock showed not only that the inefficiency or social welfare cost of, say, a tariff consists of the *Harberger triangle* and can increase with the *Tullock rectangle*, but also that the pure transfer involved in the creation of tariffs or other privileges will lead market participants to expend resources in lobbying and political activities:

> These expenditures, which may simply offset each other to some extent, are purely wasteful from the standpoint of society as a whole; they are spent not in increasing wealth, but in attempts to transfer or resist transfer of wealth. I can suggest no way of measuring these expenditures, but the potential returns are large, and it would be quite surprising if the investment was not also sizable. (ibid., p. 228)

The incentive to engage in rent-seeking activities is directly proportional to the ease with which the political process can be used for personal (or interest group) gain at the expense of others. In other words, distributional politics is viable and tends to become dominant to the extent that differential treatment is constitutionally permissible (Buchanan, 1993b, p. 6).

Tullock (1959) had shown that under any voting system which requires less than unanimous approval to implement policies, majority coalitions of interest groups will seek to obtain public provision of special interest projects. The dominant strategy for any organized interest group in a majoritarian polity is to lobby for policies which provide large benefits to its members and spread the costs among everyone else. This tendency exists even in liberal democracies. Through implicit vote trading, a coalition of interest groups, comprising a bare majority of voters, can get all or at least most of their favoured projects approved for public provision. Under certain conditions, the total costs of these projects can exceed their total benefits, while cost-spreading through the 'fisc' induces a rational ignorance of this process on the part of the disadvantaged majority. On the other hand, the asymmetric distribution of cooperative benefits leads subgroups of the collective to invest energy struggling for access to the government's coercive power. But the effort may turn out to cost more than it is worth and the end result will be that the collective's loss purchases the subgroup's gain (Schmidtz, 1991, p. 91).

Buchanan and Lee (1991) demonstrate that the gains from politically generated restrictions on markets, even to organized producing interests, are more apparent than real. The analysis demonstrates that, under plausibly realistic assumptions concerning coalitions sizes, excess burdens, organizational costs and rent-seeking outlay, a genuine utility-maximizing calculus may dictate support for constitutional prohibition of all market restrictions, by all members of the polity, including those producer interests that might be considered to be the potentially identifiable beneficiaries of cartelization.

Principal–agent theory has been used to examine the rent-seeking problem (Anderson and Hill, 1986; Merville and Osborne, 1990).[6] The principal, also the citizen, grants the agent (the government) the power of coercion. In exchange, the agent supplies the principal with public goods. Since the capitalized value of public assets is owned collectively, public good outputs of the government are like communal resources with widely diffused benefits. It soon becomes evident to vote-maximizing agents or legislators that they can maximize their political support by significantly reducing the provision of public goods to the population at large in favour of greater transfers to interest groups. These transfers are financed by general tax collections and provide concentrated benefits to designated groups. Such collusion between agents and special interest groups will invariably lead to a breaking of the constitutional contract.

Merville and Osborne (1990) use agency theory to demonstrate formally that, in majority-rule political systems, coalitions of minority factions will induce politicians to break the constitutional contract systematically in order to supply special-interest projects. Unlike contracts in private markets, political contracts are much more susceptible to this kind of opportunism.

Proposed solutions

Is the rent-seeking trap inescapable? By far the most important problem with respect to ensuring the self-enforcing character of a constitutional contract is that it must successfully constrain the power of the state itself.

Substantive restraints versus procedural rules

Generally speaking, substantive constraints on government have been dismissed as ineffective precisely because of the wide latitude they allow for reinterpretation. Gwartney and Wagner (1988a, pp. 44–9) make a strong case for procedural rules designed to uphold decentralization of governmental powers and to prevent the formation of legislative coalitions. Procedural rules will provide more effective mechanisms for self-enforcement than will substantive restraints on government. In their view, the weakness of substantive restraints derives from the politicization of the Supreme Court and the ease with which legislatures can find alternative ways to implement any given policy. They propose procedural rules requiring larger legislative majorities for legislative action at higher levels of government, thereby diffusing the power of the state to regional and local governments.

Judicial independence

Does independence of the judiciary serve the long-term public good? The traditional view of the purposes of judicial independence has been attacked as naive by law and economics and public choice scholars. Unlike many legal contracts, it is argued, there is no third-party enforcer, external to the contract, who can ensure that defectors are caught and forced to comply with the terms of the agreement. Although many countries have a nominally independent Supreme Court whose purpose is to enforce the constitution, the Supreme Court can only do this imperfectly in most cases, because the judges themselves are not totally immune from political pressures by groups wishing to subvert the original intent of the constitution. Thus, given the unreliability of third-party enforcement, and given the strong individual incentives to defect from social cooperation, the constitutional contract should somehow be self-enforcing if it is to be maintained.

The interest group theory first advanced by Landes and Posner (1975) makes the independent judiciary an integral part of the system of rent seeking engineered by Congress. However, the debate goes on. A very detailed criti-

cism of the Landes–Posner theory is contained in Boudreaux and Pritchard (1994), who argue that the theory is seriously deficient and conclude that the United States federal judiciary is truly independent of Congress and the president, and that this independence was designed by the US Constitution's framers as a means of furthering sound government.

A rule of law in politics

According to Buchanan (1993b) the direction of constitutional reform is obvious. If, somehow, the potential for differential treatment is reduced, so will be the inducement to rent-seeking behaviour. The off-diagonal solutions should simply be made impossible to achieve by the introduction of some rule or norm that prevents participants from acting or being acted upon differently, one from the other. If the off-diagonal attractors are eliminated, then the players operate with the reduced matrix shown in Figure 13.2. Thus the constitutional reform measure modifies the original prisoner's dilemma game into a reduced setting in which each player, as a member of a political coalition, knows that any choice of an action or strategy must involve the same treatment of all players or constituencies (ibid., p. 3).

		B	
		C	D
	C	3,3	X
A			
	D	X	2,2

Figure 13.2 Modified prisoner's dilemma

If and to the extent that differential treatment is replaced by equal treatment, or by the principle of generality in politics – analogous to that present in an idealized version of the rule of law – mutual exploitation will be avoided and politicians who seek to serve the 'public interest' will survive and prosper (ibid., p. 6). Thus it seems at least conceivable that rational persons, at the stage of entering into the agreement, may recognize the 'rent-seeking trap' and engage in concerted effort to escape.

However, in the hypothetical matrix construction above, the interaction was in fact assumed to occur in a state of nature, with each person holding equal prospects for membership in the majority and minority coalitions. This means that membership was assumed to be symmetrical among all partici-

pants. But this assumption may turn out to be too heroic with respect to real-world settings.

The prospects may differ among persons and groups of persons so as to create divergences in interests which may become a source of disagreement. Thus the question remains whether it is possible to modify the constitutional choice setting so as to reconcile such possible divergences. It appears that, at least from the perspective of potentially-conflicting interests among constituencies, the general problem of constitutional efficiency and survivability does not resolve itself naturally.

Veil of uncertainty and/or ignorance versus the availability of exit options

Is it possible to specify the conditons under which constitutional agreement may be facilitated in real, non-hypothetical choice situations? Is it possible to modify the constitutional choice setting so as to reconcile divergences in interests? In this respect, two lines of reasoning have been pursued in the contractarian and neo-contractarian literature. The first line of argument focuses attention on the need for a 'veil of uncertainty and/or ignorance' as a precondition for an efficient constitution.

Buchanan and Tullock (1962) had to present a convincing positive argument that unanimous consent at the constitutional level was possible at all. How can agreement on rules among persons with potentially conflicting constitutional interests be achieved? The authors' characteristic way of approaching this issue consists of emphasizing the uncertainty confronting all individuals taking part in constitutional deliberations. The existence of 'a veil of uncertainty' induces individual participants in a constitutional process to prefer rules that do not systematically favour any particular subset of citizens.

The proposed remedy involves the introduction of some means of ensuring people's inability reliably to foresee their future particularized interests, as these may be affected by different rules, thereby inducing people to make constitutional choices on some assessment of the general working properties of alternative rules, and divorced from particularized interests. Thus agreement is facilitated by whatever increases people's uncertainty about the particular effects that alternative rules can be expected to have on them. In fact the assumption of a 'veil of uncertainty' was also hidden in Buchanan (1993b), discussed above.

Buchanan's approach has affinities with John Rawls's (1971) construction, which utilizes the veil of ignorance along with the fairness criterion to derive principles of justice that emerge from a conceptual agreement at a stage prior to the selection of a political constitution. Thus in Rawls's construction, the prospect of agreement is secured by defining certain 'ideal' conditions under which constitutional choices are hypothetically made. The choosers are assumed to be placed behind a 'veil of ignorance' which makes it impossible

for them to know anything specific about how they will be personally affected by alternative rules. Ignorant about their prospective specific interests in particular outcomes, they are induced to judge rules 'impartially'. Potential conflict in constitutional interests is not eliminated, but the veil of ignorance transforms potential interpersonal conflicts into intrapersonal ones (Vanberg, 1994, p. 170).

However, the constitutionalist notion of a veil of uncertainty or ignorance, though useful as an analytical benchmark, is not very practical. It is not clear how genuine uncertainty or ignorance could be achieved in real-world constitution formation. Therefore, it has been argued that the availability of exit options can ensure a competitive setting for participants in constitutional deliberations and can even substitute for a veil of uncertainty. This condition for efficiency can be given operational substance in processes of real-world constitution formation (Lowenberg and Yu, 1992).

In order to produce an efficient social contract or constitution, deliberations must be carried out in a competitive 'constitutional environment'. This condition will be satisfied if an exit option exists for each contracting party. This conclusion is quite consistent with the Wicksell–Buchanan–Vanberg contractarian consensus test. Only in a competitive setting does unanimous agreement acquire operational substance (normative content).

Vanberg (1994) clearly recognizes that the true problem with the agreement criterion is not that it is too demanding but, rather, that it has too little normative content. A criterion needs to be specified which allows one to distinguish between constraints that are judged to make the respective individual choices involuntary, and those that do not. Vanberg's analysis reaches the conclusion that a consistent normative–individualist approach needs to rely on a combined and simultaneous application of a purely procedural, rule-oriented, as well as a substantive, avoidance/exit cost criterion. The avoidance/exit cost perspective arguably provides a more operational specification of the contractarian norm than the notion of a hypothetical contract to which Buchanan (1975, 1977) as well as Rawls (1971) appeal.

The notion of exit has thus been invoked to give more operational substance to the concept of voluntary agreement. It is derived from Albert Hirschman's (1970) classic distinction between exit and voice. Exit (and entry) is an important means by which individuals are able to express their preferences, and is precisely the method through which preferences are revealed in competitive markets for private goods. An exit option introduces an element of market-like competition into the contracting process, which limits the ability of any party to wield power over another party. It is not even necessary that this exit option be exercised, since merely the threat of its use should be enough to restrain rent appropriation. The scope for opportunism is effectively constrained by competition, actual or potential.

Furthermore, it is argued that exit options can help to solve the constitutional maintenance problem by establishing a competitive environment for post-constitutional political and market exchange (Lowenberg and Yu, 1992).

Federalism, once again

The strengthening of regional and local government relative to national government has been advocated by many scholars as an effective way to restrain the growth of legislative redistribution. The existence of separate jurisdictions with some protected powers within a constitutional federation inhibits coercive behaviour by the government. Such an arrangement facilitates migration at low cost between federal subregions and thereby enhances competition between these subregions. The resulting mobility forces competitive governmental units to supply public goods in preferred quantities and to 'price' them broadly in line with relative marginal evaluations.

The foregoing is related to the Tiebout effect (Tiebout, 1956), which says that individuals will sort themselves across communities in accordance with their preferences for the packages of taxes and public goods provided in each community. The ability of the owners of property rights to move to competing jurisdictions protects them from potential rent appropriation by a coercive government. Therefore, it is argued, a federalist constitution can effectively constrain the power of the state. In a federal system, citizens seeking political relief can vote with their feet.

The preceding paragraphs suggest that post-contractual exit opportunities might be characterized in terms of Tiebout competition between different political groupings. If the constitution permits mobility and political plurality, it will help establish and maintain a competitive political post-constitutional environment.[7]

Notes

1. A classic overview of public choice theory is contained in Mueller (1989).
2. Thus Backhaus (1995) contains an analysis of constitutional guarantees of basic rights and procedures, illustrated by three constitutions, the American Constitution of 1789 as amended in 1792, the German Basic Law of 1949 and the Dutch Basic Law of 1983. In addition, reference can be made to several case studies. Holcombe (1991) analyses the role of constitutional rules as constraints on government using three US constitutions: the Articles of Confederation (1781), the Constitution of the United States and the Confederate Constitution. Geoffrey Brennan and Jose Casas Pardo (1991) examine the Spanish constitution (1978). Sobel (1994) analyses the evolution of two international constitutions: the League of Nations Covenant and the United Nations Charter.
3. See, for example, Hefeker (1995); for some general reflections, see Eichengreen (1994).
4. The problem of time inconsistency has perhaps most notably been investigated in the context of central bank monetary policy; see in this connection the contributions of Barro and Gordon (1983a and 1983b). For a survey of subsequent elaborations and variations upon the same theme, see, for example, Walsh (2001, ch. 8, pp. 321–84).
5. In *The Power to Tax*, Brennan and Buchanan argue that the existence of potentially huge rent-seeking costs constitutes one of the important arguments for predicting that all rational

individuals, behind a veil of ignorance, would seek to constrain exploitation by revenue-maximizing government to the maximum possible extent. The only way of doing so is to minimize the rents that accrue from 'governing' – that is, by constraining Leviathan so that its surplus is minimal. Government 'surplus', or the income that accrues to government for discretionary use, is defined as $S = R - G$, that is, the excess of revenue collections over spending on specified uses. Since $G = \alpha R$, $S = (1 - \alpha)R$, where α is the proportion of total revenues to be spent on specified public goods and services (Brennan and Buchanan, 1980 [2000], ch. 2).

6. For what may now well become a standard general and formal treatment of the principal–agent model, see Laffont and Martimort (2002).
7. On the significance of the substitutability between intergovernmental competition for fiscal resources and explicit constitutional constraints on governmental taxing power, once the possibility of federalization is introduced, see Brennan and Buchanan, 1980 [2000], ch. 9. These authors' emphasis is on federal assignment as a means of ensuring that individuals have available options as among the separate taxing-spending jurisdictions, and on the effect that the potential exercise of these options has on the total fiscal exploitation in the system. Total government intrusion into the economy should be smaller, *ceteris paribus*, the greater the extent to which taxes and expenditures are decentralized, the more homogeneous are the separate units, the smaller the jurisdictions, and the lower the net locational rents. (ibid., p. 216).

References

Alchian, A.A. (1950), 'Uncertainty, evolution and economic theory', *Journal of Political Economy*, **58** (3), 211–21, reprinted in A.A. Alchian, *Economic Forces at Work*, Indianapolis: Liberty Press, 1977, pp. 15–35.

Anderson, T.L. and P.J. Hill (1986), 'Constraining the transfer society: constitutional and moral dimensions', *Cato Journal*, **6**, 317–41.

Backhaus, J.G. (1995), 'An economic analysis of constitutional law', research memorandum, University of Limburg, Maastricht.

Barro, R.J. and D.B. Gordon (1983a), 'A positive theory of monetary policy in a natural rate model', *Journal of Political Economy*, **91** (4), 589–610.

Barro, R.J. and D.B. Gordon (1983b), 'Rules, discretion and reputation in a model of monetary policy', *Journal of Monetary Economics*, **12**, 101–21.

Boudreaux, D.J. and A.C. Pritchard (1994), 'Reassessing the role of the independent judiciary in enforcing interest-group bargains', *Constitutional Political Economy*, **5** (1), 1–21.

Brennan, G. and J.M. Buchanan (1980 [2000]), *The Power to Tax – Analytical Foundations of a Fiscal Constitution*, Indianapolis: Liberty Fund.

Brennan, G. and J. Casas Pardo (1991), 'A reading of the Spanish constitution (1978)', *Constitutional Political Economy*, **2** (1), 53–79.

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

Buchanan, J.M. (1977), *Freedom in Constitutional Contract*, College Station, TX and London: Texas A&M University Press.

Buchanan, J.M. (1987), 'Justification of the compound republic: the calculus in retrospect', *Cato Journal*, **7** (2), pp. 305–12, reprinted in Gwartney and Wagner (eds) (1988b), pp. 131–7.

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

Buchanan, J.M. (1993a), 'Public choice after socialism', *Public Choice*, **77**, 67–74.

Buchanan, J.M., (1993b), 'How can constitutions be designed so that politicians who seek to serve "public interest" can survive?', *Constitutional Political Economy*, **4** (1), 1–6.

Buchanan, J.M. and D.R. Lee (1991), 'Cartels, coalitions and constitutional politics', *Constitutional Political Economy*, **2** (2), 139–61.

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

Coleman, J. (1990), 'Constitutional contractarianism', *Constitutional Political Economy*, **1** (2), 135–48.

Eichengreen, B. (1994), *International Monetary Arrangements for the 21st Century*, Washington, DC: Brookings Institution.
Gwartney, J.D. and R.E. Wagner (1988a), 'Public choice and constitutional order', in Gwartney and Wagner (eds) (1988b), pp. 29–56.
Gwartney, J.D. and R.E. Wagner (eds) (1988b), *Public Choice and Constitutional Economics*, London: JAI Press.
Hefeker, C. (1995), 'Monetary union or currency competition? Currency arrangements for monetary stability in East and West', *Constitutional Political Economy*, **6** (1), 57–69.
Hirschman, A.O. (1970), *Exit, Voice and Loyalty*, Cambridge, MA: Harvard University Press.
Holcombe, R.G. (1991), 'Constitutions as constraints: a case study of three American constitutions', *Constitutional Political Economy*, **2** (3), 303–28.
Kliemt, H. and G. Brennan (1990), 'Logo logic', *Constitutional Political Economy*, **1** (1), 125–7.
Laffont, J.-J. and D. Martimort (2002), *The Theory of Incentives – The Principal–Agent Model*, Princeton, NJ and Oxford: Princeton University Press.
Landes, W.M. and R.A. Posner (1975), 'The independent judiciary from an interest group perspective', *Journal of Law and Economics*, **18**, 875–901.
Lowenberg, A.D. and B.T. Yu (1992), 'Efficient constitution formation and maintenance: the role of "exit"', *Constitutional Political Economy*, **3** (1), 51–72.
Merville, L.J. and D.K. Osborne (1990), 'Constitutional democracy and the theory of agency', *Constitutional Political Economy*, **1** (3), 21–47.
Mueller, D.C. (1989), *Public Choice II*, Cambridge: Cambridge University Press.
Rawls, John (1971), *A Theory of Justice*, Cambridge, MA: Belknap Press.
Schmidtz, D. (1991), *The Limits of Government*, Boulder, CO, San Francisco, CA and Oxford: Westview Press.
Sobel, R.S. (1994), 'The League of Nations Covenant and the United Nations Charter: an analysis of two international constitutions', *Constitutional Political Economy*, **5** (2), 173–92.
Tiebout, C.M. (1956), 'A pure theory of local expenditures', *Journal of Political Economy*, **64**, 416–24; reprinted in T. Cowen (ed.), *Public Goods and Market Failures*, Chapter 9, London: Transaction, 1992, pp. 416–24.
Tullock, G. (1959), 'Problems of majority voting', *Journal of Political Economy*, **67**, 571–9.
Tullock, G. (1967), 'The welfare costs of tariffs, monopolies and theft', *Western Economic Journal*, **5**, 224–32.
Tullock, G. (1987), 'The calculus: postscript after 25 years', *Cato Journal*, **7** (2); reprinted in Gwartney and Wagner (eds), 1988b, pp. 139–47.
Vanberg, V.J. (1994), *Rules and Choice in Economics*, London: Routledge.
Wagner, R.E. (1988), 'The calculus of consent: a Wicksellian retrospective', *Public Choice*, **56**, 153–66.
Walsh, C.E. (2001), *Monetary Theory and Policy*, Cambridge, MA: The MIT Press.
Wicksell, K. (1896), *Finanztheoretische Untersuchungen* [Investigations into the Theory of Finance], Jena: Gustav Fischer.
Wicksell, K. (1896 [1962]), 'A new principle of just taxation', reprinted in R.A. Musgrave and A.T. Peacock (1962), *Classics in the Theory of Public Finance*, London: Macmillan, pp. 72–118.
Wiseman, J. (1990), 'Principles of political economy: an outline proposal, illustrated by application to fiscal federalism', *Constitutional Political Economy*, **1** (1), 101–24.

14 Administrative law and economics

Jean-Michel Josselin and Alain Marciano

The domain of administrative law and economics

The use of economics to understand administrative law may not be as widespread as in other areas of legal doctrine and practice. Public behaviours are none the less unambiguously susceptible to economic investigation. The objective here is to provide some general guidance as to how political economy can be used to understand the legal dimension of the state. In this respect, the domain of administrative law and economics consists of two related approaches.

The first one deals with both the efficiency and the control of administration, in a given constitutional framework. Two levels of objectives can be set therein. On the one hand, coherence of administrative behaviours and actions must be assured with regard to the goals of the state and the protection of private rights. The prominent feature is rent seeking. On the other hand, internal control is necessary at the level of the administrative agencies themselves. Bureaucratic behaviours must be contained by proper incentive mechanisms. This public choice perspective amounts to an evaluation of the outcomes of the behaviour of the administration, and can be summarized by the question: how to judge the actions of the state?

In this chapter, we shall not deal with this issue which, as is exemplified by the growing public choice literature dedicated to it, clearly requires a distinctive (notwithstanding inevitable overlap) treatment. This entry rather belongs to the second type of approach which considers the general constitutional framework as endogenous (Buchanan, 1990; Cooter, 2000). From this constitutional political economy perspective, the focus is on the procedures leading to the outcomes that are analysed by public choice theory. Therefore, the discussion will not address problems of rent seeking as such but only in so far as they are embodied in a constitutional framework. The previous question now broadens to: how to judge the state?

In this respect, the principal–agent problem proves insightful. The principal would delegate the right to judge the state. The agent would be induced to take the best actions, in order to fulfil his/her commitment. We shall first focus on two basic questions: who is the principal and who is the agent? We shall then consider the nature of the incentive mechanism. The various institutional answers to these questions provide the many ways of dealing with the power of the state.

The subsequent discussion will focus on the following problems. First, strategic behaviours of the state will be expressed as a paradox of government, administrative law being its translation in legal terms. Second, the right to judge the state will be analysed as an agency problem and the discussion will focus on the nature of the principal and the agent. Third, the debate will turn on the competition for principalship and on the way in which the agent is controlled. Finally, concluding comments will discuss the role of the constitution as a framework for that agency relation.

Strategic behaviours and the paradox of government

The political economy of rent seeking emphasizes how public bodies use their power to create monopoly prerogatives. This phenomenon can be studied as a phenomenon in itself. It is also a consequence of the juridical and constitutional system which will either be conducive to rent seeking or restrain it. In this larger perspective, strategic behaviours of the state become a paradox of government (Witt, 1992; Josselin and Marciano, 2000). It is paradoxical in that granting some power to a public entity simultaneously provides the latter with the right and means of infringing it. Therefore, the paradox of government is both an expression of rent seeking and something larger. In other words, it expresses how rent seeking is constructed and takes place in an endogenous constitutional setting.

Administrative law is the juridical translation of the paradox of government, the legal consequence of the existence of the state; it guarantees the necessary development of public actions (as defined by the constitution or by any charter) and their necessary containment as well. The way in which the state is judged indicates what its nature is meant or supposed to be. The specificity of administrative law can then be asserted only if the state deserves or demonstrates some specificity in its turn. From this perspective and despite a noticeable trend towards homogenization and cross-fertilization, two main traditions can be broadly distinguished (Josselin and Marciano, 1995). It remains common practice to separate these traditions on the basis of the duality or unity of jurisdiction. The first one is mostly substantive and embraces 'continental' systems such as the German *Verwaltungsrecht* or the French *Droit Administratif*. In these dualist systems, the state is viewed as a political body with distinctive aims and purposes. Litigation structures thus embody a separate administrative justice. The second one is mainly procedural and includes the English and Canadian conceptions. Unity of jurisdiction is the expression of a monist conception of society. A monist system grants no specific characteristics to the state. Monism thus emphasizes that no differences should be made between private and public matters. Consequently, there is no distinct view of public actions, which are to be judged in the same way as private actions. Administrative tribunals are not separate jurisdictions.

This distinction between the unity and duality of jurisdictions may prove to be very useful when we consider the agency problem associated with the control of the state. The two systems differ with respect to the means to be used to reach a shared goal, namely to control the state. While monism considers that administrative law is designed with the purpose of limiting the powers of government in order to protect the citizen, dualism assumes that administrative law is directed towards the functioning of the administration. Such a distinction influences the nature of the principal–agent relationship.

We now move on to the first component of this agency problem, that is to say, the nature of the principal and the agent.

Who is the principal? Who is the agent?

The right which is vested in the agent is that of judging the state. Who is to delegate this right? In dualist systems, because the control of the administration is at stake, the state itself is the principal. Litigation relating to *ultra vires* actions is the specific domain of administrative courts. This particular nature of the principal can be exemplified by the French Revolution. The Constituent Assembly voted two laws (in August and September 1790) establishing that no political or institutional body can judge the state, except the state itself. The judiciary deals with subjective litigation (relating to contractual or civil liabilities) brought before civil courts. But whatever concerns the very nature of the state is controlled by litigation structures close to the public power. The French jurist and economist Jean Bodin (1577), for instance, provides detailed arguments as to why the sovereign power should deserve such treatment. The French example was followed in the 1860s by the southern German states, the concept of the administrative act (*Verwaltungsakt*) being the centrepiece of their approach. Administrative discretion is then explicitly circumscribed by *quant au fond* rules (Backhaus, 2002).

In contrast, a monist system insists, in the first place, on the necessity to protect citizens who indeed are the source of the rules of law – the latter is the unintended result of the interactions between individuals. In this case, the rationale for the existence of the state is not constitutional but, rather, conventional or customary. This means that not only does the state have no capacity to impose itself as the principal but also that the principal appears to be the impartial spectator of the common law, in the sense of Hume (1992) or Smith (1976). This conception adopts different forms. Countries like England use procedural constraints which are imposed on the state, under the rule of law. The northern German states of the 1860s, notably Prussia, were much influenced by the British legal system, while in the United States citizens are largely protected against abuse of power by the state by the Bill of Rights.

Different principals may not always have different agents. However, in the case which is considered here, these agents differ substantially. The ‘dualist’

principal creates an agent, the administrative judge, who is a judge *in* the administration. He/she addresses the objective legality of administrative decisions and tries to balance public interest and the protection of citizens. Thus, the discretionary power of the state is controlled and limited by the state itself. This may seem a peculiar way to solve the paradox of government. However, substantive norms, which are a priori defined, delineate the set of possible actions of the agent. These interpretative schemes provide the Kelsenian framework (Kelsen, 1962) within which administrative courts judge and thereby administer. As far as the 'monist' agent is concerned, ordinary jurisdictions provide ordinary judges, whose set of possible actions is delimited by procedural fairness and *stare decisis*. Although in practice they are becoming increasingly specialized, these judges none the less belong to the common law judiciary. Some countries may of course have more intricate institutional arrangements.

Other countries have a quite different view of what is covered by administrative law. Such is the case in the United States, where regulation is at the core (Rose-Ackerman, 1992). The aim is an efficient control of economic activities, particularly but not only with regard to public utilities. In principle, separation of power implies that the judiciary should control the legislative and executive powers. In practice, numerous independent and specialized agencies give rise to many agents. Congress grants them a quasi-jurisdictional status, but at the same time they may endorse decision-making systems that are similar to common law procedures. Thus the legislator and the impartial spectator may have to share the principalship.

Having defined the prominent characters of this agency problem, we shall now briefly outline the incentive mechanisms and the competition for control of the agency, with reference to the founding cases of France and England.

Incentive mechanisms and competition for principalship

Cardinal Richelieu imposed the Edict of Saint-Germain-en-Laye in 1641; Napoleon Bonaparte created the *Conseil d'État* (Council of State) in 1800. Between these two dates, the French Revolution and the 1790 laws are not historical accidents. Alexis de Tocqueville insisted that, on the contrary, there is a striking continuity. We may interpret it here in the economic terms of competition for principalship and rent-seeking opportunities. We shall also show that monist countries are confronted by this competition, although in different terms.

Control of power requires direct influence on administrative litigation structures. Since the thirteenth century, the French monarchy had been trying to avoid the supervision of ordinary tribunals and regional parliaments. To this end, it created specialized jurisdictions, not unlike the current US agencies, and a King's council. These agencies were sold to *intendants*, the power of

which was strengthened by the Edict of Saint-Germain-en-Laye and by subsequent decrees in October 1656 and July 1661. Their avowed aim was to exclude parliament from the structures of control. From a microeconomic point of view, this relation between the *intendants* and the king was rather informal; the contractual link was loose. Since offices were both sold and grounded on loyalty, moral hazard in the contract was reciprocal. On the one hand, there was no effective compatibility constraint, which was a clear disadvantage to the king. On the other hand, magistrates could be dismissed at any time. However, this simple threat could not in practice prevent rent seeking and shirking by magistrates, and some contemporary economists rejected this solution. For example, Jean Bodin is explicit about the mercantile threat that would endanger political life. The proposed alternative solution consisted in bringing judges under the direct and explicit authority of the king, and the revolutionary laws endorsed this process. Napoleon created the Council of State as a dependent agent and this situation prevailed until the end of the nineteenth century when the functioning of the council was modified. The 1872 law acknowledged the necessity of a competition between two principals, namely the government and parliament. The process was completed in 1889 when the Cadot decision increased the professional specificity of administrative justice. The designation of the members of the Council of State was no longer the prerogative of one or other of the two principals.

The lack of competition for office largely explains the inefficiency of the pre-Revolutionary system. Moreover, French parliaments were regional and they retained less power than the English parliament did. The privilege of the ruler was all the more likely to be preserved, and this remained the case after the Revolution. The situation in England was quite different: competition for rents took the form of a duopoly, consisting of the king and parliament (Ekelund and Tollison, 1980). Since the state had no special prerogatives in itself, the problem was not that of controlling it; rather, the two protagonists competed in order to obtain as many rents as possible. The outcome of this mercantile competition that had begun in the Middle Ages shows that parliament was the winner, and monism and unity of jurisdiction is a consequence of this victory. Ordinary courts of justice deal with every kind of litigation, either private or quasi-private. For instance, the dissolution of the king's Privy Council (and subsequently of the court of the Star Chamber) in 1642 gave most power to common law courts, particularly to the King's Bench, which is a *civil* court of justice. In many respects, the pre-eminence of parliament will improve the agency contract. Prior to its domination, judges in charge of the administrative litigation were 'justices of the peace', who were not paid. They were also placed under the control of the king's Privy Council. Their participation constraint could not but distort their decisions, which were often driven by rent seeking. Their incentive compatibility con-

straint was quite inessential, because of the king's supervision. Conversely, parliament tried to introduce financial incentives so as at least to bind the participation constraint and at best to contain strategic behaviours within the incentive compatibility constraint.

One of the dimensions of competition for power is thus the status of administrative law. We have shown how it resolves into an agency problem. How is this problem embodied in the constitutional framework?

The constitutional framework

Constitutions require social contracts. A contract shapes the state (Vanberg, 1994) and consequently delineates its legal characteristics, which is why administrative law is mostly a product of the constitution, if the latter exists. In countries with a common law tradition, constitutional conventions usually do not confer any legal status on the sovereign power. When the state is judged, the impartial spectator is the principal, and he/she adopts the same procedures as he/she does with private individuals. In contrast, countries with a social contract tradition must provide a tight constitutional link. Since usually the principal is the state itself, there must be a strict guarantee by the constitution, so as to protect citizens and to ensure common interests rather than rent-seeking activities. It is a necessary link that shapes the objective function of the principal: what he/she maximizes is under the control of the constitution. This constitutional setting is essentially Kelsenian and it has been adopted by Germany and Austria, among others. Since the Second World War, especially among European nations, there has been a continuous trend towards the constitutionalization of administrative law. For example, in the Netherlands, the constitutional revision of 1983 introduced a provision (article 107.2) according to which general principles of administration must be laid down by statute. The case for this trend is further outlined by many other countries. In Italy, the 1947 constitution stipulates the liability of civil servants (article 28), control of public firms (article 43) and the organization of legal actions against the administration (article 113). The Spanish constitution (1978) introduces fundamental rights of citizens (articles 10 to 55), although the Crown prerogatives remain extensive. The statutory nature of the Council of State is provided by article 107. In Greece, article 20.2 of the 1975 constitution declares the right to a hearing in the case of a dispute with the administration.

The case of the United States is also significant. Since the 1930s, the ongoing debate on the constitutionality of the 'independent agencies' exemplifies the competition for principalship between Congress and the president (Bermann and Lindseth, 1995, pp. 525–7). Whatever the outcome, the increasing number of agencies and their progressive emancipation may amount to a continuous renegotiation of the initial constitutional contract, but without

the veil of ignorance that could at least partly prevent rent seeking. From this perspective, the European Union takes the middle ground. Whereas the French tradition largely inspired the early juridical construction, it has largely evolved from this continental approach towards the Anglo-Saxon regulatory precepts. Whatever the outcome, what is at stake is the control of the future European state.

In more general terms, the different kinds of administrative law illustrate the various answers to the unsolved principal–agent problem described in section two of the 1776 Virginia Bill of Rights: 'that all power is vested in, and consequently derived from the people; that magistrates are their trustees and servants, and at all times amenable to them'.

References

Backhaus, Jürgen (2002), 'Legal and economic principles for the common administrative law in Europe', in A. Marciano and J.-M. Josselin (eds), *The Economics of Harmonizing European Law*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, 240–52.

Bermann, George and Peter Lindseth (1995), 'Réflexions sur le droit administratif aux Etats-Unis' ['Observations on administrative law in the United States'], in Conseil d'État, *Rapport Public*, Paris: La Documentation Française, pp. 515–28.

Bodin, Jean (1577), *Les Six Livres de la République* [The Six Books of the Republic], Paris: Jacques du Puis, Libraire juré à la Samaritaine (original edition).

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

Cooter, Robert (2000), *The Strategic Constitution*, Princeton, NJ: Princeton University Press.

Ekelund, Robert and Robert D. Tollison (1980), 'Economic regulation in mercantile England: Hecksher revisited', *Economic Inquiry*, **18**, 567–99.

Hume, David (1992), *Treatise of Human Nature*, Buffalo, NY: Prometheus (first published 1739–40).

Josselin, Jean-Michel and Alain Marciano (1995), 'Constitutionalism and common knowledge: assessment and application to the design of a future European constitution', *Public Choice*, **85**, 177–83.

Josselin, Jean-Michel and Alain Marciano (2000), 'Displacing your principal: two historical case studies of some interest for the constitutional future of Europe', *European Journal of Law and Economics*, **10**, 217–33.

Kelsen, Hans (1962), *La Théorie pure du droit* [The Pure Theory of Law], Paris: Dalloz (*Reine RechtsLehre*: first published 1934, rev. edition 1960).

Rose-Ackerman, Susan (1992), *Rethinking the Progressive Agenda: The Reform of the American Regulatory State*, New York: Free Press.

Smith, Adam (1976), *The Theory of Moral Sentiments*, Indianapolis: Liberty Press (first published 1759).

Vanberg, Victor (1994), 'Social contract theory', in P.J. Boettke (ed.), *Elgar Companion to Austrian Economics*, Aldershot, UK and Brookfield, US: Edward Elgar, pp. 337–42.

Witt, Ulrich (1992), 'The emergence of a protective agency and the constitutional dilemma', *Constitutional Political Economy*, **3**, 255–66.

15 Property

Thomas J. Miceli

Introduction

The concept of property is fundamental to both law and economics. The law defines and protects the bundle of rights that constitute property, thereby creating the legal framework within which resource allocation and wealth distribution take place. The economic approach to property law emphasizes its role in promoting an efficient allocation of resources. Accomplishing this goal generally involves creation and protection of individual rights in property so as to encourage exchange and investment, though, in some cases, transaction costs, scale effects, or other factors may favour third-party or collective control of property rights. The purpose of this entry is to show how legal rules define and protect the efficient property rights regime.

The discussion begins with the fundamental problem of assigning property rights. It then turns to an examination of legal rules governing transfers (voluntary exchange) and violations (involuntary exchange) of property rights. The theme is that these rules are designed to minimize the impact of transaction costs and uncertainties about ownership. The final topic concerns the optimal scale of ownership: when is efficiency served by individual ownership and when by common or collective ownership? The principal tradeoff is between the incentives of private ownership and the scale economies of common ownership.

Origin and assignment of property rights

In a seminal article on the emergence and function of property rights, Demsetz (1967) pointed out the fundamental link between property rights and externalities. Externalities are defined to be costs or benefits that one party imposes on others while not taking account of them in his or her own decision making. The presence of externalities implies the existence of unrealized gains from trade, a situation that economists call 'market failure' (Buchanan and Stubblebine, 1962). Demsetz argued that these gains will remain unrealized as long as the cost of internalization exceeds the gains. However, 'property rights develop to internalize externalities when the gains of internalization become larger than the cost of internalization' (Demsetz, 1967, p. 350). Often internalization occurs through bargaining, or mutually beneficial exchange. However, when transaction costs are high, internalization may still be achieved by coercive means, subject to qualifications noted below.

The Coase theorem

Coase (1960) provided the seminal discussion of the relationship among the assignment (or definition) of property rights, transaction costs and the efficient internalization of externalities. (Coase framed his discussion in the context of external costs, but the same logic applies to external benefits.) Coase began by noting that economists traditionally viewed external costs as requiring government intervention (coercion) to ensure the efficient outcome. According to this 'Pigouvian' view, the government first identifies the 'cause' of the external harm (for example, a rancher whose cattle stray onto a farmer's land and damage his crops) and then imposes a tax or fine on that party, thereby inducing him to curtail his activities to the efficient level (for example, by reducing his herd size or fencing his land).

Coase's work provided an important critique of the Pigouvian view. Specifically, he noted that it is based on a particular assignment of property rights – namely, the victim of the external cost (the farmer) is awarded the right to be free of the harm. Thus, if the rancher wants to allow his cattle to stray, he must 'purchase' the right by paying the tax. And, if the tax is set equal to the farmer's damages, the rancher will allow the efficient number of cattle to stray, either by reducing his herd or erecting a fence, whichever is cheaper.

Suppose, in contrast, that the rancher is awarded the right to allow his cattle to stray without penalty (the case of an 'open range'). The traditional view would hold that the rancher will expand his herd inefficiently because he will ignore the cost suffered by the farmer. Coase argued, however, that if transaction costs between the farmer and rancher are low enough, the parties will engage in a bargain whereby the farmer pays the rancher to reduce the number of straying cattle to the socially efficient number. In other words, the farmer will purchase the right to be free of damage from inefficiently straying cattle.

This example shows that, regardless of how initial property rights are assigned in externality settings, the efficient outcome can be achieved without government intervention if transaction costs are low. This conclusion is known as the Coase theorem. Another way of stating the theorem is that the initial assignment of property rights, as determined, for example, by the rule of liability, is irrelevant for the final allocation of resources (Demsetz, 1972).

The irrelevance of the law with respect to resource allocation when transaction costs are low was one important insight of Coase's analysis; another has to do with the notion of causation in externality settings. As noted above, the traditional view of externalities identifies *the* cause of the external harm and taxes or fines that party. This implies a one-way causal relationship from the 'injurer' (rancher) to the 'victim' (farmer). Coase pointed out that causation is actually reciprocal in the sense that both parties simultaneously cause the harm. That is, the presence of both parties is necessary for harm to occur

– the absence of either eliminates the harm. (Thus both parties are 'causes-in-fact' of the harm: ibid., p. 25.)

It follows that the designation of the 'injurer' and 'victim' are, in a sense, products of the initial assignment of the right. To illustrate this point, consider the case of *Spur Industries* v. *Del E. Webb Development Co.* (1972). This was a nuisance case brought by a residential developer to shut down a feedlot that emitted offensive odours. The court ordered the lot to shut down but, because it was pre-existing, the developer was required to pay its relocation costs. Thus the developer would only shut down the feedlot if he valued elimination of the nuisance more than the cost of relocating the feedlot to its next-best site. This is the efficient relocation criterion, assuming that the feedlot did not foresee residential development when it made its initial location choice (Posner, 2003, pp. 62–3). Note that, in this case, the court's award of damages to the feedlot for its losses implicitly assigned the role of 'victim' to the feedlot (by virtue of its having been 'first in time') and 'injurer' to the developer.

Although the foregoing examples have shown that the initial assignment of rights does not matter for resource allocation in a zero transaction cost world, it *does* matter for the distribution of wealth. Specifically, the party who is endowed with the property right receives something of value which the other party must purchase if he or she values it more highly. Thus the initial assignment of rights dictates the direction of monetary payments that will occur in moving to the final (efficient) allocation.

To this point we have considered a world in which transaction costs are zero or low. In most real-world settings, however, significant transaction costs are present (see, for example, the case studies in Ellickson, 1991). In those cases, the assignment of rights may have allocative effects, suggesting that property law needs to be sensitive to the bargaining costs of the parties to a dispute. Legal rules can sometimes be chosen to minimize the cost of bargaining in an effort to facilitate private resolution of disputes. For example, property rights can be defined in clear and simple terms, as when they are assigned by the rule of first possession (Wittman, 1980; Epstein, 1986; Lueck 1995). In other cases, transaction costs may prohibit bargaining altogether, regardless of the clarity of property rights. In that case, the court, when given the opportunity to resolve a dispute over property rights, should assign them initially to the party that values them most (assuming that the court can observe the parties' valuations – a point discussed below). Such an assignment mimics the outcome of bargaining without requiring the expenditure of transaction costs. (For further discussion and extensions of the Coase theorem, see Frech, 1979; Coleman, 1982; Cooter, 1982; and Holderness, 1989.)

As a final point regarding the assignment of property rights, note that Coase's analysis, while emphasizing the role of bargaining as opposed to

third-party intervention in externality cases, nevertheless maintains the view of the state 'as the sole creator of operative rules of entitlements among individuals' (Ellickson, 1991, p. 4). This view ignores the importance of social customs and norms for defining rights and allocating resources in many settings. Economic analysis of these situations often appeals to the notion of 'spontaneous order' (or 'order without law') to explain the emergence of socially desirable outcomes in the absence of state or legal intervention (see generally Barzel, 1989, and Ellickson, 1991).

Legal protection of property rights

The preceding section focused on the *assignment* of property rights; this is one role of the law. Another is to provide rules that both govern voluntary transfers and provide legal remedies for violations (involuntary transfers) of those rights. Calabresi and Melamed (1972) developed an economic theory of rules for transferring rights that is a natural extension of Coase's analysis of the externality problem. They distinguished two rules for protecting legal entitlements or rights: *property rules*, which allow rightholders to enjoin all attempts to acquire the right on terms that they deem unacceptable; and *liability rules*, which only allow rightholders to seek monetary compensation (as set by a court) for seizures of the right.

The key distinction between the two rules is that the rightholder's *consent* is required for a transfer under a property rule, but consent is not required under a liability rule. The tradeoff is that, on the one hand, consent guarantees that any exchanges that occur will be mutually beneficial and therefore efficient, but, on the other, the transaction costs involved in obtaining consent will sometimes be large enough to prevent the completion of otherwise efficient exchanges. This tradeoff suggests that property rules are the preferred remedy when transaction costs are low because they facilitate mutually beneficial bargaining between private parties. In contrast, liability rules may be preferred when transaction costs are high because they allow the court to coerce exchanges when bargaining is not possible. The choice of liability rules in high transaction cost settings is not unqualified, however, because the court is generally not able to observe subjective values, which may result in some inefficient transfers, if, for example, the court sets the wrong damage amount. Thus, the cost of potentially inefficient exchanges under liability rules needs to be weighed against the cost of forgone transactions under property rules. This proposition regarding the choice between property rules and liability rules, along with the Coase theorem and its corollaries stated above, form the core of the economic theory of property law. (For further refinements of the economic choice between property rules and liability rules, see Polinsky, 1980; Coleman, 1988, ch. 2; Kaplow and Shavell, 1996.)

Controlling externalities trespass, nuisance and regulation

Externalities are a form of involuntary invasion or acquisition of one's property rights. The above theory suggests that the law should provide remedies for these invasions so as to resolve them in the most efficient manner. The principal common law remedies for externalities are embodied in the laws of trespass and nuisance. Under trespass, a property owner can seek to enjoin unwanted invasions of his or her property, whereas, under nuisance, a property owner is entitled to relief (damages or an injunction) only if the harm is substantial and the nuisance-creating activity is judged to be unreasonable (inefficient).

Consistent with the above theory, Merrill (1985) has argued that transaction costs broadly explain this distinction. Specifically, he argued that when transaction costs between the parties to the dispute are low, trespass generally governs, whereas when the transaction costs are high, nuisance generally governs. Trespass is efficient when transaction costs are low because, like a property rule, it forces the parties to bargain over the allocation of property rights. In contrast, nuisance is preferred when transaction costs are high because the parties may not be able to resolve the dispute on their own. Thus the court dictates the conditions under which relief is granted by conducting a reasonableness test to determine whether or not the nuisance-creating activity is cost-justified. In doing so, the court in principle imposes a value-enhancing outcome. The case of *Boomer* v. *Atlantic Cement Co.* (1970), which concerned a cement company that imposed costs on several neighbouring landowners, illustrates the economic argument for choosing a court-imposed damage remedy in externality settings involving high transaction costs between injurers and victims.

Externalities are often resolved, not by trespass and nuisance (private remedies), but by direct government regulation (a public remedy). In the same way that the choice between trespass and nuisance can be explained in terms of transaction costs, the decision to use a public rather than a private remedy can also be explained (Ellickson, 1973). In some externality settings, the costs will be dispersed across a large number of victims, as in the case of many types of environmental harms. (Such harms are generally referred to as public nuisances.) The problem in these cases is that no one victim may suffer enough harm to justify the cost of pursuing a private remedy, even though the *aggregate harm* exceeds the benefit of the activity (that is, a nuisance suit would have succeeded in obtaining a remedy). In that case, the government acts as an agent of the victims by regulating the offending activity (Landes and Posner, 1987, ch. 2). The justification for government regulation is thus high transaction costs *among victims*, which precludes a private remedy.

Table 15.1 summarizes the transaction cost argument for choosing among trespass, nuisance and public control. Specifically, it shows that a private

Table 15.1 Optimal remedies for externalities

			Transaction costs between injurer and victim(s)	
			Low	High
Transaction costs among victims	Low	private	trespass	nuisance
	High	public	market acquisition	eminent domain

remedy is preferred when transaction costs among victims are low, and a public remedy is preferred when these costs are high. It also shows that, in the case of a private remedy, trespass is preferred when transaction costs between the injurer and victim(s) are low, and nuisance is preferred when these costs are high.

Government acquisition of rights and compensation

The preceding argument provides an economic justification for the government's acting as an agent of private individuals in regulating or acquiring private property from an owner who values the property less than the group collectively. This justification does not, however, explain why the government has the right to *coerce* such an exchange rather than having to bargain with the owner. In other words, once the government solves the collective action problem among those seeking a property right by acting as their agent in dealing with the owner of the right, why is the government not required to bargain with the owner for a mutually beneficial exchange?

The answer is that transaction costs between the government and the owner of the property may be high, thereby preventing successful bargaining. This is the economic justification for the government's power of eminent domain, as shown by the bottom right entry in Table 15.1 (Epstein, 1985; Posner, 2003, p. 55). The typical source of the transaction costs when the government acquires property is the hold-out problem. Consider a situation in which the government is assembling a large number of properties for the purpose of building a public good like a highway. Once the assembly of properties has begun, it becomes very costly to seek an alternative route. Knowledge of this fact confers monopoly power on individual owners whose land is in the path of the project. They therefore have an incentive to hold out for prices in excess of the opportunity cost of their land. This problem (a form of rent seeking) creates the potential for very high transaction costs if the government has to bargain with each owner individually (that is, if each

owner has property rule protection). Replacing the owners' right to consent with a right to compensation (that is, substituting liability rule protection for property rule protection) reduces the transaction costs and allows beneficial public projects to go forward. This occurs, however, at the risk of undercompensating owners for their land, which may result in overacquisition of private property for public use (Munch, 1976).

This discussion implies that when the government does not face a hold-out problem, it should be required to bargain with individual owners. (This is indicated by the entry labelled 'market acquisition' in Table 15.1.) The public use requirement may be one way that the court limits the government's use of its taking power in such cases (Epstein, 1985). In addition, Fischel notes that eminent domain is 'self-limiting' in that, when the hold-out problem is absent, the transaction costs of using the market are generally lower than the cost of using eminent domain (Merrill, 1986a; Fischel, 1995, p. 74).

A transaction cost argument also provides a basis for coercive government *regulation* (as opposed to acquisitions) of property. Specifically, the transaction costs involved in government acquisition of individual property rights that are more valuable to the public than to the owner (for example, the right to engage in land uses that degrade the environment or that threaten an endangered species) will ordinarily preclude acquisition by mutually beneficial exchange. Thus landowners do not have the right to block government acquisition of those sticks in their bundle of rights that would impose widespread costs on others.

In contrast to physical acquisitions, however, courts have generally not given property owners the right to *compensation* for these 'partial acquisitions', except when the reduction in the value of the land is total or substantial. (See *Pennsylvania Coal Co.* v. *Mahon* (1922) and *Lucas* v. *South Carolina Coastal Council* (1992).) Uncompensated regulations have been justified by courts as legitimate police power actions aimed at protecting public safety and welfare (*Mugler* v. *Kansas*, 1887). The problem is that failure to pay compensation creates the threat of over-regulation by the government (Fischel and Shapiro, 1988). At the same time, economists have shown that full compensation for lost value creates a moral hazard problem whereby property owners ignore the external costs that certain uses of their property may impose, thus causing them to overinvest in those uses (Blume et al., 1984). Economic approaches to the compensation question propose compensation rules that balance these and other factors. (See, for example, Michelman, 1967; Sax, 1964, 1971; Fischel, 1995; Miceli and Segerson, 1996.)

Title assurance and the transfer of property

The origin and assignment of property rights is a necessary first step for promoting exchange. The process of exchange then conveys the rights of

ownership to subsequent owners, including the right of future disposition. However, a potential impediment to transactions is uncertainty about ownership. For example, the potential for theft, fraudulent transfer or error in recording ownership of the property creates the possibility of a break in the chain of title, in which case the 'owner' and the 'possessor' will not be the same person. As a result, a possessor who acquired the property in a legitimate transfer may find his or her rights of ownership challenged by a previously defrauded owner. In that case, the law of property must decide whom to favour: the last rightful owner, or the current possessor. Legal rules for establishing and assuring title provide the answer to this question (Baird and Jackson, 1984).

Possession versus public recording of title

In a world of uncertainty, there are two methods for assuring title. One is simply to vest title in the *possessor* of a piece of property. The benefit of this type of system is that the cost of verifying title is low since a would-be purchaser need not be concerned with the sequence of transactions by which the current possessor acquired the property – the fact of possession itself assures title. The cost is that owners are at risk of losing their property without their consent, as by theft, thus encouraging unproductive investment in security and, in the case of land, discouraging productive investment in improvements.

The other system for assuring title to a piece of property is a public record that records all legitimate transactions involving the property. Thus any would-be purchaser can, in theory, trace the ownership of the property back to the point where it was first transformed into private property. Searching the title record thus assures the purchaser that he or she is dealing with the legitimate owner and that no claimants will subsequently appear to challenge his or her ownership. In this way, it protects owners from non-consensual loss of title and discourages theft, since a wrongfully displaced owner can reclaim title.

A recording system is costly to administer, however, making it impractical for most types of property. The efficient system, therefore, balances the risk of involuntary transfer under a possession-based system against the cost of verifying ownership via a public record. In view of this tradeoff, a recording system is comparatively better for property that is valuable, not transferred often, relatively immobile or easy to identify, and when the ability to divide or share ownership of the property is valuable (Baird and Jackson, 1984). Real property is the paradigmatic example of this type of property. (The Statute of Frauds, which requires written rather than oral contracts for the transfer of land, represents the contractual counterpart to the recording system: Friedman, 1985, p. 278.)

In contrast, a possession-based system is preferred for property that is frequently transferred and not easily described or identified. The classic example is money. Most forms of personal property are intermediate between land and money and therefore could conceivably be protected by a filing system. Thus, when personal property is sufficiently valuable (cars, for example), a recording system is cost justified.

An interesting combination of recording and possession systems for land is the Torrens registration system, under which the government, following a title investigation, certifies the owner's title against any future claims (Janczyk, 1977; Shick and Plotkin, 1978). Legitimate claimants therefore cannot seek recovery of the land but only compensation out of a government fund financed by registration fees. In contrast, owners typically purchase private title insurance under a recording system to protect against the risk of losing their property to a legitimate claimant. (See Miceli and Sirmans, 1995a, and Miceli et al., 2002 for a comparison of the Torrens and recording systems.) Land registration has succeeded in England, but it has not been widely used in the United States (Cribbet, 1975; Friedman, 1985, p. 434; Bostick, 1987).

Title protection and economic development

Several economic studies have examined the link between formal title protection and economic development. De Soto (2000), for example, has claimed that the lack of a government-backed title system in most developing countries accounts for their failure to match the economic growth of Western market economies. Empirical evidence largely supports the hypothesis that formal title enhances land value and promotes investment (Besley, 1995; Alston, et al., 1996; Miceli et al., 2001a).

Involuntary transfers: adverse possession

The foregoing has emphasized the role of title assurance in promoting efficient exchange of property. In general, this involved removal of impediments to voluntary exchange such as uncertain ownership. The discussion of property rules versus liability rules, however, suggested that, in some cases, efficiency may be better promoted through involuntary exchange. This is true when the transaction costs of voluntary exchange are so high that otherwise efficient transactions are forgone. For example, the doctrine of adverse possession provides a means whereby a trespasser can acquire title without the owner's consent by occupying the land continuously for a statutorily determined period of time (usually ten or more years). In effect, adverse possession places a time limit on an owner's right to exclude trespassers from his or her property. It thus represents a 'time-limited property rule' (Miceli and Sirmans, 1995b).

The economic benefits generally attributed to adverse possession are twofold. First, it clears title to land after passage of the statutory period, thereby

lowering the costs to would-be purchasers of discovering the legitimate title-holder and removing the risk of a future claim (Epstein, 1986; Merrill, 1986b; Netter et al. 1986). Second, adverse possession prevents land from being left idle for long periods of time. One should be cautious, however, in invoking these justifications, first, because modern land records are ordinarily easily accessible to would-be buyers and, second, because land left idle is not necessarily being used inefficiently, given the option value of future development.

A more satisfactory explanation for adverse possession is that it minimizes the costs of boundary errors, which are the principal source of claims. If we suppose that an encroacher's valuation of the mistakenly occupied land grows over time, a rule that awards title to the occupier after a certain period of time limits the loss that he or she would have to endure if expelled (Stake, 1995). Offsetting this is the risk that setting the statutory period too short would encourage intentional encroachment. The optimal statutory period thus balances these two effects (Ellickson, 1986; Miceli and Sirmans, 1995b; Baker et al., 2001).

The optimal scale of ownership: private versus common property

The discussion so far has focused primarily on private, or individual, ownership of property rights. There are certain circumstances, however, in which common, or public, ownership of rights might be optimal. The difference between private and common ownership is that, under the latter, all individuals in the community can exercise the rights of ownership. That is, members of the community cannot be excluded, whereas members outside the community can be. (Common ownership is thus a 'restricted access' property regime, as opposed to an 'open access' regime where no one is excluded: Lueck, 1994; Anderson and Swimmer, 1997.) In contrast, under private ownership, the individual has the right to exclude all others from use of his or her property. (Private ownership is thus an extreme case of restricted access.) The primary disadvantage of common ownership stems from externalities – the fact that individual users will not take full account of the effects of their activities on other users. At the same time, common ownership can offer advantages associated with scale economies and risk sharing. The optimal scale of ownership involves a balancing of these effects.

Private versus common ownership of land

Demsetz (1967) made the seminal argument for private ownership. When land is communally owned, each member of the group will have an incentive to overwork it because he or she does not bear the full costs of doing so. As a result, the resource value of the land (its fertility, stock of game or deposit of minerals) will be diminished too quickly (Hardin, 1968; Cornes and Sandler, 1986, pp. 128–31). Negotiation among the members of the group and/or

efforts to limit the number of users can curtail this tendency (Ostrom, 1990), but that involves the cost of negotiating multiple agreements and of policing those agreements. In contrast, individual ownership of land results in efficient use of parcels, since owners will fully internalize the costs and benefits of their activities. (See Anderson and Lueck, 1992, and Besley, 1995, for empirical evidence in support of this proposition.)

Of course, externalities remain under private ownership, as when the actions of individual owners spill over onto neighbouring parcels. Again, these effects can be resolved through Coasean bargaining. However, one expects that the costs of bargaining will be considerably lower under individual ownership since only those owners whose parcels are affected by the spillover need bargain. In contrast, all members of the group must participate in any bargain under common ownership since none can be excluded from use of the property. Further, private ownership allows another solution to externalities, namely, consolidation of parcels affected by spillovers. In some cases this will be a lower-cost way to internalize externalities compared to bargaining; it depends on the costs of bargaining compared to the scale economies and governance costs associated with consolidation (Coase, 1937).

Still, common ownership offers some advantages over private ownership. This is true for goods whose consumption is non-rivalrous or public and for large-scale externalities (Ellickson, 1993, pp. 1334–5). In these situations, multilateral bargaining is prohibitively costly, and the scale of the externality is too large to make consolidation into single ownership desirable. Thus some level of public ownership may be the least-cost option. This typically requires coercion by the governing body to prevent free-riding by members of the group (governance costs), as when taxes are raised to provide public goods, and expenditure of resources to enforce the boundaries of the common (exclusion costs) (Field, 1989; Ostrom, 1990; Lueck, 1994, 1995; Anderson and Swimmer, 1997).

Publicness represents a technological basis for group ownership. Two other reasons are based on preferences. The first is the desire of owners to spread the risk that their land may be damaged or unproductive (Ellickson, 1993, pp. 1342–4). A sole owner bears the entire risk, whereas under group ownership the risk is spread across all members since no individual's return is tied to a particular parcel. Of course, this approach to sharing risks is not ideal because it may create a tradeoff between risk sharing and incentives. Thus we would expect group ownership to decline in importance as other methods for sharing risks such as insurance markets are developed.

A final reason for communal ownership of land is a group's preference for egalitarianism, or equal sharing of wealth among its members. This represents a case where *private* (as well as public) goods are publicly owned. As a consequence, Ellickson (ibid., p. 1351) notes that pursuit of egalitarianism

'increases the need for pervasive controls against shirking, and may prompt the most skillful workers to consider greater rewards outside the commune'. Indeed, Cosgel et al. (1997) argue that the optimal scale for equal sharing of wealth within communal societies may be limited by consideration of these factors.

Temporal division of ownership

Common ownership is a form of divided ownership in which all members of the group possess rights in the object at a point in time. Property rights can also be divided through time, as in the case of leases. However, incentives for inefficient use exist under temporal division as well (Epstein, 1986, pp. 707–10; Miceli et al., 2001b; Posner, 2003, pp. 71–5,). For example, tenants will not internalize the full costs and benefits of their actions since their rights to the property are limited in time (and possibly scope). The allocative costs of dividing ownership, however, may be offset by the benefits, one of which is specialization. For example, a lease allows landlords to specialize in bearing the risk of changes in the market value of property, and tenants to specialize in combining labour and capital inputs with land to produce output.

Modern landlord–tenant law has redefined the obligations of parties to residential leases by increasing the rights of tenants relative to landlords (Hirsch and Law, 1979; Rabin, 1984). The Coase theorem and its corollaries suggest that this trend may enhance efficiency (rather than being merely redistributive) if transaction costs between landlords and tenants in residential settings have become high enough to preclude bargaining and if tenants value the rights more than landlords (Posner, 2003, pp. 484–5).

Property rights in information

Information has the characteristics of a public good because the cost of producing it does not increase in proportion to the number of consumers. Thus information would seem to be well-suited to public or common ownership. This assignment of rights, however, would likely lead to too little production of new information because producers would have difficulty appropriating its full social value. At the other extreme, assigning exclusive rights in information to its producer would limit its use if it is costly for those who value information to locate and bargain with the producer. This is the basic tradeoff involved in assigning property rights in information.

Patent law addresses this tradeoff by awarding innovators exclusive rights to an invention for a limited period of time (17–20 years) (Kitch, 1977). Similarly, a copyright is granted to the creator of an original idea, but that right is also limited in duration (the creator's life plus 50 years in the United States) and scope (quotations and limited copying for educational rather than commercial uses are generally allowed under fair use) (Landes and Posner, 1989).

Conclusion

This entry has reviewed the economic approach to property law. Owing to space limitations, the scope of the discussion was necessarily limited. In particular, it was primarily in the context of real property and focused on a few fundamental principles, including the role of property rights in promoting efficient exchange and investment, and on optimal ownership regimes in the presence of externalities, economies of scale and transaction costs. The basic theme, however, was general – namely, that economic theory provides a unifying framework for understanding many aspects of property rights institutions.

Acknowledgement

I wish to acknowledge the comments of Bob Ellickson, Derek Johnson, Dean Lueck, Kathy Segerson, C.F. Sirmans and an anonymous reviewer.

List of cases

Boomer v. *Atlantic Cement Co.*, 26 N.Y.2d 219, 257 N.E.2d 870 (1970).
Lucas v. *South Carolina Coastal Council*, 112 S.Ct. 2886 (1992).
Mugler v. *Kansas*, 123 U.S. 623 (1887).
Pennsylvania Coal Co. v. *Mahon*, 260 U.S. 393 (1922).
Spur Industries v. *Del E. Webb Development Co.*, 494 P.2d 701 (Ariz. 1972).

References

Alston, L., G. Libecap and R. Schneider (1996), 'The determinants and impact of property rights: land titles on the Brazilian frontier', *Journal of Law, Economics and Organization*, **12** (1), 25–61.
Anderson, C. and E. Swimmer (1997), 'Some empirical evidence on property rights of first peoples', *Journal of Economic Behavior and Organization*, **33** (1), 1–22.
Anderson, T. and Dean Lueck (1992), 'Land tenure and agricultural productivity on Indian reservations', *Journal of Law and Economics*, **35**, 427–54.
Baird, D. and T. Jackson (1984), 'Information, uncertainty and the transfer of property', *Journal of Legal Studies*, **13** (2), 299–320.
Baker, M., T. Miceli, C.F. Sirmans and G. Turnbull (2001), 'Property rights by squatting: land ownership risk and adverse possession statutes', *Land Economics*, **77**, 360–70.
Barzel, Y. (1989), *Economic Analysis of Property Rights*, Cambridge: Cambridge University Press.
Besley, T. (1995), 'Property rights and investment incentives: theory and evidence from Ghana', *Journal of Political Economy*, **103** (5), 903–37.
Blume, L., D. Rubinfeld and P. Shapiro (1984), 'The taking of land: when should compensation be paid?', *Quarterly Journal of Economics*, **99**, 71–92.
Bostick, C.D. (1987), 'Land title registration: an English solution to an American problem', *Indiana Law Journal*, **63**, 55–111.
Buchanan, J. and W. Stubblebine (1962), 'Externality', *Economica*, **29**, 371–84.
Calabresi, G. and A.D. Melamed (1972), 'Property rules, liability rules and inalienability: one view of the cathedral', *Harvard Law Review*, **85**, 1089–128.
Coase, R. (1937), 'The nature of the firm', *Economica*, (n.s.), **4**, 386–405.
Coase, R. (1960), 'The problem of social cost', *Journal of Law and Economics*, **3**, 1–44.
Coleman, J. (1982), 'The economic analysis of law', in J. Pennock and J. Chapman (eds), *Ethics, Economics and the Law: NOMOS XXIV*, New York: New York University Press, pp. 83–103.

Coleman, J. (1988), *Markets, Morals and the Law*, New York: Cambridge University Press.
Cooter, R. (1982), 'The cost of Coase', *Journal of Legal Studies*, **11** (1), 1–33.
Cornes, R. and T. Sandler (1986), *The Theory of Externalities, Public Goods and Club Goods*, Cambridge: Cambridge University Press.
Cosgel, M., T. Miceli and J. Murray (1997), 'Organization and distributional equality in a network of communes: the Shakers', *American Journal of Economics and Sociology*, **56** (2), 129–44.
Cribbet, J. (1975), *Principles of Property Law*, Mineola, NY: The Foundation Press.
De Soto, H. (2000), *The Mystery of Capital*, New York: Basic Books.
Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review*, **57** (2), 347–59.
Demsetz, H. (1972), 'When does the rule of liability matter?', *Journal of Legal Studies*, **1**, 13–28.
Ellickson, R. (1973), 'Alternatives to zoning: covenants, nuisance rules and fines as land use controls', *University of Chicago Law Review*, **40**, 681–782.
Ellickson, R. (1986), 'Adverse possession and perpetuities law: two dents in the libertarian model of property rights', *Washington University Law Quarterly*, **64** (3), 723–37.
Ellickson, R. (1991), *Order without Law*, Cambridge, MA: Harvard University Press.
Ellickson, R. (1993), 'Property in land', *Yale Law Journal*, **102** (6), 1315–400.
Epstein, R. (1985), *Takings: Private Property and the Power of Eminent Domain*, Cambridge, MA: Harvard University Press.
Epstein, R. (1986), 'Past and future: the temporal dimension of property law', *Washington University Law Quarterly*, **64** (3), 667–722.
Field, B. (1989), 'The evolution of property rights', *Kyklos*, **42**, 319–45.
Fischel, W. (1995), *Regulatory Takings: Law, Economics and Politics*, Cambridge, MA: Harvard University Press.
Fischel, W. and P. Shapiro (1988), 'Takings, insurance and Michelman: comments on economic interpretations of "just compensation" law', *Journal of Legal Studies*, **17**, 269–93.
Frech, H. (1979), 'The extended coase theorem and long run equilibrium: the non-equivalence of liability rules and property rights', *Economic Inquiry*, **17**, 254–68.
Friedman, L. (1985), *A History of American Law*, 2nd edn, New York: Touchstone Books.
Hardin, G. (1968), 'The tragedy of the commons', *Science*, **162**, 1243–8.
Hirsch, W. and C. Law (1979), 'Habitability laws and the shrinkage of substandard rental housing stock', *Urban Studies*, **16**, 19–28.
Holderness, C. (1989), 'The assignment of rights, entry effects and the allocation of resources', *Journal of Legal Studies*, **18** (1), 181–9.
Janczyk, J. (1977), 'An economic analysis of land title systems for transferring property', *Journal of Legal Studies*, **6** (1), 213–33.
Kaplow, L. and S. Shavell (1996), 'Property rules versus liability rules', *Harvard Law Review*, **109** (4), 713–90.
Kitch, E. (1977). 'The nature and function of the patent system', *Journal of Law and Economics*, **20**, 265–90.
Landes, W. and R. Posner (1987), *The Economic Structure of Tort Law*, Cambridge, MA: Harvard University Press.
Landes, W. and R. Posner (1989), 'An economic analysis of copyright law', *Journal of Legal Studies*, **18**, 325–63.
Lueck, D. (1994), 'Common property as an egalitarian share contract', *Journal of Economic Behavior and Organization*, **25**, 93–108.
Lueck, D. (1995), 'The rule of first possession and the design of the law', *Journal of Law and Economics*, **38** (2), 393–436.
Merrill, T. (1985), 'Trespass, nuisance and the cost of determining property rights', *Journal of Legal Studies*, **14** (1), 13–48.
Merrill, T. (1986a), 'The economics of public use', *Cornell Law Review*, **72**, 61–116.
Merrill, T. (1986b), 'Property rules, liability rules and adverse possession', *Northwestern University Law Review*, **79**, 1122–54.

Miceli, T., H. Munneke, C.F. Sirmans and G. Turnbull (2002), 'Title systems and land values', *Journal of Law and Economics*, **45**, 565–82.
Miceli, T. and K. Segerson (1996), *Compensation for Regulatory Takings: An Economic Analysis with Applications*, Greenwich, CT: JAI Press.
Miceli, T. and C.F. Sirmans (1995a), 'The economics of land transfer and title insurance', *Journal of Real Estate Finance and Economics*, **10**, 81–8.
Miceli, T. and C.F. Sirmans (1995b), 'An economic theory of adverse possession', *International Review of Law and Economics*, **15** (2), 161–73.
Miceli, T., C.F. Sirmans and J. Kieyah (2001a), 'The demand for land title registration: theory with evidence from Kenya', *American Law and Economics Review*, **3**, 275–87.
Miceli, T., C.F. Sirmans and G. Turnbull (2001b), 'The propert-contract boundary: an economic analysis of leases', *American Law and Economics Review*, **3**, 165–85.
Michelman, F. (1967), 'Property, utility and fairness: comments on ethical foundations of "just compensation" law', *Harvard Law Review*, **80**, 1165–258.
Munch, P. (1976), 'An economic analysis of eminent domain', *Journal of Political Economy*, **84**, 473–97.
Netter, J., P. Hersch and W. Manson (1986), 'An economic analysis of adverse possession', *International Review of Law and Economics*, **6**, 217–27.
Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.
Polinsky, A.M. (1980), 'On the choice between property rules and liability rules', *Economic Inquiry*, **18** (2), 233–46.
Posner, R. (2003), *Economic Analysis of Law*, 6th edn, New York: Aspen.
Rabin, E. (1984), 'The revolution in residential landlord–tenant law: causes and consequences', *Cornell Law Review*, **69**, 517–84.
Sax, J. (1964), 'Takings and the police power', *Yale Law Journal*, **74**, 36–76.
Sax, J. (1971), 'Takings, private property and public rights', *Yale Law Journal*, **81**, 149–86.
Shick, B. and I. Plotkin (1978), *Torrens in the United States: A Legal and Economic Analysis of Land Registration*, Lexington, MA: D.C. Heath & Co.
Stake, J. (1995), 'Loss aversion and involuntary transfers of title', in R. Malloy and C. Braun (eds), *Law and Economics: New and Critical Perspectives*, New York: Peter Lang, pp. 331–60.
Wittman, D. (1980), 'First come, first served: an economic analysis of "coming to the nuisance"', *Journal of Legal Studies*, **9** (3), 557–68.

16 The European Union's institutional design

Elisabetta Croci Angelini

Introduction

The Treaty establishing the European Economic Community (TEC) was signed in Rome by the six founding countries (Belgium, France, Germany, Italy, Luxembourg, and the Netherlands) on 25 March 1957. After the first enlargement to Denmark, Ireland and the UK in 1973, the second to Greece in 1981, and the third to Spain and Portugal in 1986, the cooperation among the 12 member states was fostered by a stronger agreement accomplished by the first revision to the Treaty of Rome, the Single European Act (SEA) in force since 1987, and later by the European Union Treaty (TEU) signed in Maastricht in 1991. The European Union (EU) came into existence in 1993 and consists of three pillars: the European Community (EC), the Common Foreign and Security Policy (CFSP) and Justice and Home Affairs (JHA). Austria, Finland and Sweden joined the EU in 1995. Two more revision treaties were agreed: in Amsterdam in 1997 and in Nice in 2000. The most recent enlargement involved the entry of eight Eastern European countries (the Czech Republic, Slovakia, Slovenia, Hungary, Estonia, Latvia, Lithuania and Poland) plus Cyprus and Malta, in May 2004. The process of European integration may be described as a sequence of successful widening and deepening operations, yet this representation overlooks the considerable modifications undertaken over the years.

The integration process was initiated with the application of the incremental tactic envisaged by Jean Monnet, pointing to the development of a European interest above the national interests. The strategy was to pursue common policies in the areas where cooperation could be exploited to mutual advantage. The self-enhancing character of the succession of agreements consisted in each common policy creating spillovers, from which a new common project entailing positive net benefits for all members could be devised.

This strategy is behind the functionalist view of the EU integration process whereby the EU is a continuous process of accumulation of cooperative agreements, based on two commitments: (i) the aim of further integration, stated in the preamble of the treaties of Rome (TEC) and Maastricht (TEU); and (ii) the acceptance of the *acquis communautaire* (the set of EU treaties, laws, rules and practices), whereby the accession countries are required to accept the EU obligations as such. However, the accession of new members is usually preceded by many years of negotiation, followed by a transitory

period. The six founding countries had a 12-year transitory period, divided into three phases and ending in 1969, while Spain and Portugal were given a six-year period. The UK, having changed government just after acceding, then succeeded in renegotiating its terms of accession. Negotiations for the 2004 enlargement were begun in 1998 and the transitory period is planned to end in 2006. On rare occasions some exceptions, exemptions or delayed applications to the *acquis communautaire* have also been sanctioned.

In the absence of a supranational state or federal government, the EU has developed a system of laws, rules and decisions made by the four main institutions: the Commission, the Council, the Parliament and the Court of Justice.

The European Commission, composed of executive commissioners and headed by their president, is appointed by the national governments and approved by the Parliament. It has the important task of proposing common policies, ensuring the correct implementation of the treaties, and representing the EU in international trade negotiations (formerly within the General Agreement and Tariffs and Trade: GATT; and now in the World Trade Organization: WTO). In the EU Council, each national government is represented by the relevant minister (foreign affairs, finance, agriculture and so on). The Council's decision-making activity is endorsed by the European Council, the quarterly summit of the heads of state and governments of the member states, and includes the participation of both presidents of the Commission and the European Parliament. The European Parliament (EP) currently comprises 730 members elected by the national constituencies and organized in transnational political groups. The EP's powers, whose prerogatives have been continuously strengthened since the SEA and through to the Nice Treaty, are threefold: (i) supervisory: mainly trying to influence the Commission's proposals and Council decisions; (ii) legislative: it has a co-decision power along with the Council; and (iii) budgetary: its vote is decisive for the approval of the EU budget both *ex ante* and *ex post*. The Court of Justice (ECJ), which is composed of one judge for each member state, has the task of clarifying the interpretation of the treaties and EU law, and adjudicates over disputes which may arise among the member states, the EU institutions, as well as with any other concerned parties.

The EU institutional architecture is not characterized by a complete separation of powers among the four main institutional bodies. On the contrary, they exercise mutual monitoring through a complex system of checks and balances, finally reaching consensual decisions. According to principal–agent theory, this interdependence creates a conflict of interests – between the Commission and the Council (for example, the conflict over the possible revision of the Stability and Growth Pact: SGP) or between the Council and the Parliament (for example, the EU budget) – from which the appropriate

incentives for each institution and the officials' accountability should follow. The Commission has the agenda-setting power that enables it to control the integration process by putting forward proposals to be approved by the Council and the Parliament. Most of the EU legislation proposed by the Commission is approved under the co-decision procedure, whereby the Council and the Parliament can amend each other's changes to the proposal. If there is disagreement, the matter is passed to the Conciliation Committee, which makes the final decision regarding adoption or rejection. The power of the Council as a whole has been endorsed by the European Councils, which have acted as an important, though informal, decision-making forum and have strengthened the power in the hands of national governments. The Court of Justice is acknowledged to have acted as a main engine of integration by passing judgments that in many cases have fostered the integration mechanism. The doctrine of direct effect and supremacy of European over national law, which is needed to guarantee the coherence of the system, has been applied since its inception, and national courts accept its authority.

Methods of integration

An integration process among a number of national economies is characterized by the evaluation of benefits and costs. Benefits consist mainly of economies of scale and the internalization of reciprocal externalities. Costs are related to the heterogeneity of national preferences with regard to public goods provision.

EU integration has consisted of a collective action for delivering common policies with the characteristics of public or club goods to the member countries. So far, the most important achievements have been: (i) the Single Market, which is an area of free circulation of goods, services, capitals and individuals, separated by tariff as well as non-tariff barriers from the rest of the world; and (ii) The European Monetary Union (EMU), which has delivered the euro as a club good after the fixed (but adjustable) exchange rate agreement established in 1979 by the European Monetary System (EMS) delivered the public good of monetary stability to its members.

Both these achievements have certainly provided the firms and economic sectors of the member states with the advantage of benefiting from economies of scale. The Single Market for internal free trade has spread over larger markets, and the single currency for external trade has signalled the end of expectations of exchange rate variations as well as the likely upgrading of the euro to the rank of second international reserve currency, competing with the US dollar. In addition, the EMU has given the participating countries the advantage of the internalizing the externality of frequent devaluations which caused considerable macroeconomic instability in Europe during the last decades. The cost of the EU integration process is more difficult to assess. In

the past decades, an increasing heterogeneity across the EU member states has resulted from diverging fiscal policies due to public primary deficits and public debt accumulation. The coordination device of the SGP was created to foster convergence to sound national fiscal stances, after the enforcement of the Maastricht criteria came to an end with the monetary union.

However, two issues emerged connected to the lack of EU centralization of fiscal policies. First, the deflationary bias caused in the macroeconomic governance by the tight European Central Bank (ECB) monetary policy and by the restrictive orientation imposed on national fiscal policies by the SGP. Second, the absence of any internalization of the externalities provoked by the very different levels and composition of both income taxes and social expenditures.

In addition, the most recent EU enlargement is likely to result in further costs in the EU integration process. First, the EMU as a club good could be burdened by congestion costs. The higher average inflation of the accession countries might undermine the EMU monetary–fiscal policy mix as an ECB interest rate vote could result in a level that is too high for the incumbent economies' macroeconomic equilibrium (de Grauwe, 2003). Second, the extension to the accession countries of the EU common policies devoted to foster real convergence in the backward areas, could magnify the redistributive problems stemming from externalities. In particular, the welfare state reforms undertaken in the EMU countries are bound to take into account the structural change in the labour markets which will be created by the migration of workers coming from the Eastern countries.

The interpretation of the EU integration points to two approaches to integration: supranationalism and intergovernmentalism. The supranational approach to integration tends to pursue the idea of a European entity, defined in terms of social and cultural identity, which should be translated into a political and economic being. The intergovernmental approach instead conceives integration mainly as the coordination among national interests. In Table 16.1, various features of the EU institutions are classified under these approaches.

Table 16.1 Features of EU institutional bodies

	Appointment	Composition	Task	Decision rule
Commission	Intergovernmental	Intergovernmental	Supranational	Supranational
Council	Intergovernmental	Intergovernmental	Intergovernmental	Mixed
Parliament	Supranational	Supranational	Supranational	Supranational
Court of Justice	Intergovernmental	Intergovernmental	Supranational	Supranational

The balance of power in the integration process has been proceeding as a 'pendulum' (Wallace, 1996). Sometimes, intergovernmentalism, mainly aimed at finding a mere compromise among the conflicting national interests, has prevailed and so the Council has gained power *vis-à-vis* the Commission. On other occasions, supranationalism has been promoted by allowing a somewhat undefined common interest to prevail over the arrangement of national interests. Despite its composition on a national basis, like the Council, the ECJ has aptly endorsed the supranational attitude towards the EU integration process, in line with the scope of its responsibility. This was especially evident during a long stagnation of the integration process, in the 1970s, when political will was lacking and decision making impaired. In fact, after the Luxembourg compromise *de facto* impeded the regular application of the majority rule, decisions were made only after long-drawn-out negotiations whereby consensus was finally obtained by eliminating national vetoes in the Council. The impasse ended when the SEA required qualified majority voting over matters related to the completion of the internal market. The different voting weights assigned to the ministers sitting in the Council reflect their countries' size in terms of population. This criterion underlines the intergovernmental character of the institution, whereby the voters are the member states – on an equal basis when voting under the unanimity rule, or with weighted votes under qualified majority rule – and it is assumed that each government adequately represents the preferences of the citizens of its country. The voting rules within the Council were reformed by the Nice Treaty, which required a triple majority for any proposal to be accepted: (i) the usual majority of at least 71 per cent of the Council's weighted votes; (ii) the majority of two-thirds of member states; which may also be asked to correspond to (iii) at least 62 per cent of the EU population.

In 2002 the constitutional design of the EU integration process was addressed by the Convention on the Future Europe, which produced a Draft Treaty establishing a constitution to be submitted to the Intergovernmental Conference on the Future of the Union.[1] On the one hand, the convention can be regarded as an expression of the intergovernmental method. First, through its procedure: a constitution is not usually provided by an international treaty, although the existing treaties are regarded as having a constitutional status. Second, the EU's identity, formerly seen as a 'process creating an ever closer union' and based on the European Community and its common policies, as stated in Article 1 of the TEU, has become an agreement on policy coordination and on a limited number of 'competences' attributed to the EU by the member states. Third, if a state is defined by the relationship between representation and taxation, then the failure to reach unanimous agreement on the transition to majority voting on fiscal matters is a clue to the intergovernmental grip over political integration. On the other hand, the convention maintains

a supranational attitude in some ways: by incorporating the Charter of Fundamental Rights; by declaring the legal status of the EU, thus adding the legitimacy of the EU institutional entity and that of EU citizenship to the national constitutions; and by trying to find ways of avoiding the limits to integration if a sufficient number of countries so desire. The convention also recognizes the right to secession from the integration process. This proposal, which can be explained by the increased recourse to the majority rule, is meant to alleviate the institutional shock following a possible separation of a member state from the EU.

Modes of integration: enhanced cooperation

Special attention has recently been given to the idea of enhanced cooperation, in that a proposed common policy could be endorsed and pursued by a group of member states, as a second-best solution, rather than waiting for the first best from participation by all member states. The veto power on the launch of an enhanced cooperation, conferred by the Amsterdam Treaty on non-participating countries, has been removed by the Nice Treaty, as decisions and commitments taken in an enhanced cooperation are not part of the *acquis communautaire*, and so are not binding for the opting-out member states. When a common policy is decided by a subset of countries, the opting-out member states lose in terms of vote-trading power as well as in their power to block future initiatives.

The question is whether this should be acknowledged as a damage requiring compensation. In addition, should the opting-out countries be given the right to be consulted on the enhanced cooperation dealings, since they retain the option of a late access and might experience spillovers following its implementation? Moreover, a late participation could be conceived as free-riding wait-and-see behaviour, opting for participation if there are net benefits. A possible solution to the compensation dilemma is to compare enhanced cooperation with the two extremes of no common policy (decentralization) and cooperation among all member states (centralization). When the indicator of the preferences' dispersion is lower for enhanced cooperation than for centralization, and it does not damage the opting-out countries, this mode is the efficient solution. In fact, even if negative spillovers arise for the opting-out countries and compensation should be paid by the enhanced cooperation member states, then the amount is likely to be lower than compensation under centralization.

This reasoning suggests that the more profound question about this method of integration deals with the tradeoff between the subsidiarity principle and the commitment of the treaties to a closer union. Were this principle to be applied, a stop to further integration would become an advantageous policy for member states favouring the status quo. If instead the viability of the

process of integration prevails over the subsidiarity principle, then the correct procedure would be to allow enhanced cooperation and then negotiate to solve the compensation issue at the moment of the delayed accession. There was no compensation for those member states which declined to take part in the two major cooperative enterprises put forward outside the EU framework: the monetary integration process from the EMS to the EMU and the Schengen *acquis* from the Agreement to the Convention and its inclusion in the Amsterdam Treaty. Monetary union was initiated by a subset of countries which were subsequently joined by other member states, when it became evident that the exchange rate agreement was conveying the benefit of a slow but clear deflationary trend. This enhanced cooperation was successful in aggregating an increasing number of member states in the cooperative effort of establishing the public good of monetary stability. However, in the 1990s the prolonged tightness of both monetary and fiscal policies imposed by the need to comply with the Maastricht Treaty and participate in the EMU was probably responsible for the hysteresis effect in the EMU labour markets that yielded structural unemployment. The Schengen Agreement was signed by the founding EC member countries (with the exception of Italy), to eliminate controls at their common borders and allow free circulation. Many other countries joined later, including two outsiders: Iceland and Norway. The UK and Ireland never joined, but have taken part in some activities after the Schengen *acquis* was included in the Amsterdam Treaty. The latter status applies to the accession countries on a temporary basis. In addition to the enhanced cooperation initiatives within the EC, this mode will also apply to the second and third pillars. The CFSP includes the organization of a common military force, either under the Western European Union (WEU) or linked to the North Atlantic Treaty Organization (NATO). Despite the pure public good feature of defence, the paradox might be that a subset of member states launches an initiative for an army with positive spillovers over the remaining non-contributing countries which are opting out for political reasons.

The decision to abolish the veto power by the opting-out member states makes enhanced cooperation a sort of substitute for the extension of majority voting to the most controversial EC areas (tax system, social policies and environment).

Modes of integration: open coordination

Open coordination differs from enhanced cooperation in that all member states are required to participate from the beginning. This mode consists in the formulation of objectives and procedures aimed at boosting convergence to a common standard in a particular domain.

The preamble of the Treaty of Rome places a common standard for labour and welfare conditions among the goals of the EC. The social policy started

in the early 1960s was funded by the common budget with limited, ad hoc funds according to a 'key' included in the treaty, and was especially concerned with immigrants' right to non-discrimination. Yet the European legislation of the Single Act and the Maastricht Treaty was concerned solely with competition policy, by stimulating deregulation, liberalization and privatization processes. Social policies[2] aimed at the catching up of disadvantaged or stagnating areas (cohesion funds and structural funds) have been incorporated into the social cohesion policies. The convergence process was hampered by the depressed growth environment surrounding these policies. In 1999 the launch of the EMU definitively abolished the autonomy of national monetary and exchange rate policies, a process started by the need to comply with the EMS fixed exchange rates. Alternatively, the limited room for manoeuvre left to national fiscal authorities by the constraints of the SGP has undermined macroeconomic stabilization after negative demand and supply shocks, with negative implications for the EU growth rate. The Lisbon European Council (2000) then encouraged growth by the implementation of a series of microeconomic policies coordinated by Brussels and pursued at the national level. Education, training, research and development, social protection and social inclusion were declared appropriate domains for the implementation of open coordination. The aim is to commit national governments to 'modernize the European social model' by voluntary cooperation through the exchange of plans for active policies with such objectives as promoting equality of opportunity in the labour market, employability (to improve skills and labour incentives), entrepreneur spirit (to foster propensity to risk and investment by deregulation), adaptability (lower job protection), and common standards as for social inclusion.

In the future, these active policies put forward by a subset of the EU member states to improve the employment rate might be complemented by the revision of the rigid SGP rules. Since these rules are part of the European legislation and thus overrule national legislation, any attempt by countries who promote open coordination are induced to rely more on further deregulation than on re-regulation processes aimed at increasing employability by training programmes which might conflict with the 3 per cent deficit/GDP ceiling.

Mutual recognition versus harmonization

The principle of mutual recognition was introduced by the ECJ with the Cassis de Dijon judgment in 1979, which allows every good legally produced in one member state to circulate freely in all others. This principle has mainly been used for fostering trade in domains where information is available and consumer sovereignty may prevail, whereas harmonization is pursued when, due to health and safety reasons or severe asymmetric information, the matter

cannot be left to the market. Mutual recognition is unlikely to evolve in complete harmonization because in many domains where sunk costs are considerable, the national interests of the member states mutually conflict and each country refrains from complying with the adoption of another country's national standard as the EU common standard.

In the area of labour market institutions, partial harmonization has been endorsed by the EP encouraging a threshold as for the minimum wage, working day length, vacation periods and so on. The Commission is trying to harmonize the regulation of private pension funds in terms of transparency, portfolio management and precautionary requirements. The Social Charter of 1989 states a commitment by member states on health and safety issues. However, to enable the smooth development of the 'negative' integration consisting of the abatement of any obstacle to the free circulation of goods, services, capital and workers, there has been a virtual stasis in the harmonization on social protection. Mutual recognition in the labour market would mean allowing on the same market the existence of different wages and regulations linked to the rules existing in the worker's country of birth or residence rather than according to the worker's country of employment. In a country with a high wage and high job protection, mutual recognition would be likely to foster a downward pressure on wages and regulations. The ECJ has repeatedly stated that diversity of retributive and regulatory treatment of employees in the same working condition is illegitimate within a country, being at odds with the principles of equal treatment and competition. The convention recognizes European citizenship (art. 8), forbids discrimination and promotes social justice and social cohesion (art. 3).

Mutual recognition has been applied partially to welfare benefits. The welfare system broadly consists of social protection (health care, unemployment benefits, poverty subsidies), and social security (the pension system and invalidity). In the actuarially fair domains, where contributions correspond to benefits, each country is responsible for its own share of social insurance benefits in proportion to the contributions received although only one country actually pays (and is reimbursed by the others via a clearing system). Being actuarially unfair, health care and unemployment benefits are excluded, while retirement benefits can be cumulated across countries (although in the individual balance between contributions and benefits of the pension system there may be a certain degree of redistribution). The present EU welfare policy on the one hand has a limited harmonization – the introduction of a common standard in the area of absolute deprivation, such as a safety net against social exclusion and poverty – and on the other hand expects each member state to provide social protection according to its own tradition and coverage. Another goal is the progressive move towards actuarially fair national welfare systems. The presumption is that the welfare reforms aimed at reducing social expenditures

will improve the balance between contributions (and the share of fiscal revenues devoted to social protection) and monetary and in-kind benefits. The application of mutual recognition to social protection would represent an incentive to lower taxes and welfare benefits, as a country with high taxes and benefits is particularly exposed to system competition.

The fiscal system is a domain where spillovers are more widespread. While economic theory underlines the advantages of centralization in order to internalize the reciprocal externalities, the negotiations over coordination are complex and even steps towards improving harmonization have been only partially successful, since under unanimity voting any harmonization proposal may be opposed by veto. As for personal income, taxation follows the system in the country of birth, even if taxes are paid in the country where the person actually works. This regime may become an incentive to workers' mobility towards countries with high-tax and high-welfare benefits, since workers from low-tax countries would gain by being taxed in their own country and enjoying the high level of welfare services and in-kind benefits of their country of immigration (if citizenship is maintained, this would apply only to wages and salaries). The orientation in the member states has been towards the application of taxation in the country in which income (wages, profits) is received rather than in the country of residence.

Although the creation of the single market should have fostered coordination in the taxation of goods, real harmonization of indirect taxation is still to come. In fact, the member states have resisted the Commission's attempt to switch from the destination to the origin principle in both the excise and the value-added tax rates. The origin principle would amount to extending the logic of the single market to the fiscal domain, as it would imply that goods from all countries should be considered on an equal basis in each market. The main reason for the member states preferring goods to be taxed at the rate of the consumption country is to avoid fiscal competition, with an obvious cost in terms of a decrease in inflows.

As for financial assets, a centralized fiscal regime would respond to the need to internalize spillovers stemming from the high mobility of financial capital. Financial interests and dividends are taxed in the country but non-residents enjoy complete exemption. To progressively introduce the principle of residence, member states are required to exchange information on financial assets in the portfolios of individual residents of another member state. Although the phenomenon of country bias (despite capital liberalization, the so called 'home bias share' that is, the continuing high share of domestic financial assets in the saver's portfolio) mitigates the problem of capital flight, fiscal competition represents too dangerous a threat on public budgets. Under unanimity voting, the veto threat to any proposal of harmonization has brought about a provisional agreement consisting in the progressive conver-

gence across tax bases and the option between two alternative regimes of capital taxation (to limit tax evasion by exchange of information or adopt a system of tax withholding, that is, taxation at the source). Yet, member states whose financial markets attract many capital inflows shopping around for the best fiscal treatment, are resisting the implementation of the agreement.

The democratic legitimacy of EU institutions

In the future developments of the European integration process the question of the EU's so-called 'democratic deficit' is likely to surface. It is important to observe that the term 'democracy' is compatible with a wide range of possible definitions. This indeterminacy enables us to look at this issue from several vantage points.

The most popular appraisal of the democratic deficit undermining the EU integration process argues against EU institutional organization by observing that only the EP has direct accountability to the voters in the national constituencies. The procedural weakness of the decision process originates in the fact that hitherto, two non-elected bodies (the Commission and the Council) have been responsible for a great part of the legislation process. The lack of democratic legitimacy of the market-building process and realization of the four liberties upheld by the ECJ, and the lack of accountability of the ECB, have also been criticized. According to the view of perfect substitutability between the market and the voting mechanisms, a Commission concentrating its efforts towards the completion of the single market, and an ECB committed solely to monetary stability, aim at 'levelling the playing field' for the deployment of competitive forces, and are considered to confer democratic legitimacy on the marketplace. This view can be questioned by arguing that democracy should not be narrowly identified with voting. Many members of democratic institutions, who are appointed rather than elected, derive their democratic legitimacy indirectly through another elected institution that is responsible for appointing them. Just as the Commission's president and members are nominated by the intergovernmental method and approved by the EP, in many European countries there is no constitutional obligation for the government to be composed of elected representatives only. Although from the EU perspective the Council's decisions can be considered barely accountable, nevertheless the Council indirectly derives its democratic legitimacy from the accountability of its representatives at the national level.

However, another problem arises with the increased use of majority voting in the Council, whereby an entire country may be outvoted and bound to comply with the very same legislation it opposed. In such cases, the democratic deficit is strictly interwoven with the intergovernmental method and makes accountability impossible within the Council, while the same problem does not affect the democratic substance of the EP decision making which

abides by the supranational method. The democratic control about how the representatives in the EP interpret their mandate is validated by voting in the national constituencies. On more general grounds, the democratic accountability of an institution-building process with no federal state yet in place is quite different from the democratic accountability within a nation-state. A related question is the accountability of the quasi-jurisdictional bodies. The practice of empowering independent institutions to represent groups of interest is increasingly popular in the EU decisional process. The recognition of political accountability to agency-type regulation has been defended on the questionable grounds that the EU is just a regulatory state (Majone, 1996).

Another, less discussed, aspect of the democratic deficit is economic democracy. A higher level of social cohesion in the EU should be pursued through the improvement of the well-being of 'disadvantaged individuals' (the unskilled, the poor, immigrants, the disabled and so on). Public policy oriented to equality of opportunity could play an important role in fostering economic democracy with positive fall-out on the quality of the public discourse. The record of EU integration on social policy is not outstanding.

The market-building process of negative integration through the abatement of tariff and non-tariff barriers to trade has scarcely been complemented by the market-correcting process of positive integration (Scharpf, 1999). This view stems from two structural changes which are soon to be implemented: (i) after the liberalization of the national markets promoted competition inside the integrated market, the stringency of anti-inflationary monetary policies, the loss of the devaluation instrument after the passage to the EMU, and the anti-state-aid legislation have reduced the financial capacity of the national states to set up fiscal stabilization policies, to provide financial support to the national strategic companies, and to devise education programmes dedicated to improving human capital; and (ii) fiscal competition might also put pressure on the national systems of social security and health care. In contrast to market competition, where the exit of non-competitive firms strengthens the functioning of the market, a competitive tendency across fiscal systems may weaken social cohesion by provoking a welfare system squeeze. In order to avoid capital flight, governments may be compelled by fiscal competition to reduce taxes and transfers in order to keep the rich, reject low-income and needy people and attract capital from abroad. The complementarity of market institutions and social protection institutions in fostering both efficiency and well-being might be undermined. The competition among the welfare systems may have differential effects on investors and the less-mobile unskilled workers and may undermine the long-term efficiency of the EU countries. While the EU has accomplished the objective of liberating the market-building competitive forces, the competition system may ultimately cripple the implementation of the market-correcting legislation. A possible strategy aimed

at fostering positive integration could consist in a coordinated effort by those member states with similar systems of social protection to launch their own harmonization process.

On the political side, the opposition to a federal Europe relies on the argument that no European *demos* exists. This position is controversial. Equating *demos* with *ethnos* fails to distinguish between the pre-political cultural and historical ties and the common political objectives that a group of communities such as the European peoples may consciously set up as a result of public discourse taking place in the domestic and supranational democratic environments. A European constitution could be conceived as the self-recognition of choice of belonging to consensual norms and values (Habermas, 2001). An indicator of the democratic accountability of the EU governance system could be found in European citizens' satisfaction with the consequences on their well-being of the implemented policies. Many domains so far excluded from the competence of EU institutions include several of the most important and controversial issues: the fiscal system, the welfare state, home security and immigration (Scharpf, 2002, 2003). From the consequentialist perspective, a poor EU performance in terms of positive integration indicates that EU citizens consider issues in the realms of domestic security and the welfare state so important for their well-being that policy makers do not dare to remove power from the national legislative bodies lest it provoke an uncertain harmonization outcome at the supranational level.

A theoretical appraisal of EU integration

Institutional design has probably been the most important device for the EU member states to agree on common policies. The design, as a provider of appropriate incentives to the participating countries, can be presented in game-theoretic form as the achievement of the optimal common policy in coordination games among member states where a conflict of interest hinders the Pareto-optimal solution. Many common policies have been transformed from prisoner's dilemma (PD) games to coordination games of mutual advantage just because an institution made the cooperation strategy more profitable than the defection strategy. However, the widening number of both member states and common policies has intensified the integration process to the extent that the increasing complexity of the bargaining has made any agreement a difficult compromise in terms of the distribution of benefits and costs across member states. In general, institutions have been successful in achieving cooperative agreement on the proposed common policies when the member states were under a 'veil of ignorance' concerning the future division across the states of benefits and costs deriving from their implementation.

An example of institution design is the regulation of competition in the manufacturing, public utilities and services markets. Competition policy has

played a relevant role in sustaining the implementation of the Single Market by a variety of community laws, deliberations and judgments regarding the monitoring and enforcement of rules not only for competitive market structures, but also for corporate governance, mergers and acquisitions. The problem is that public ownership in strategic sectors (energy, financial institutions and so on), oligopolistic market structures, and national regulations operating as non-tariff barriers are very common in the EU countries, but unevenly distributed across sectors and member states. In game-theoretic terms, the desire to gain the highest free-riding pay-off – to get the best advantages of competitive European markets, while scarcely cooperating due to a slow domestic process of privatization and liberalization – has led many governments to a ruthless defence of their 'national champions'. The EU institutions have been struggling in order to protect the integration process and avoid a situation whereby an efficient solution to the coordination game of constructing an environment of competitive markets could be jeopardized by the cooperation–failure pay-off matrix of the prisoner's dilemma. The Commission has endeavoured to influence the privatization and liberalization processes inside the EU economies by endogenously influencing these processes through the provision of the necessary institutional infrastructure. The Commission's strategy can be likened to fostering the pay-off structure of a 'chicken game' where the conflict of interests dominating the coordination nature of the players' interaction is offset by the alteration of the PD game with a penalty greater than the pay-off of mutual defection. In Figure 16.1, for the sake of simplicity the pay-off matrix of the chicken game refers to two players only. We can think of the Commission as aiming at reducing the pay-off for defection from further liberalization. When the pay-off matrix of the EU competition game is modified from a PD to a chicken game, the possibility of mutual efforts is strengthened.

Deliberations opposing state aid to strategic sectors, as well as pecuniary sanctions to punish monopolistic practices (collusion in price formation, mergers to control competition, low transparency and low compliance with regulations by companies operating in the service sector and so on) can be

	Cooperate	Defect
Cooperate	4,4	–2,6
Defect	6,–2	–5,–5

Figure 16.1 Pay-off matrix of the EU

formally devised in terms of a penalty for being suckered (–2) greater than the pay-off associated with mutual defection (–5). Since free-riding behaviour is discouraged, in contrast to the PD game, mutual defection ceases to be the rational strategy. However, the Commission's efforts to enhance competition were undermined by the incentives to free-riding continuously created by the externalities across the various privatization and liberalization processes going on in the member states where a conflict of interest is aggravated by the variable degree of public ownership in the service sectors across the states. The generalization of an *n*-player chicken game, with the number of players who pre-commit to defection not exceeding the maximum number of individuals who may free ride on the cooperation of others without causing the non-provision of the public good (Taylor, 1987), may represent the formal explanation of the still lacking rationale for cooperative behaviour.

Three further institutional designs – group asymmetry, procedural technologies regulating cooperative games with conflict of interest among member states and the commitment device of 'tying-one's-hands' – can also help explain the EU integration process.

Group asymmetry

A first aspect of institutional design, which has been studied by both Olson (1965) and Keohane (1984), is group asymmetry as a facilitator of successful cooperation. Collective actions may be asymmetric in the sense that they may be laid out in a form that allows the common policy to deliver both public advantages for all countries and private advantages for single member states. An asymmetric strategic interaction was exploited by the countries launching a fixed but adjustable exchange rate system in Europe (the EMS) to put a stop to the accelerating inflation. The strategic interaction among the European countries – which consisted in the beggar-thy-neighbour situation of competitive devaluations and resulted in the PD suboptimal outcome of stagflation – was transformed in order to establish low inflation as a public good in Europe. Soon after its inception in 1979, the EMS cooperative agreement took the semblance of a hegemonic agreement, due to the asymmetric solution given to the $n - 1$ problem (since among n participants $n - 1$ bilateral parities are formed, just one central bank is left free to conduct its own monetary policy, while the monetary policy autonomy of the remaining $n - 1$ central banks is constrained by the commitment to defend the exchange rate). An assurance game was set up in a leader–follower structure where a cooperation pledge from Germany – the leader country whose central bank was soon singled out as having the best performance in curbing inflation – created the positive net benefits from which the remaining countries, the followers, derived their incentives for participating in the collective action. During the EMS period, with tight bilateral bands, the Bundesbank virtually dictated the

money supply for the whole area, while the remaining central banks were compelled to peg their exchange rate to the German mark in order to take advantage of Germany's reputation as the best performer for low inflation. This privileged position gave the German manufacturing sectors a private advantage to benefit from real depreciation *vis-à-vis* the remaining EMS countries.

Linkage across games

The EU institutions may also make recourse to an institutional design exploiting the interaction between games representing different areas in which cooperation could be pursued to advantage. A first case of linkage deals with the need by the largest member states to dodge the free-riding problem by organizing the implementation of common policies as a nested game with alternative asymmetric positive net benefits. The largest member states were able to overcome the conflict of interests by having one (or a subset) of them contributing or providing the most in a certain game, and another one (or a subset) doing so in another game. A condition for establishing the linkage between the two games is that the common policy proves to be mutually advantageous by the summation of the pay-off matrices. It can be shown that the case of connection between two games is linked to the kind of admitted institutional technology.

In Figure 16.2, the pay-off matrices of two games (I and II) are considered, with the pay-off values stemming from the status quo (SQ), enhanced cooperation (EC) by a subset of two out of three member states (A,B,C) and a common policy (CP). The pay-off values of the game sum show that complete integration (that is, CP) – which compels member states to participate jointly in both policies – is inferior in terms of total welfare to enhanced cooperation. Therefore, the game sum shows that the outcome is different depending on whether or not enhanced cooperation is admitted as a possible

	Game I				Game II				Game sum		
	A	B	C		A	B	C		A	B	C
SQ	20	10	5	SQ	10	5	10	SQ	30	15	10
EC	25	30	15	EC	10	35	35	EC	35	65	50
CP	20	25	20	CP	25	30	25	CP	45	55	45

Key:
SQ = status quo, EC = enhanced cooperation, CP = common policy.

Figure 16.2 Games: pay-off matrices

option. In the example, EC can be a superior alternative to CP in both games (that is, in the sum game). In fact, game I shows a possible enhanced cooperation between A and B, and game II shows a possible enhanced cooperation between B and C. In case of non-admission of enhanced cooperation, the maximizing option is definitely the common policy (the sum of the pay-off values for CP is higher than for SQ). However, CP no longer enjoys a preferential status *vis-à-vis* EC.

In fact, the Nice Treaty has amended the Amsterdam Treaty, whereby an enhanced cooperation could have been stopped by a veto from some other country. A subset of at least one-third of the total number of member countries can now submit a project to the Council which will be decided by qualified majority voting. The EU integration process might be weakened by this procedural change. It is easy to show that once enhanced cooperations are admitted, the CP outcome is undermined. In being compelled to enter the CP in both games, countries B and C lose with respect to entering an EC (in game I, between A and B; in game II between B and C). Even if negotiation costs were nil, the gain of 15 accruing to country A in passing from SQ to CP is insufficient to compensate the other two countries for renouncing EC. A fails to negotiate compliance with a CP because it is unable to cover the compensation of 10 due to B (5 in each game) plus 10 due to C in game II. Country B might then convince C to join in an EC in game II, so that A will have to renounce the CP and join B in an EC in game I. This example demonstrates that the view according to which the EU integration process has gained flexibility, by the availability of the option by a subset of countries of autonomously choosing to start an EC, is flawed. As a matter of fact, the larger set of options magnifies the incentive to exert the threat of power and bargain on compensation.

A commitment to 'tie one's hands'

This commitment consists in the exploitation by a country of the strategic interaction which can often be established between the outcome of a domestic game (the political elections in a society split between reformers and anti-reformers) and the EU game opposed by some member states (see Figure 16.3).

Assume that a common policy among three countries (A, B and C) is expected to enjoy substantial economies of scale only if game I is changed in game I′. In the initial situation of Game I an EC may be preferred by countries B and C, thus leaving country A in SQ. Suppose that country C – the only country in Game I of Figure 16.3 to show a lower pay-off for CP than for EC – could improve its pay-offs by getting rid of some inefficiency which negatively affects its capacity to exploit the economies of scale expected from a fully-fledged CP (for example, poor educational institutions

	Game I				Game I'		
	A	B	C		A	B	C'
SQ	5	5	10	SQ	5	5	20
CP	40	25	20	CP	40	25	40
EC	10	25	30	EC	10	25	35

Key:
SQ = status quo, EC = enhanced cooperation, CP = common policy.

Figure 16.3 Games: tied hands matrices

hindering the level of human capital), as column C′ of game I′ shows. The pay-off values of game I′ indicate that A could gain its much higher CP pay-off and C would prefer CP, to EC with B. Yet, some social groups in country C oppose the needed efficiency-enhancing reform (a structural change in education) because they would be negatively affected by the progressive taxation which is required to fund it. The institutional mechanism capable of rendering the common policy feasible is a change in the 'contract' between the agent (the coalition of parties in charge of the government in country C) and the principal (the constituency). Suppose that the coalition makes the political choice to take a bet in the polls by tying its hands by making the commitment to implement the reform despite the domestic distributive conflict. On the basis of the expectation of being able to convince the opposition voters of the long-term 'mutual advantage' of a higher social welfare, the coalition centres its electoral programme on the structural change in education as the first reform due to the 'national interest' to abide by an external constraint imposed by the supranational EU institutions and take advantage of the strengthening of the integration process.

Had the elections been won and the reform passed, the strategic interaction among the three countries would be modified. While B remains indifferent between CP and EC, the pay-off of both countries A and C is now higher in the case of a CP as C no longer prefers to set up an EC with B. The supranational solution of a CP is finally taken. Because B is aware that the institutional mechanism implemented in the domestic political arena of country C has changed the strategic environment, B is forced to enter the agreement for the CP in order to avoid the worst outcome (the SQ pay-off). However, in contrast to the example given, this top-down approach to the approval of common policies by the national constituencies may be undermined by asymmetric information. A government may have a vested interest in the proposed policy and public officials may be tempted to manipulate the voters' opinion

by distorting statistics and reports produced by governmental research agencies to their advantage.

Notes

1. The Constitution was approved by the European Council on 28 October, 2004 in Rome and at the time of going to press is being ratified by the national parliaments of the 25 member states.
2. The Treaty of Amsterdam included the Social Policy Agreement in Title XI of the TEC, despite the UK opting out of the 1989 Community Charter of the Fundamental Social Rights of Workers (the Social Charter). Its objectives cover employment promotion, proper social protection, dialogue between management and labour, the development of human resources, and combating exclusion (Article 136).

References

De Grauwe, P. (2003), *Economics of Monetary Union*, 5th edn, Oxford: Oxford University Press.

European Council (2000), *Presidency Conclusions*, Lisbon, 23–24 March.

Habermas, J. (2001), 'Perchè l'Europa ha bisogno di una Costituzione?' ['Why Europe needs a constitution'], in G. Bonacchi (ed.), *Una Costituzione senza Stato*, New Left Review, **11** (Sept/Oct), p. 5–26.

Keohane, R. (1984), *After Hegemony: Cooperation and Discord in the World Political Economy*, Princeton, NJ: Princeton University Press.

Majone, G. (1996), *Regulating Europe*, London: Routledge.

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

Scharpf, F.W. (1999), *Governing Europe. Effective or Democratic?*, Oxford: Oxford University Press.

Scharpf, F.W. (2002), 'The European social model: coping with the challenges of diversity', *Journal of Common Market Studies*, **40**: 645–70.

Scharpf, F.W. (2003), 'Problem-solving effectiveness and democratic accountability in the EU', Max Planck Institute für Gesellshaftsforschung Working Paper 03/1, February.

Taylor, M. (1987), *The Possibility of Cooperation*, Cambridge: Cambridge University Press.

Wallace, H. (1996), 'The institutions of the EU: experience and experiments', in H.Wallace and W. Wallace (eds), *Policy-Making in the European Union*, Oxford: Oxford University Press, pp. 37–68.

17 Subsidiarity

Jürgen G. Backhaus

Introduction

The principle of subsidiarity, since it became part of the Maastricht Treaty and thereby of European constitutional law, has received a lot of attention. Its relationship to ecological issues, however, has rarely been explored. Subsidiarity is a perfectly generalizable principle of organization. It can apply to all areas of policy: financial, agricultural, technological, education, defence, economic development and, of course, environmental policy. The principle of subsidiarity is an organizing principle. Taken as such, it is silent about the specific purpose, direction or content of a particular policy. Whatever may be the purpose of the policy, the principle of subsidiarity requires that it be carried out within the smallest viable context in which the objective can successfully be attained. When a task is too complicated for a small unit such as an office or a firm to carry out successfully, that unit has to be augmented to the point where the task can be performed effectively. Likewise, if an organization is too large to handle particular problems successfully, as its procedures may be too cumbersome, as it lacks sufficient detailed information or experiences repeated recurrence of problems it has tried to settle, a different organizational form must be found, preferably an existing one, which is closer to the problem at hand and able to carry out the policy in question. It goes without saying that along with the shift in responsibility will go the access to resources with which to carry out the task.

Aspirations and policy goals: enforced and self-enforcing rules and policies

In constitutional economics, we can distinguish between different types of approaches, notably those that need to be enforced and those that are designed to be self-enforcing. Given the difficulties that enforcement agencies have in enforcing rules laid down and sometimes agreed beforehand, there is always a preference for self-enforcing rules.

Rules have a better chance of being self-enforcing if there is a close match between the functions of a particular unit and the means and responsibilities to serve this function, where the rules have the purpose of ensuring a proper use of the means and responsibilities with a view to performing the functions in question. The crucial issue then is to define that unit which is most appropriate for fulfilling a particular function. Very often in constitutional

public choice, this question is not addressed head on. Typically, we start from the assumption that there is a state or a commonwealth, without exploring the question of which domain this commonwealth or state should actually occupy, and in relation to what other public bodies. Therefore the following section begins with an in-depth discussion of the subsidiarity principle. It is shown, in particular, that the principle has a strong history in European thought, and that it therefore deserves to be relied upon systematically in constitutional economics.

Subsidiarity

The principle of subsidiarity has earned a solid position in European intellectual history. Since, as a result of its incorporation into the Maastricht Treaty,[1] it has become international law, it is worth briefly retracing its intellectual history.

Historical context and theory

This short subsection tries to clarify the meaning of the principle of subsidiarity – which is far from unambiguous – and to explain some of its implications in the context of European unification and the accompanying legal discourse. Perhaps one of the earliest formulations of the principle of subsidiarity, which is at the same time one of the first formulations of a full-scale state welfare programme, can be found in Christian Wolff's *Principles of Natural Law*, published in 1754. There, in section 1022, he states the following:

> In order to lead one's life with decency, it is necessary that the destitutes and the beggars be provided with what they basically need, and that therefore the subjects will not be burdened with too much charitable giving. It has to be carefully determined what the natural law prescribes for charity. Houses of order are to be established where those can work who, although they are able to work would rather prefer to keep begging. Similarly, houses for the destitute need to be established, where those will be fed who cannot work for their own livelihood, and who have no relatives or friends who could take care of their needs: also hospitals, where the destitute will be fed and sometimes cured. Likewise orphanages, where poor orphans will be educated: and finally schools for the destitute, where the children of poor parents will be taught free in matters that are necessary and useful to them.[2]

Several elements of this exposition are particularly noteworthy. First is the scope of the programme of the welfare state, including measures to alleviate the consequences of sickness, unemployment, poor health, orphanhood and, more generally, poverty, in which case access to public schools is guaranteed. Yet the rationale for this government welfare programme, as Wolff explicitly states, is to keep the burden of the welfare taxes to be borne by citizens at a minimum. The welfare state intervenes if and because this is the least expen-

sive way to alleviate the problem. Equally noteworthy is the precise circumscription of the instance under which the welfare state has to intervene: that is, if those in need 'have no relatives or friends who could take care of their needs'. The welfare state, therefore, is only subsidiary to the traditional bonds of family and friendship. Yet it would be wrong to think of the principle of subsidiarity as conceived by Wolff as one built around individualism. It is not the individual who is considered the smallest unit of support (of others), since that would not have been a realistic description of the political economy of his time. Rather, Wolff puts foward a different economic constitution of his welfare state, with the big households comprehending the extended family as the nuclear entity of which the larger economy is composed:

> It can be readily seen that single houses cannot sufficiently provide themselves with what gives satisfaction of needs, comfort, and pleasure, in fact even what is needed for their welfare, nor can they safely profit from their rights and from what they can rightfully expect from others, nor can they be sure to protect themselves against the violence of others. It is therefore necessary, to provide through common forces what single houses cannot get by themselves. For this purpose societies have to be formed. (Section 836)[3]

We realize, then, that in Wolff's conception the principle supporting the unit of the welfare state was the house, the classical household or micro economy; only if it failed did a larger politico-economic unit have to intervene. The interesting question is what form this intervention was supposed to take. The answer is already implicit in Wolff, but a more explicit statement can be found in the writings of another authority on political thought in continental Europe, Robert von Mohl, writing about a hundred years after Wolff:

> Of course, from the need to have anyone provided with basic necessities, it by no means follows that it is the state that has to do the providing. To the contrary, the state will look at this need as it does at any other demand on the part of his citizens. In particular, the state has to lend a helping hand through policing and regulating if this cannot be accomplished through private efforts. As a rule, this will not be necessary if the general provisions have been taken that allow the citizens to earn income and wealth and to use their means effectively.[4]

In brief, Mohl (1844) suggests that the mere existence of needs of its citizens is an insufficient reason for the state to directly satisfy those needs. In general, the state, rather than satisfying the needs of its citizens directly, has the task of creating the conditions under which the citizens can accumulate sufficient wealth to satisfy their needs themselves. Note that this application of the principle is twofold, and twice correct. On the one hand, Mohl correctly identifies the state as the better provider of the (mostly legal)

infrastructure which allows its citizens to prosper and accumulate wealth in order to satisfy the needs of themselves and their dependants. He correctly assumes that the provision by the state will detract from this objective, since the provision by the individual households will be more cost-effective than the provision by the state which would, by necessity, have to be financed through taxes to be borne by the tax-paying households. Consequently, the state is responsible for providing the infrastructure in which the individual households can prosper. Since, on the other hand, provision of welfare services and the satisfaction of needs can be better directed and more effectively rendered by the (larger) households, it is to them that the task also falls, relieving them of the otherwise necessary tax burden. In applying the principle of subsidiarity, it is of pivotal importance continuously to consider both sides of the coin, expenditure and revenue. The determination of the primary and the subsidiary service entity can only be accomplished if both expenditure and revenue factors have been properly and completely taken into account.

A more recent (and perhaps the most widely read) formulation of the principle of subsidiarity is contained in a papal encyclical named after the first two Latin words, *Quadrogesimo anno*, in section 79 of the 1931 encyclical. The text reads:

> And since what an individual can accomplish through his own initiative must not be taken away from him and accorded as a collective task to the state, so similarly it violates the principle of justice that the bigger and higher authority claim a task that smaller communities can accomplish well. This would be extremely disadvantageous and confusing for the entire social order. Every social activity, to be sure, is subsidiary by its own nature and on its own terms. It is supposed to support the different organs of the bigger social body, which however may not absorb or destroy the smaller entities.[5]

In this formulation, the subsidiarity principle appears rather matter-of-factly as a restatement of an old principle in political theory and also church doctrine. Second, it receives an organic twist which blurs its sound economic interpretation. If a community can reasonably discharge its duties, a larger community (of which it is conceivably a part) should not take over these duties as a matter of (ethical) principle. In this formulation, no mention is made of the costs and benefits to the larger and the smaller community, respectively, or to the costs and benefits facing other smaller communities being part of the larger whole. It is not surprising that the formulation from the encyclical, rather than the traditional formulation, has given rise to much controversy and to contradictory applications.

Yet the economic core of the principle is readily crystallized. In order to optimize the performance of the larger political entity, primary liability for the solution of problems lies with the smallest functional unit. This need not

be the lowest functional unit in a hierarchical sense. This smallest functional unit can be a single unit or a group of such units; as shown in the quotation from Wolff (Section 972) these units can be linked by friendship, neighbourhood, a common religion, history or some other such link. The concept of joining several small units may be referred to as 'lateral subsidiarity'.

In the clear case of insufficiency of a particular level, the nearest functional one needs to be found. The search can go either up or down, depending on where one starts. If a particular province, for example, cannot adequately solve, for instance, an environmental problem, it might consider cooperating with (i) another province (lateral subsidiarity), (ii) the nation-state or the European Community (upwards subsidiarity), or (iii) the local communities inside the province, where the problem is most urgent (downwards). The situation might arise where the provincial government takes care of the environmental problem for the smaller communities, whereas the larger communities or those where the environmental problem is most urgent take the matter into their own hands.

It should be emphasized that this principle provides a pattern of thought in order to delineate public responsibilities. The historical survey shows the principle of subsidiarity to be a long-standing constitutional principle which actually dates back to the very beginnings of constitution writing based on natural law systems. The natural law source also makes the principle an integral part of church teachings. In its original formulation, the principle of subsidiarity is a constitutional principle determining the definition and delineation of rights and responsibilities of private bodies, on the one hand, and public bodies, on the other. For example, this is still the case with the role of the principle in the German Basic Law (article two). In the Maastricht Treaty, the principle is used primarily (but not exclusively) to define the relative responsibilities of the European communities, on the one hand, and member states or parts of member states (in the German and Belgian case, federal states or '*gewesten*') on the other. In the tug-of-war between the community and the member states, where the community tries to expand its programmes while these attempts are sometimes met with resistance on the part of at least some of the member states or parts thereof, the principle has assumed a substantial importance. It is enforceable and regularly enforced.[6]

In brief, we can interpret the principle of subsidiarity as an economic principle of functional organization. First, the function needs to be clearly defined; second, the organizational unit needs to be defined in terms of a constitutional economic analysis determining which unit can best fulfil this function. A third step will necessarily consist in finding a legal counterpart to the organizational unit defined in this way.

Lateral subsidiarity

Perhaps the most difficult concept in this context is the notion of lateral subsidiarity. All minds are trained to think in terms of analogies and parallels, and it is difficult to conceive of putting bodies of different legal and political status together into one common cooperation. Yet, in the environmental context where historical and political developments transcend geographical and ecological reasons, such cooperations with a binding legal form are perfectly normal. By way of example, the Lake of Constance is a condominium of one republic (Austrian), two federal German states (Bavaria and Baden-Württemberg) and the Swiss canton of Thurgau. In cooperations between highly centralized countries such as France or the Netherlands and highly decentralized countries such as Belgium and Germany, many lateral forms of subsidiarity can develop.

Conclusion

In this entry the principle of subsidiarity has been developed as an organization principle to be used in constitutional environmental economics. Constitutional economics is that part of economics in which we try to set out the structures in which decision-making processes with respect to ecological units can take place. The principle of subsidiarity has been shown to help in defining relevant organizational units in which political decision making can take place in order to ensure an efficient and sustainable management of specific units.

It is by no means unusual for a central government to have to cooperate with a state government (such as between the Netherlands and the state of North Rhine Westphalia, where such cooperations are common between The Hague and Düsseldorf), a district or even a county or city. In Switzerland and Belgium, many tasks which are normally reserved for the central government fall to the canton (Switzerland) or *de gewest* (Belgium). Germany has had a traditional system, since the reforms by Freiherr von Stein and Carl August Hardenberg in the first decade of the nineteenth century, emphasizing local independence and responsibility, even the residual right of taxation. It is true that central governments are often reluctant to enter into consultations with foreign governments at lower levels. But this is what the principle of subsidiarity requires. Logically, the subsidiarity principle requires the cooperation of the smallest viable units and, if a country does not provide for local authority, that authority has to be found at whichever is the lowest level. In highly centralized countries, such as France and the Netherlands, the principle of subsidiarity will therefore point towards the centre, whereas in highly decentralized countries, such as Belgium, Germany and Switzerland, it points to local bodies.

Notes

1. The most detailed statement of the principle can be found in article 3 B (2) of the EC Treaty: 'In some areas which do not fall within the exclusive competence, the community shall take action, in accordance with the principle of subsidiarity, only if and insofar as the objective of the proposed action cannot be sufficiently achieved by the member states and can therefore, by means of the scale or effects of the proposed action, be better achieved by the community'.
2. Das Leben gehörig hinzubringen, wird auch erfordert, dass man vor die Dürftige und Bettler besorge, was zur Nothdurft des Lebens nöthig ist, und, damit die Unterthanen nicht gar zu sehr mit Allmosengeben beschwehret werden, ist in sorgfältige Betrachtung zu ziehen, was das Naturgesetz von den Allmosen feste setzt (§ 488ff.). Daher sind Zuchthäuser aufzubauen, worinn diejenigen zur Arbeit angehalten werden müssen, welche, ob sie gleich Arbeiten könnten, doch lieber betteln wollen; ingleichen Armenhäuser, worinn man die dürftigen ernähret, die sich durch Arbeiten das nicht zu erwerben im Stande sind, was sie zur Lebensnothdurft gebrauchen, und keine Anverwandten oder Freunde haben, welche sich ihrer Bedürfnisse annehmen könnten; noch ferner Kranckenhäuser, worinn krancke Arme theils ernähret, theils geheilet werden; So auch Waysenhäuser, worinn man arme Waysen erziehet; endlich Armenschulen, in welchen man die Kinder armer Eltern umsonst in demjenigen unterrichtet, was ihnen zu wissen nöthig und nützlich ist.
3. Wir erkennen sehr leicht, dass eintzele Häuser sich selbst dasjenige nicht hinreichend verschaffen können, was zur Nothdurft, Bequemlichkeit und dem Vergnügen, ja zur Glückseligkeit erfordert wird, noch auch ihre Rechte ruhig geniessen, und was sie von andern zu fordern haben, sicher erhalten, noch auch sich und das ihrige wider anderer Gewaltthätigkeit schützen können. Es ist also nötig, dasjenige durch gemeinschaftliche Kräfte zu erhalten, was eintzele Häuser vor sich nicht erhalten können. Und zu dem Ende müssen Gesellschaften errichtet werden. (§ 836).
4. Natürlich folgt aus dieser Notwendigkeit des Besitzes genügender Lebensbedürfnisse noch keineswegs, dass sie gerade der Staat zu liefern habe. Im Gegenteil wird er sich zu diesem Bedürfnis wie zu jedem anderen Verlangen seiner Bürger verhalten. Er hat also namentlich polizeiliche Hilfe nur dann zu gewähren, wenn bei der grösseren Allgemeinheit eines Bedürfnisses auch die Unmöglichkeit vorliegt, dasselbe mittels Privatanstrengung zu befriedigen. Und in der Regel wird Letzteres nicht der Fall sein, wenn die allgemeinen Massregeln zur Erwerbung von Vermögen gehörig getroffen sind und benützt werden.
5. Wie dasjenige, was der Einzelmensch aus eigener Initiative und in seinen eigenen Kräften leisten kann, ihm nicht entzogen und der Gesellschaftstätigkeit zugewiesen werden darf, so verstösst es gegen die Gerechtigkeit, das, was die kleineren und untergeordneten Gemeinwesen leisten und zum guten Ende führen können, für die weitere und übergeordnete Gemeinschaft in Anspruch zu nehmen; zugleich ist es überaus nachteilig und verwirrt die ganze Gesellschaftsordnung. Jedwede Gesellschaftstätigkeit ist ja ihrem Wesen und Begriff nach subsidiär; sie soll die Glieder des Sozialkörpers unterstützen, darf sie aber niemals zerschlagen oder aufsaugen. (Quoted by Herzog, 1963, p. 339–423).
6. However, I am not aware of any rulings on the part of the European Court of Justice. There is extensive jurisdiction by the German constitutional court in Karlsruhe.

References

Herzog, R. (1963), 'Subsidiarititätsprinzip und Staatsverfassung' ['The principle of subsidiary and the constitution of state'], *Der Staat*, **3**, 339–423.

Mohl, Robert von (1844), *Die Polizeiwissenschaft nach den Grundsätzen des Rechtsstaates* [The Police Science According to the Principles of the Constitutional State], Tübingen.

Vaubel, R. (1995), 'A public-choice view of the Delors Report', in R. Vaubel and T.D. Willet (eds), *The Political Economy of International Organizations. A Public Choice Approach*, Boulder, CO: Westview Press.

Wolff, C. (1754), *Grundsätze des Natur- und Völckerrechts* (Principles of Natural Law), Halle: Renger, III, 2.1.

PART IV

LABOUR LAW AND ECONOMICS

18 Labour contracts

Don Bellante

In the United States and in much of Western Europe, only a minority of private sector workers are covered by collectively bargained contracts. Even in the unionized sector where formal explicit contracts exist, the contracts are incomplete, in that they do not specify all of the terms of the employment relationship. Outside the unionized sector, the 'contract' between firm and employee tends to be unwritten, implicit and therefore for the most part unenforceable in courts of law. The primary task of the social scientist is to explain why the implicit, incomplete form of contract is the overwhelmingly dominant form, while the secondary (but more difficult) task is to explain the consequences of the dominance of implicit over explicit labour contracts.

In one sense, the explanation as to why labour contracts tend to be implicit seems obvious: if they are made explicit, they are generally enforceable only on the employer, with employees free to abandon the employment contract at will. This lack of symmetry markedly distinguishes the explicit labour contract from the commercial contract, and employers thus would seem to have no incentive to offer explicit contracts. This explanation is not entirely satisfactory, however, as, even under existing legal conventions, employers might well have an incentive to offer explicit contracts even though such contracts would not bind employees. As long as workers value the security that a formal contract would offer, there are gains to be made by the employer in the form of a lower pay level that would be necessary in order to attract the requisite quantity and quality of labour, in accordance with the theory of compensating wage differentials (Rosen, 1986). Moreover, it is likely that workers would place a very high value on such contractual guarantees. The offering of such contracts would be consistent with profit maximization (and the Pareto-efficiency criterion) as long as the perceived additional costs of formal contracting imposed on the employer are less than the benefits. The general absence of explicit contracts in the private sector therefore must be explained by an overwhelming inability of formal contracts to generate economic value that is sufficient to overcome the transaction costs associated with the offering of explicit contracts. The explanation can be derived from the literature on the economics of organization in general, and on transaction cost economics in particular.

The very concept of implicit contracts, though, originated in the literature of macroeconomics in attempts to explain the downward inflexibility of

wages in the presence of a cyclically induced excess supply of labour (Gordon, 1974; Azariadis, 1975). The argument is that, in most settings, an implicit contract evolves between employer and employee with the employer promising not to take advantage of temporary declines in the demand for labour by lowering wages, as might occur in a spot or auction market. By making this implicit promise, the employer gains an assurance that, during temporary shortages of labour, employees are unlikely to quit to take advantage of a higher spot price of labour, as the employer prepared to offer the spot price will not be prepared to continue it once the temporary high demand for labour is gone. Since the greater stability of wages owing to this implicit promise comes at the expense (to employees) of great instability of employment, the question remains as to why this particular tradeoff necessarily enhances worker welfare. The answer partly involves the asymmetry of information between employer and employee. If the employer offered employment stability instead of wage stability, the employer would have an incentive to 'cheat' by acting as though the demand for labour were lower than its actual level. There is no comparable incentive to cheat if wage stability is offered: given the stable wage, the employer has no incentive to hire any less than the quantity of labour that the current level of demand renders profit maximizing. And when long-run shifts in the structure of demand are sufficiently large to dictate movement to a lower wage, workers presumably will be less inclined to think that firms have exploited workers' imperfect information if the wage cuts have been preceded by large employment reductions (Frank, 1986).

Whether or not the rationale for an implicit promise of wage stability over employment stability is convincing, the rationale still gives no answer as to why an implicit rather than explicit contract is the dominant form. One answer lies in the problem of specification of a formal contract in the presence of bounded rationality (Simon, 1951). That is, the employment relationship is long term, and it is impossible for participants in the relationship to foresee all or even most of the contingent circumstances that may arise over the course of the relationship. Nor is it possible to specify in advance what will be the optimal or even feasible solutions to all contingencies that can be foreseen. Hence it is not possible to specify completely all terms covering all possible contingencies in the contract – any labour contract, whether explicit or implicit, is necessarily incomplete. However, the inability to specify complete contracts puts the explicit contract at a distinct disadvantage. If, for example, lifetime employment were assured provided that certain objective conditions were met by the employee, the firm would have no legally defensible grounds to dismiss the employee if all of those objective conditions were met. The firm would be unable to protect itself against unforeseen contingencies, beyond the control of either firm or employee, that made the assurance of lifetime employment no longer viable.

The greatest impediment that contract incompleteness presents to the formation of explicit contracts, however, concerns those circumstances that are within the control of the employee, and which present incentives for employees to engage in post-contractual opportunism. There may be an understanding, for example, that the employee will 'work hard', and there may be an inarticulate understanding on the part of both parties as to what 'work hard' means, at least under present circumstances. Even so, it is impossible to articulate in a formal contract the specific, court-enforceable meaning of the phrase and a means of proving in a court whether the agreed-upon effort was in fact provided. The employee, knowing this, may face an incentive to live up to all terms of a formal contract that can be clearly specified, but shirk on those terms that are not so specified. Combined with the potential litigation costs to employers where terms are specified, but only loosely so, the prohibitive cost or sheer impossibility of specifying all aspects of the employment relationship in most cases causes explicit contracting costs, in turn, to be prohibitive.

The employee's incentive to engage in post-contractual opportunism necessitates the use of resources by the firm to ensure that employee behaviour is in line with the pre-contractual expectations of the firm. The expenses of closely monitoring employees' activities should thus be seen as the cost of an inability to specify the labour contract completely. There are, however, substitutes for monitoring that, in given circumstances, are more efficient. One is the payment of what are called 'efficiency wages' (Shapiro and Stiglitz, 1984). As with the concept of implicit contracts, the efficiency wage concept also originated in attempts to explain wage rigidity and unemployment at the macroeconomic level. The concept's ability to explain macroeconomic phenomena has been questioned, for example by Bellante (1994), but the concept has found useful applications at the level of the firm. The central idea of the concept is that employers may economize on monitoring costs by providing a rent in the form of a higher than market wage, the payment of which reduces shirking to a greater extent than the expenditure of the same amount on monitoring. The reduced shirking comes about because the rent that would be lost through dismissal makes shirking more costly to the employee.

Of course, substitution of efficiency wages for monitoring cannot be complete, as the fear of loss of the rent is not effective if there is a zero probability that shirking will be detected. Besides reducing the incentive to shirk, the payment of an efficiency wage should reduce labour turnover and its associated costs, a benefit that does not result from monitoring expenditures. More precisely the benefits and costs of both monitoring and efficiency wage payments are respectively subject to decreasing marginal returns and increasing marginal costs. Hence there is an optimal division of expenditure on monitoring and efficiency wages, though in some settings a corner solution

involving no efficiency wage is optimal. It can be argued, however, that not all payments of efficiency wages are the result of a deliberate, optimizing decision of the firm. The ability of unions to obtain higher rents has been well documented, and is extensively surveyed in Lewis (1986). Despite the inability of firms to resist incorporation of these rents into explicit collectively bargained contracts, the firm may, *ex post*, reap the benefits of efficiency wages, even if the benefits do not cover the costs. These results may explain the counterintuitive finding of Freeman and Medoff (1984) that unions raise productivity levels of workers.

Another alternative to monitoring is incentive-based pay. There are severe limits to the use of incentive-based pay, so that its use tends to be fairly narrow. While pay for performance evolved as a partial mitigation of the consequences of the inability to specify labour contracts completely, incentive-based pay would reward effort rather than performance, but effort cannot be directly measured. Performance is subject to influences that are beyond the control of the employee; thus employees are subjected to risk in any real-world incentive pay scheme. Assuming risk aversion on the part of employees, a compensating differential must be paid, usually in the form of a higher average compensation than otherwise would be necessary. *Ceteris paribus*, the larger the random element in performance, and the greater the degree of workers' risk aversion, the higher must be the compensating wage differential. Whether and to what extent an incentive-based pay system is efficient depends also on the responsiveness of employee effort to a particular incentive scheme and, in turn on the responsiveness of firm profits in increased employee effort. It is possible in complex jobs for employees to respond too strongly to incentives: if only some aspects of the desired job behaviour can be brought into the incentive scheme, employees have an incentive to neglect the non-compensated element (Milgrom and Roberts, 1992, pp. 232–7). The extent to which performance can be accurately measured varies quite considerably from one job setting to another, and such measurement difficulties can be a source of workplace friction. Resistance to incentive schemes is also founded in the possibility that these schemes will be used to induce employees to reveal their capabilities to employers, who will then revise or eliminate the incentive scheme. The potential for this information-gathering use of incentive schemes, or in general anything that makes incentive schemes less attractive to workers, will make such schemes more costly to the firm.

The choice of the use of monitoring, efficiency wages or incentive pay (or the optimal mix of the three) obviously depends on conditions peculiar to a given employment relationship, but the conditions described above lead to some recognizable patterns. For example, sales commissions are used far more extensively with outside than with inside sales personnel in light of the relatively lower cost of monitoring the latter. Another pattern is that piece-

rate compensation, though rare, tends to be used most widely in agriculture, where discerning individual contributions is not difficult and where, within wide limits, product quality is not significantly affected by the worker response to the quantity incentive. It is further apparent that, in manufacturing, particularly in small, owner-managed firms, monitoring costs will be comparatively low. Hence in such firms there is a preponderant use of monitoring and the payment of fixed hourly wages that, at least for unskilled labour, contain little or nothing in the way of efficiency wage premia. However, in large multi-plant manufacturing firms, monitoring is less cost-effective because of the more intense principal–agency problem (for example, the monitors must be monitored). Hence the intensity of the use of efficiency wages can be expected to rise with firm size, and this may explain the often observed (beginning with Lester, 1968) positive relationship between wage levels and firm size. In any event, it is clear that the costs of all three devices for coping with potential shirking come under the general category of agency costs, in that they arise out of the incongruence between the objectives of firms and their employees (Okuno-Fujiwara, 1987).

The necessarily implicit nature of most employment contracts, combined with the lack of enforceability on employees, raises other problems of opportunistic behaviour. In most labour markets involving complex jobs, the firm makes a number of investments in its workers. Chief among these is training, which may be either formal or on-the-job in nature. The more important distinction is whether training is general or specific. If the training is sufficiently general in nature, firms will face little incentive to provide it. Hence training in general skills such as accountancy, nursing and so on will rarely be provided by firms. When training would provide firm-specific skills, that is, skills that do not have much value to firms other than the one providing them, a potential 'hold-up' problem exists on both sides of the employment relationship. If the employee were to undertake all of the costs of specific training on the basis of a promise of future pay increases that would fully compensate for the cost of the training, the very specificity of the training would leave the employee with no market alternative to ensure that the promise would be kept. In principle, an explicit contract to pay future compensation would protect workers from being held up. But, for the reasons mentioned previously, the inability to specify completely all the terms of employment renders explicit contracting generally infeasible. Alternatively, the firm that provides and completely pays for specific training is unable to obtain a commitment from employees to stay long enough to ensure that the investment will be returned. The firm facing an inability to obtain commitment could still invest in some provision of specific training, but only up to the point where the return over the expected duration of employment is sufficient to return the investment. As compared to an ideal situation where

commitment could be obtained from workers, this situation will produce less than the optimal (that is, value-maximizing) level of investment in specific human capital.

In most respects, this problem of specific human capital is no different from the general asset-specificity problem faced by firms subject to the hold-up problem in dealing with other firms. For example, a railroad and a phosphate mining firm would both be reluctant to pay for the laying of track that would be of no use other than for hauling phosphate from the location of a specific firm's mine unless some mutual protection against hold-up could be provided. Given bounded rationality, this is a difficult problem, and thus not all potential net value-creating investments will be made. However, the strongest solution to the asset-specificity problem in commercial relationships between firms is not available in employment relationships. The potential hold-up problem between firms arising out of asset specificity can be solved by merger or vertical integration (Joskow, 1985), but, since this solution is not available in employment relationships, other solutions have to be employed. Basically, there are two solutions: reliance on reputation as a contract enforcement mechanism, and the construction of (usually) implicit contracts that are said to be 'self-enforcing'.

It can be argued that the effect of loss of reputation can, under proper conditions, provide a sufficiently strong incentive not to renege on the terms of an implicit contract (Klein and Leffler, 1981). It can also be argued that potential loss of reputation provides much more protection against reneging for the employee than it does for the employer. The value of reputation depends on repeated transaction. An employee will transact with only a few employers over a finite working life; a corporation deals with many employees over its potentially perpetual life. Efficiency considerations, supported by game-theoretic analysis, suggest that implicit contracts tilt the risk towards the party with the least to lose through loss of reputation (Kreps et al., 1982; Kreps and Wilson, 1982). Hence, to the extent that reputational concerns affect the division of specific training costs between employer and employee, they will tend towards requiring the employee to bear the cost, most likely in the form of a below-market wage during the period of training. It should be noted, however, that reputation considerations affect, not just payment for specific training, but the entire structure of implicit labour contracts.

If potential loss of reputation offers some protection for employees, it does so imperfectly; moreover, it offers little or no protection for the employer. Primary protection against opportunistic behaviour comes from structuring implicit relational contracts in such a way that they are 'self-enforcing'. That is, both the employer and employee must face a strong incentive to honour the terms of the implicit contract. Consequently, firms and employees typically share the costs of training, but the firm's share is in large part deferred.

The necessity of self-enforcement is one of the explanations offered by economists for the fact that earnings tend to rise with tenure to a greater extent than can be explained by rising productivity over time. In this explanation, employees accept lower pay during the initial phases of employment (during which a high proportion of the employee's specific training is received) in exchange for increasingly higher wages in future periods. Pensions that are fully funded but for which 'vesting' is deferred are a very common means of deferred compensation that allows firms strongly to induce commitment by workers and to ensure a return on the firm's investment. The United States now in general requires full vesting to take place within five years. This change is seen as nullifying existing implicit contracts and is expected to reduce the willingness of firms to offer training (Bellante and Porter, 1990).

For implicit contracts involving deferred compensation to be self-enforcing, several conditions should be met. First, the deferred payments must be sufficiently large in discounted present value terms to match the present value of alternative employment opportunities that do not offer such training, and at the same time sufficiently small to yield at least a competitive return on the firm's investment in training. The employer thus receives some assurance that employees have a motive not to leave the firm. Moreover, it induces job applicants to reveal their private knowledge of their own *ex ante* expectations about commitment to staying with the firm (Salop and Salop, 1976). The fact that the training provided will cause productivity to rise over time (though not as rapidly as wages) gives the employer a somewhat reduced incentive to renege by firing employees in later stages of employment, in addition to whatever effect potential reputation loss provides. Moreover, replacing older workers with new employees would necessitate additional investment in training and other start-up costs of employment – costs that are sunk, so far as current employees are concerned. Further protection against opportunistic behaviour on the part of employers is provided by explicit seniority policies. And, by making employees more willing to undertake training and accept the associated deferment of wages, such policies also serve the purpose of firms by reducing the lifetime wage premium that must be paid to such workers.

A second condition that must be met is that employees will not 'over collect' from the firm by staying with it too long. Mandatory retirement policies can be seen as meeting this condition (Lazear, 1979). The desires of national governments in recent years to prohibit mandatory retirement have curtailed the use of this device. Alternatively, pension plans can be structured to induce retirement by providing a schedule of benefits whose expected present value begins to decline at the age for which the firm desires retirement to take place. Lazear (1983) has provided evidence that a large fraction of firms paying pensions do in fact structure them in this manner.

Despite all that can be done to make implicit contracts self-enforcing, market conditions may change so as to make their abrogation seem necessary, perhaps even a prerequisite for the firm's survival. However, the costs to the firm of abrogation are obvious and likely to be severe. A question worthy of further investigation is whether and to what extent mergers and spin-offs are in some instances devices for minimizing the costs of abrogation through unilateral 'rewriting' of the implicit contract.

References

Azariadis, C. (1975), 'Implicit contracts and underemployment equilibria', *Journal of Political Economy*, **83**, 1183–202.
Bellante, D. (1994), 'Sticky wages, efficiency wages and market processes', *Review of Austrian Economics*, **8**, 21–33.
Bellante, D. and P. Porter (1990), 'A subjectivist economic analysis of government-mandated employee benefits', *Harvard Journal of Law and Public Policy*, **13**, 657–87.
Frank, J. (1986), *The New Keynesian Economics*, New York: St. Martin's Press.
Freeman, R. and J. Medoff (1984), *What Do Unions Do?*, New York: Basic Books.
Gordon, D. (1974), 'A neo-classical theory of Keynesian unemployment', *Economic Inquiry*, **12**, 431–59.
Joskow, P. (1985), 'Vertical integration and long term contracts: the case of coal-burning electric generating plants', *Journal of Law, Economics and Organization*, **1**, 33–79.
Klein, B. and K. Leffler (1981), 'The role of market forces in assuring contractual performance', *Journal of Political Economy*, **89**, 615–41.
Kreps, D., P. Milgrom, J. Roberts and R. Wilson (1982), 'Rational cooperation in the finitely repeated prisoner's dilemma', *Journal of Economic Theory*, **27**, 245–52.
Kreps, D. and R. Wilson (1982), 'Reputation and imperfect information', *Journal of Economic Theory*, **27**, 253–79.
Lazear, E. (1979), 'Why is there mandatory retirement?', *Journal of Political Economy*, **87**, 1261–84.
Lazear, E. (1983), 'Pensions as severance pay', in Z. Bodie and J. Shoven (eds), *Financial Aspects of the United States Pension System*, Chicago: University of Chicago Press, pp. 57–85.
Lester, R.A. (1968), 'Pay differentials by size of establishment', *Industrial Relations*, **7**, 57–67.
Lewis, H.G. (1986), *Union Relative Wage Effects: A Survey*, Chicago: University of Chicago Press.
Milgrom, P. and J. Roberts (1992), *Economics, Organization and Management*, Englewood Cliffs, NJ: Prentice-Hall.
Okuno-Fujiwara, M. (1987), 'Monitoring cost, agency relationships and equilibrium modes of labor contracts', *Journal of Japanese and International Economies*, **1**, 147–67.
Rosen, S. (1986), 'The theory of equalizing differences', in O. Ashenfelter and R. Layard (eds), *Handbook of Labour Economics*, New York: North-Holland, pp. 641–92.
Salop, J. and S. Salop (1976), 'Self-selection and turnover in the labour market', *Quarterly Journal of Economics*, **90**, 619–28.
Shapiro, C. and J.E. Stiglitz (1984), 'Equilibrium unemployment as a worker discipline device', *American Economic Review*, **74**, 433–44.
Simon, H.A. (1951), 'A formal theory of the employment relationship', *Econometrica*, **19**, 293–305.

19 Company board representation

Jürgen G. Backhaus

Introduction

Corporations can have widely differing forms. Against the backdrop of the peculiar German system of codetermination, we can illustrate the law and economics rationale behind different types of board composition. Boards are split into two, the managing and the supervisory boards. Works councils elect the powerful chairmen, who can end up on supervisory boards and, more importantly, the steering committees. In addition, investment bankers may join forces with employee representatives in controlling management appointments. How did this system come about, which events shaped it, and why is it being continued by globally operating companies?

Codetermination as an interlocking system

Codetermination is an interrelated set of institutions which consists of five major elements (Table 19.1). At the plant level, it involves the representation of workers in works councils. Works councils need to be heard in many matters referring to the daily practice of work and they need to be involved in

Table 19.1 Codetermination as an interlocking system

	Works council	Collective bargaining	Two-tier board	Bonding of investment	Competition between locations
Works council	x	1	2	3	4
Collective bargaining	5	x	6	7	8
Two-tier board	9	10	x	11	12
Bonding of investment	13	14	15	x	16
Competition between locations/ *Gewerbesteuer*	17	18	19	20	x

discussions with respect to such issues as hours, time and form of payment, and the hiring and firing of individual workers. They need to agree on economic matters that concern them existentially, in particular if a plant is to be closed. In such a case, a social plan needs to be implemented. Obviously, the works councils have no say over what is being produced, what the prices are, what kinds of investment are being made, and so on. But what is crucial is that they determine the exit costs and thereby the entry costs of a potential investor.

Next to the works councils, codetermination provides for the representation of labour representatives on the supervisory boards of corporations. The supervisory boards appoint management and supervise management activities. Under different codetermination acts, workers have either one-third or one-half of the seats of the supervisory board. In the supervisory boards, the long-term investment plans of companies are determined. Third, codetermination may involve the appointment of one member of the managing board of the company, who is in charge of personnel affairs and has the special trust of the labour representatives on the board. In this, and only in this sense, labour can participate in the managing of some corporations, but the personnel director will always be in the minority. Under German corporate law, the board acts jointly.

The system has become the centre of political disagreement with respect to efforts at forming the type of a European Company 'Societas Europea' (SE). The German Company Board representation would have to somehow fit in with the structure of the European Union.

The acts establishing codetermination in Germany were passed by the German federal parliament in 1951, 1952, 1965 and 1976. The 1976 act, for instance, which generalized the codetermination model that had originally in its strong form been limited to the so-called 'mountain' industries (mining and steel), carried a majority in parliament of more than 90 per cent. In the context of European Union harmonization attempts of corporate law, Germany is trying to introduce codetermination as an element into European corporate law; this is being resisted by some other European Union members. The reason for Germany's stance in this matter is the fact that codetermination is part of the German consensus and departing from it is not supported by any one of the stronger political movements in the country.

From a law and economics point of view, the question remains why Germany opted for this peculiar form of company organization, which is at substantial variance with the classical model of the business firm. The answer to this question has to be given in two parts: the first needs to address the issue of why codetermination was introduced in the first place; the second needs to address the issue of why codetermination is being clung to as a generally desired form of corporate organization.

In 1949, Germany still lay in ruins. The reconstruction effort had not really taken off, because the institutional prerequisites (stable money, stable legislative framework, labour harmony, a supporting tax system and consistent economic policy) were not yet quite in place. Codetermination was one aspect of this package of implementing the institutional prerequisites for the reconstruction of post-Second World War Germany. Since the physical plants had been largely destroyed or were still subject to dismantling, productive capital was actually embodied in human capital. The human capital, of course, needed to be won over in any investment decision. It is for this reason that human capital was systematically introduced into the organized forms of capital representation in the corporation. Only in this way did it seem possible to reinvest unpaid wages and dividends into the company. In 1951 and 1952, pursuant to the government programme, West Germany introduced the system of codetermination as we now know it. It was chosen in preference to a number of different organizational alternatives which would all have been much less compatible with the functioning of a market economy.

In 1965, with the reform of the act on joint-stock companies, codetermination was again expanded into sectors not hitherto covered. It was further expanded in 1976. It is difficult to conceive of an interest group-driven scenario of parliamentary legislative decision making that could explain the consistency of the record and the large supporting majorities. The system has clearly proved workable, but it has interesting implications with respect to which industries can expand in Germany and which are driven abroad, typically with the active cooperation of the labour representatives on the supervisory board – Volkswagen do Brasil probably having been the first highly visible example of the approach. In a nutshell, the system is geared towards company structures with a near parity of human and physical capital. Only some lines of production lend themselves to this model, and herein lies a clear strategic choice of the codetermination legislation, although it may not have been consciously made.

Codetermination as a model cannot be superimposed on any country. Legislators keen on copying Germany's legislation should carefully consider Böhm's (1951) argument in order to find out whether the conditions that prevailed in Germany and made codetermination a worthwhile legislative experiment would prevail in the candidate country as well. Otherwise, Böhm's warning of a collusion of capital and labour against the consumer might well be heeded with profit.

In reviewing the parliamentary debate in 1951, one is surprised to note how little time was spent in hammering out the details of the legislation. Indeed, much of the relevant knowledge was probably shared by the members of parliament, but it was going to take almost 30 years before, in 1979, the German Constitutional Court emphasized the extent to which the different

institutions making up the German codetermination system are interdependent, even interlocking.[1] The system makes sense only because its five different pillars support each other: taking one element away can result in serious shortcomings. The five elements have been listed in Table 19.1, and their interdependence will be discussed in turn in the sequence of the numbers indicated.

The works councils are the oldest element. They were introduced during the First World War originally in the ammunition industry in order to allow workers, who threatened to strike, to participate in decisions about the organization of their working environment, schedules and the like. Today the works councils are an omnipresent part of the management structure of the German company. They are also present in the public service with slightly different functions, since civil servants in Germany do not have the right to strike.

1. The works councils relate to the institutions of collective bargaining in terms of the division of tasks. All those aspects works councils deal with are, in principle, *not* subject to collective bargaining. For instance, collective bargaining will set the base wage. Works councils will discuss with management when and how the wage is paid. This division of the task is extremely important since works councils, in order to be effective, have to be able to reach their goals by convincing management of the superiority of their proposed solution in the interest of the entire company. Works councils participate in decision making in the sense that they cannot be ignored: rarely can they take decisions on their own. Only in one case (when the firm is to be liquidated) can works councils, by withholding their agreement to the social plan, impose serious damage on the firm's owner. The institution of collective bargaining, on the other hand, by definition uses powerful sanctions on either side. The workers can strike, the companies can lock them out. In both cases, serious losses can be imposed. Unions may not be able to strike because they may not be able to pay the striking workers their compensation, which typically they are understood to do. This division of tasks is extremely important, in particular as we compare the German situation with the American one. In the United States, collective bargaining will also cover work rules, a practice which can result in very formalistic agreements about which disputes are likely to arise. On the other hand, union leaders used to collective bargaining may find it difficult to cooperate in company boards when given the chance, as in the case of Eastern Airlines (under the chairmanship of Frank Bormann).
2. The works councils are often the recruiting ground for the labour member in the supervisory board of German cooperations. It is here we

realize that the two-tier system is a precondition for the possibility of codetermination. If the works councils were to send representatives into the managing board, they would end up negotiating with themselves. Incidentally, the labour directors mentioned above were always recruited from the outside in order to avoid such a conflict. By being able to rely on the works councils, the supervisory boards have a second channel of information at their disposal, the first one being management which, as they also have to control it, may be withholding information that could be damaging. It is for this reason that the system of codetermination can strengthen owners' rights in the firm, at the expense of managers. In this sense, codetermination is one conceivable answer to the problem involving the division between ownership and corporate control.

3. Bonding is a device to protect sensitive investments in specific assets. When highly specific human capital is deployed in the production process, the owners may be interested in seeking protection against the loss that can occur if inappropriate management decisions are being taken. The works councils can provide such protection, on the one hand through the day-to-day consultation with management, on the other hand through their links with the supervisory board. Here both labour and capital can seek to protect the value of the investment by monitoring management that, through its operation, determines the marginal productivity of both capital and labour.
4. In Germany (as in the United States) local communities depend for their tax revenue on the profitability and presence of corporations within their jurisdiction. There exists a symmetry in the interest of the local communities and the employees of corporations, who typically are tied to these communities by virtue of their home ownership, family and other ties, and generally their roots. German workers tend to be much less mobile than their American counterparts. Thus, the works councils also act in the interest of the local communities from which they come in guarding the presence of the corporation at its current location. This, by the way, does not prevent codetermined firms from making sometimes large investments overseas, if these investments are perceived by labour representatives as implicitly enhancing the stability of their own jobs.
5. By virtue of the division of tasks between the system of collective bargaining, on the one hand, where employers' associations and the trade unions negotiate the base wage, and the system of works councils, on the other, where the day-to-day operation of the working routines is being discussed, the works councils are relieved of those issues which necessarily would result in acrimonious disputes. Thus the two systems support each other.

6. Since labour representatives sit on the boards of corporations, the unions possess much more detailed information about specific companies than their American counterparts. They are therefore unlikely to overshoot in their bargaining, and are more inclined to take a long view and maximize lifetime earnings rather than per annum wages.
7. The first German post-Second World War parliament gave the hair-thin majority of the governing coalition 'an overwhelming majority', a very important demonstration of a consensus living in the post-Second World War West German democracy. The families of the workers typically being tied to the location, title needs to be given to all these stakeholders. Unions make these investments, because the investment they make in wage concessions can be guarded through the codetermination structures.
8. Very often, trade unions also guard the interests of the corporation in the towns where they operate. They thereby help in creating a business environment that is hospitable to the corporate activities.
9. The two-tier board can function effectively because, through the works councils, information can be gathered in a systematic manner. The board is able to collect information which sometimes is not even available to management.
10. Since the supervisory board and the managing board are separate, the managing board is free to bargain with the unions, who are not represented there. A single-tier system would not allow for this separation of interests.
11. The two-tier system, again, helps in safeguarding investments against excessive demands from either side. This is true for both labour and capital.
12. Although, through the corporate city tax (*Gewerbesteuer*), the German city is in fact a stakeholder in its corporations, and although the city is expected to furnish services from basics to schools, sometimes even theatres and symphony orchestras, no attempt seems to have been made to repeat the features of the Hesse Organization Act and grant the city seats on the supervisory board. The instrument that is used is typically a plain contract. This is a public contract, but many communities preferred to contract privately through development boards (*Entwicklungsgesellschaften*).
13. Since bonding of investments is one of the key strategies used under codetermination, the works councils chairmen have a specific role that is far beyond that of a shop steward, for instance, in the British system. They are selected, not for their aggressive negotiation skills, impressive confrontational methods or rhetoric, but rather for guaranteeing smooth working relationships and an avenue for resolving the

many conflicts that necessarily develop when people have to work together, certainly if highly committed and highly skilled people are involved.

14. Also collective bargaining is conducted differently when bonding of investments is a part of the set of strategies. Tactics need to be more subtle, not only in view of strike funds, but also in view of the human capital investments at stake.
15. Bonding is an attractive alternative strategy because the supervisory board allows for keeping an eye on the specific investments made.
16. Since bonding strategies are available, cities, in turn, are able to make long-term commitments, which after all they have to finance out of tax revenues. These investments can go far beyond what is usually practised in the United States by, for instance, a county development board. For example, establishing schools, theatres, public swimming pools and the like requires not only the initial investment, but also the continued commitment to cover the sometimes substantial operating costs.
17. Cities, through the works councils, often have an avenue by which they can put their priorities on company board agendas and demand firm decision making as well. This is important when issues of city planning and work organization overlap (as with bus schedules). Short communication lines can substantially cut transaction costs.
18. When a strike is being called, the mayor holds his breath. The city can lose twice, not only because of the reduced income and therefore likely increased level of entitlements that need to be paid out – in Germany the entitlements have to be paid by the cities, not by the state or the federal government – but also because depressed corporate revenues mean depressed corporate tax revenues, putting the city budget into a squeeze. It is obvious that through various channels a moderating influence is present.
19. Most cities in Germany operate a bank, the so-called *Sparkasse*. This is a leftover from traditional cameralist policies. These banking institutions also get involved, as they can in Germany, with locally operating companies, since commercial and investment banking can be done through one and the same institution in the German banking system. In this way, the city can exert influence through the supervisory board if, as is customary, the major local banks and firms take seats on this board. The city mayor, in turn, typically appoints the bank manager. Often, the city also relies to a not insignificant extent on the bank revenues for its own budgetary requirements.
20. Since the interests of the city and the workers (its citizens) overlap to a large extent or are even fully parallel, the city where the company is operating is also interested in keeping the strategy of bonding invest-

ments viable. It tries to create an environment that makes the strategy attractive.

In this (as well as in many another unmentioned) way, the five pillars of codetermination support the entire structure in an interlocking fashion. Removing any one of them will lead to multifarious consequences, and these consequences are very difficult to predict beforehand. Codetermination creates a specific actor in the labour market that is strongly committed to human capital formation and to the location where the works are sited. This is a different position from that of the typical American or multinational corporation, which tends not to maintain close ties to any specific location. The consequence is that codetermination tends to open up a different arena where, next to labour market exchange, decisions can be taken. These decisions refer to shaping the production function, not to the formation of wages. The latter distinction also points to the reason why codetermination has to be the result of legislative decision making and cannot conceivably be the outcome of collective bargaining. It is not a bargain, it is rather a constitution in which production functions can be shaped.

The paradox of monitoring

The theory of the classical firm assigns to one factor of production, typically capital, the role of deciding the monitor of, first, its own performance and, second, the performance of the other factor, that is, labour. Monitoring is in itself a productive task, since it prevents shirking, that is, a suboptimal use of the means of production, typically in the interest of one of those involved in the process of production but not coincident with the purpose of the operation. Oliver Williamson calls this the problem of subgoal pursuit. The marginal productivity of the factors of production, both labour and capital, depends on managerial services, and in particular the marginal productivity of labour depends on managerial inputs, as the marginal productivity of management depends on the willingness of workers to be efficiently managed. The marginal productivity of either one of these factors determines, ultimately, its remuneration. Both wages and managers' earnings depend, ultimately, on their marginal productivity, measuring errors notwithstanding. Hence, when in both line work and management there is an element of discretion, a system of monitoring needs to be found that does not depend on any one of these factors.

If we look at the situation as it existed in West Germany in 1949, when the newly elected government with its tiny majority announced its intention to pass the legislation that would introduce codetermination, setting up an efficient governance structure amounted to a substantial organizational and legislative challenge. Six elements of that situation stand out:

1. To begin with, German companies in 1949 had virtually no access to international capital markets. Since there were no local capital markets either, investments had to be made out of profits not yet realized. These profits, in turn, obviously depended on the size of investment. The larger the investment could be, the larger could be the marginal and total productivity of labour, and hence the wage. However, no capital markets being available and investments having to be paid out of gains not yet being made, the larger the net profits of the firm the larger the wage, but the larger the wages fund, the smaller the net profits of the firm and hence the smaller the investments to be made and hence the smaller the marginal productivity of labour. This strange paradox called for a resolution.
2. Owing to both the war damage and the ensuing dismantling campaigns, many firms found themselves with insufficient or no equipment. The firms' book assets having been reduced by war and ensuing inflation, currency reform and depreciations to next to nothing, the real assets consisted in the firm's history, a site which often resembled a pile of rubble, and a willing workforce. The history also included the memory of how to produce, of how to organize, how to link up with former suppliers and customers, and how to work with the local government.
3. Some of these former customers, however, were not currently available. German companies had to re-establish their previous international customer relationships, itself a time-consuming process, and a difficult one as, because of the war, technological advances had not been realized in civilian production.
4. The reputation of former owners and managers was often tainted. Immediately after the war, some of the owners had been detained in camps, and what had been a business relationship with the government was now seen as criminal complicity. Hence firms often also lacked the credit that they had formerly enjoyed through the reputation of their owners. In part, the new managers, who had to a large extent also been recruited from the labour movement, formed the reputation on which credit could be extended.
5. On the other hand, firms in 1949 invariably met strong market demand, as citizens were trying to rebuild their lives, as either refugees or survivors of the bombing raids. Many people had lost their household possessions and were eager to rebuild a normal life. The strong materialism characteristic of the 1950s and 1960s in Germany bears ample witness to these initial shortages.
6. Finally, political life was reconstituting itself from the bottom up. Only in 1949 had a limited federal government been formed. In 1945, first local, then regional and later state governments had been formed. Only

> in 1948 was the currency re-established with the consequent emergence of markets. In 1949, the war and post-war period was still very much dominating people's minds. The experience of the newly created market, in many ways the first free market since 1914, was still very fresh.

What, then, is the appropriate governance structure of corporations to adopt in order to rise out of such precarious conditions? It is obvious that the classical firm, under these conditions, is not the appropriate answer. The monitors themselves needed to be monitored.

The solution of bonding

Under the above circumstances, both owners and workers with their representatives had an interest in rebuilding the companies and creating an investment fund through wage restraint. Thereby, of course, the workers had to become co-owners, but the solution suggested by the Free Democrats of gradually giving them ownership certificates such as stock in exchange for wage concessions was not a feasible alternative, since comonitoring of the investment had to start right away. How could the workers have accepted stock in worthless companies that could only gain in value through their commitment to create credit? The solution suggested by the (legally trained – all four had a doctorate) managers in 1947 was a stroke of genius. By building on the concept of the two-tier board, labour could be invited to assume an equal position on the supervisory board and management could proceed, as professionally as possible, under the joint supervision of labour and capital. In fact, the investment of the workers had been bonded. Indeed, only through the bonding process was it forthcoming. In this sense, codetermination can be assigned an important place in explaining the 'economic miracle' West Germany lived through in the 1950s and 1960s.[2]

The stark consensus that marked the position of all parties in the parliamentary debate thereby finds its logical explanation. The situation required a new solution which was indeed hit upon, and there was no relevant alternative. This, however, may be said not to have been true afterwards. To that question, finally, we now turn.

Path dependence

Ultimately, the question needs to be addressed as to why the institution of codetermination was not only continued but also extended to cover the entire German corporate landscape (of companies having more than 1000 employees). The answer has to be given in different parts. To start with, two common misperceptions have to be avoided. The first misperception concerns the firm's capital access at the time when codetermination was introduced. Henry Hansmann (1996, p. 2) writes:

> I can, however, hazard a theory of my own: that the firms involved would probably have to raise new capital at some point, and that they might want to sell large amounts of stock for the purpose – stock that would dilute the worker's earnings and control.
>
> Codetermination assured them 50% control, and through that control they could assure themselves wages yielding an appropriate fraction of the companies' earnings.

This alternative theory is factually wrong on two points. First, companies for some time to come had no access to capital markets and therefore had to rely on self-financing. International capital markets were not accessible to German corporations (at the time of the deliberations still suffering under denunciations) and domestic capital markets were extremely thin during 1950–55. Second, codetermination did allow workers control over the investments in terms of siting and (through co-decision making in the works councils) the technologies used, and therefore could indeed have been used in order to shape any specific capital increases through issuing new stocks or bonds. Codetermination, however, does not allow any access to decisions over wages.

Hence, in order to explain the further extension of codetermination, we have to stick to the specific German case, which is not generalizable as it depends on a historical starting point not to be found anywhere else in the industrialized world. The first part concerns the act of 1965, which was part of the corporate reform act needed because the national socialist government had written the leader principle into corporate law. This complicated reform package also included all those elements of codetermination that the speaker of the Free Democratic Party in the parliamentary debate had outlined in 1951, including the one-third representation of labour in the supervisory boards. That act covers every joint-stock company with more than 1000 employees.

The act of 1976 was different. It was the result of the big coalition between the original two large parties that had in 1949 supported codetermination, although only one of them had been part of government. The price for taking part in government that the Social Democrats exacted was the extension of codetermination over all industries beyond coal and steel. Here the echelon was finally decided to be a minimum of 2000 employees. However, Karl Biedenkopf as early as 1965 had found that codetermination in the industries covered worked, in that it did not lead to frictions and, conceivably, had benefits. Since 1965, companies had been aware of the pending legislation. Easy escape routes were built early on. The act requires specifically corporate forms which can be changed, it is only applicable to corporations of a specific size which can be undercut by reproportioning or regrouping enterprises, it requires a location of the seat of the company within the German jurisdiction which can be easily avoided (for instance by locating headquarters in Ger-

man-language Switzerland) and even a simple change in the by-laws of the supervisory board can render codetermination under the act of 1976 meaningless. All of these activities took place to a certain, but not very large, extent.

When the act was passed in 1976 with more than 90 per cent of the vote in parliament, and without even a hint of union action, it was widely perceived as a logical evolutionary development in German corporate law. No perceptible effect on the stock exchange has ever been documented. There was litigation surrounding the implementation and there was litigation before the constitutional court in order to test the future of such legislation. However, the general perception is that there is not going to be a future in this legislation. Codetermination as a part of German corporate law is firmly in place, and the only point of further argument is to defend it against future developments in corporate law initiated by the European Commission or the European Parliament. Since, under the Maastricht Treaty, that treaty has to be read through the German Basic Law, and since codetermination has been tested against the German Basic Law, through the constitutional court case against the codetermination act of 1976 and the ensuing verdict, codetermination has been chiselled in stone for the time being and under the restrictions mentioned above.

The further development of codetermination after 1951 can only be partly seen as path dependent. The other and more important part is the experience the German industry has gained with the institution of codetermination, and that experience has not varied much from the original conditions. Since in Germany the different stakeholders are expected to invest heavily in the company (the stakeholders, including the workers, the local community, capital owners), of course title needs to be given to all these investment commitments. Since such title is difficult to imagine, codetermination may have proved to be the most adequate vehicle to protect investments that can be jeopardized when industrial strife, antagonism, intolerance or managerial hubris endanger specific assets, these being either specific human capital investments or specific capital investments.

Conclusion

This entry has tried to render an account of the history of German codetermination from the point of view of the reason for its actually coming about. It has suggested that the solution, although innovative, was virtually the only one German decision makers in industry, trade and politics could hit upon. As it turns out, the solution was suggested, independently or not, by four academically trained industry leaders in the industrial heartland of West Germany. Although the implementation through various legislatures of the solution required a circuitous route, the final outcome seems logical enough.

However, it is an outcome that has a very specific origin. It is not clear under which circumstances German codetermination could be an attractive or even an interesting model for corporate America.

The legislation on codetermination created a type of player in the labour market which was substantially different from employers operating under different rules. Codetermination emphasizes the specific location of the works, it has an impact on the product mix, it is not irrelevant with respect to technology choice, it requires heavy investment in research and development, as well as a long-term view of human capital formation, where the expectation is that the human capital will remain with the firm as long as the firm exists. A player in such circumstances will be more steady in the labour market than a player not similarly constrained and motivated. Codetermination may create, in the short run, fewer jobs, but can be expected, in the longer run, to sustain more of them than the comparable corporation.

Notes

1. BundesVerfassungsGerichtsEntscheidung (BVerfGE) 50, 1979, 290.
2. An American commentator has cast the matter differently: the workers, afraid of future stock issuance and a consequent dilution of their shares, preferred the statutory board seats to shares. This may be a good start for didactical purposes. Yet German companies tend not to rely on stock issuance for their expansions, not even today, and certainly not then (in 1951), when there were as yet no viable capital markets.

Bibliography

Ackerman, Bruce A. (1983), *Reconstructing American Law*, Cambridge, MA: Harvard University Press.

Backhaus, J.G. (1979), *Ökonomik der partizipativen Unternehmung I* [Firms with Labor Participation], Tübingen: Mohr/Siebeck.

Backhaus, J.G. and H.G. Nutzinger (eds), *Co-determination: A Discussion of Different Approaches*, Heidelberg and New York: Springer.

Bainbridge, Stephen (1996), 'Participatory management within a theory of the firm', *Journal of Corporation Law*, **21** (4).

Bentham, Jeremy, *An Introduction to the Principles of Morals and Legislation* (printed 1780, published 1978); *Of Laws in General* (written before 1789, published 1970); *The Works of Jeremy Bentham* (11 vols), ed. J. Bowering, Edinburgh, 1838–43; *The Collected Works of Jeremy Bentham*, ed. J. Burns, London, 1968.

Böhm, Franz (1951), 'Das wirtschaftliche Mitbestimmungsrecht der Arbeiter im Betrieb' ['The worker's economic right of co-determination in the firm], *ORDO*, **4**.

Buchanan, James M. and Dwight R. Lek (1982), 'Politics, time and the Laffer curve', *Journal of Political Economy*, **90**, 816–19.

Frey, Bruno S. and Hans Jürgen Ramser (1986), 'Where are the limits of regulation?' [Zeitschrift für die gesamte Staatswissenschaft], *Journal of Institutional and Theoretical Economics*, **142**, 571–80.

Hansmann, Henry (1996), 'Comments', Conference on Employees and Corporate Governance, Columbia Law School, 22 November.

Hart, H.L.A. (1982), *Essays on Bentham, Studies in Jurisprudence and Political Theory*, Oxford: Clarendon Press.

Michelman, Frank I. (1980), 'Constitutions, statutes and the theory of efficient adjudication', *Journal of Legal Studies*, **9**, 431–61.

Priest, George L. (1980), 'Selective characteristics of litigation', *Journal of Legal Studies*, **9**, 399.
Priest, George L. and Benjamin Klein (1984), 'The selection of disputes for litigation', *Journal of Legal Studies*, **13**, 1–55.
Tullock, Gordon (1980), 'Efficient rent seeking', in J.M. Buchanan, R.P. Tollison and G. Tullock (eds), *Toward a Theory of the Rent Seeking Society*, College Station, TX: Texas A & M University Press.

20 Employment security through dismissal protection: market versus policy failures

Christoph F. Buechtemann and Ulrich Walwei[1]

Introduction

Employment security and the impacts of legal restrictions on employment terminations by firms have been a regularly recurring theme and controversial issue in the labour market policy debates of most Western industrialized countries. Whereas the proponents of employment security regulations have upheld the view that restraints on employers' dismissal behaviour are necessary for establishing parity and fairness between the labour market parties and for stabilizing employment over the business cycle, the critics have blamed employment security regulations for slowing down necessary workforce adjustments, increasing fixed labour costs, reducing the allocative efficiency of labour markets and thus, at least in part, accounting for sluggish employment growth and persisting high levels of long-term unemployment in Europe as compared to the United States.

In the 1980s, such criticism, though frequently rooted in abstract notions rather than firm empirical evidence about the functioning of labour markets, spurred several European governments to introduce new legislation selectively relaxing legal dismissal and lay-off restraints and/or widening legal 'loopholes' allowing a circumvention of statutory dismissal protection, for instance, through facilitating the use of temporary workers or encouraging early retirement of older workers. Despite these changes, which have left the basic systems of statutory dismissal protection largely intact, the debate about the allegedly adverse employment and labour market impacts of employment security regulation has continued with undiminished intensity, receiving additional fuel from the substantial labour shedding during the latest recession of the early 1990s. This is witnessed by the fact that the OECD's *Employment Outlook* for 1993 (OECD, 1993, pp. 95ff.) and the 1993 edition of the EU Commission's *Employment in Europe* report (see Commission of the European Communities, 1993, pp. 173ff.) each devote a whole chapter to the issue.

The debate on employment security has stimulated a great deal of research effort during the past two decades. However, most attempts to determine the actual labour market impacts of existing employment security regimes have produced highly conflicting results. According to the present findings, strict

employment protection regulation has hardly had any impact on the global level of unemployment. However, the results provide certain evidence that the composition of unemployment can change if regulation is stricter. Unemployment among prime-age men would be lower, though at the expense of younger workers. The level of employment is not really affected, either, according to the result. What is more significant here too, however, are the structural effects, since strict employment protection obviously improves the employment prospects of prime-age men and worsens those of younger workers and of prime-age women. With respect to the flow values there are clearer links. Thus a high level of employment security reduces the fluctuation on the labour market (and raises the stability of employment), lowers the risk of becoming unemployed but simultaneously increases the duration of unemployment when it does occur (see, for example, OECD 1999). In particular, most previous research has not been well grounded in theoretical terms, lacking a clear-cut notion of the specific nature of modern employment relationships and the potential source of market failure inherent therein, as well as about the differential impacts and implications of contrasting regulatory approaches and legal approaches.

Defining the issue

The concept of 'employment security' has several dimensions. In the American debate, commentators commonly speak of *job security*, which – in a strict sense – refers to a worker's probability of keeping a particular job with a particular employer, including the specific work tasks, skill requirements, compensation and non-wage attributes (such as seniority entitlements) of that particular job. Job security, in this sense, presumes narrowly defined job categories or job descriptions as they have underlain the traditional system of work rules in the US unionized sector.

Employment security, by contrast, is a broader concept and has both a macroeconomic and a microeconomic dimension. Macroeconomic employment security relates to the availability of employment opportunities in the economy as a whole and could be defined as a worker's probability of staying in employment or finding new employment as long as he or she wishes to participate in the labour force. The degree of macroeconomic employment security is primarily dependent on the overall state of the economy and is influenced by macroeconomic policies and only very indirectly by labour legislation. Microeconomic employment security refers to the worker's probability of not being unjustly or arbitrarily dismissed by the employer and thereby relates to a continuing employment relationship with a particular firm (rather than to a particular job within that firm). Unlike 'job security', microeconomic employment security may involve (and in fact often depends upon) the possibility of internal job reassignments as well as changes in work tasks, working conditions and compensation.

It is evident that these different dimensions of employment security are closely interrelated and highly interdependent: in a stable economic environment of economic growth and tight labour markets (as characterized the period of the 1960s and early 1970s, when most employment security regulations in European countries were enacted), the provision of microeconomic employment security involves low costs for firms and few benefits for workers since, in the case of a job termination, alternative employment can easily be found. Macroeconomic policies, by smoothing demand fluctuations, may thus perform a vital role in facilitating the provision of microeconomic employment security by firms. The same is true with regard to income security policies which, for instance, by subsidizing firms' labour hoarding during cyclical downturns or temporary restructuring in the form of short-time working compensation or unemployment benefits, may likewise strongly reduce the costs borne by firms for providing employment security. Finally, microeconomic employment security is inversely related to job security inasmuch as it directly depends on the fiexibility or variability of other elements of the employment relationship: not only workers' job assignments and skills brought to the job, but also working time and wages.

A further important distinction in this context is the one between 'institutional' and '*de facto*' employment security. The latter term is commonly used to describe empirical findings that, even in the absence of legal or collectively negotiated dismissal and lay-off restraints, many workers enjoy a high degree of factual employment security involving long-term employment relationships extending through variable states of the economy. *De facto* employment security thus refers to the fact that both firms and workers may have purely economic motives for continuing employment relationships over longer periods (without any externally imposed constraint). From mere *de facto* employment security (which may be better termed 'employment *stability*', since it does not preclude individual workers – at any point in time – from facing insecurity as to the further continuation of their employment relationship), we distinguish institutional employment security through explicit or implicit (though enforceable) rules and provisions imposing restraints on firms' ability to dismiss workers 'at will'. Such restraints may be 'endogenous', that is, the outcome of unconstrained 'efficient' contracting between firms and workers (or their representatives) or 'exogenous', that is, imposed by law or collective agreements, as in most European countries and large parts of the US public sector. Whereas endogenously evolved employment security rules and standards (for example, in the form of individual contract terms or company agreements to follow a 'no lay-off' policy) are generally considered less problematic in economic terms,[2] the controversy over employment security has focused largely on institutional dismissal or lay-off restraints imposed by legislation and/or industry-wide collective agreements.

Finally, it should be noted that in most cases institutional employment security does not mean absolute security against job loss in the sense of guaranteed employment.[3] In modern market economies, employment security, for the vast majority of workers, can only mean security from unjustified or arbitrary employment terminations.[4] That is, employment terminations are tied to certain standards, rules and procedures, the concretization and 'strictness' of which tend to vary across different groups of workers, from situation to situation, as well as from country to country. Consequently, employment security regimes must be viewed as a continuum rather than as a fixed point, ranging from legal pre-notification requirements, seniority rules for lay-offs and basic 'fairness' criteria for disciplinary dismissals all the way to comprehensive systems involving judicially enforced 'just cause' requirements, mandatory consultations with worker representatives and public authorization procedures. From an analytical perspective, the need for policy interventions in this area essentially hinges on two questions: first, whether, from the viewpoint of maximizing individual and collective welfare, the nature of modern labour markets and modern employment relationships requires special arrangements regarding the conditions and terms of employment terminations; and second, whether the definition and enforcement of such terms and conditions can be achieved more efficiently[5] through public policy (legislation) than through mere private contracting by the individual labour market parties.

Modern labour markets and the nature of the employment relationship

In fact, developments in theoretical labour market analysis since the mid-1970s have provided several arguments that can be used to answer both the above questions and, at the same time, yield criteria for formulating hypotheses about the efficiency of different employment security regimes. One major strand of theory production has been inspired by the empirical finding of prevailing long-term job attachments and delayed or incomplete employment adjustment to demand fluctuations and has proposed several arguments why both firms and workers may, indeed, share an ('endogenous') economic interest in open-ended, stable employment relationships extending over longer time periods. Central to most of these arguments has been the notion of market imperfections due to asset specificity and idiosyncratic exchange: that is, exchanges between workers and firms (as opposed to transactions in classical commodity markets) involve specific irreversible investments in relationship-specific capital (sunk costs) that generate economic rents for both sides (dual monopoly) as long as the exchange between the two parties continues, thereby generating a mutual interest in long-term relationships.[6]

More specifically, most theoretical models in this tradition have focused on fixed employment costs (turnover costs) which have to be incurred at the

beginning of the employment relationship (costs for screening job candidates, hiring, specific training and matching workers to jobs). Other models have emphasized workers' higher risk aversion as compared to firms as a rationale for *implicit contracts* distributing income streams over longer periods. And even other theories, known as *efficiency wage* or *effort regulation* models, have focused on information asymmetries and on the costs involved in monitoring work effort in employment relationships and have provided arguments that seniority-based pay and deferred compensation schemes may be efficiency enhancing. *Internal labour market* theories, finally, have combined several of the above explanations in order to account for the observed dualism of workforce adjustment inertia in some parts of the economy (primarily larger firms) and more rapid adjustments involving external labour turnover in others (for example, occupational labour markets, as prevailing in the traditional crafts sector, or secondary labour markets employing primarily casual labour). In this view,

> [Internal labour markets] develop as technological considerations require firm-specific investments and asymmetric information necessitates the monitoring of work effort: workers make sunk investments in training and monitoring by accepting deferred compensation. At the same time the firm makes sunk investments in shared training costs and in intangibles such as reputation in the labour market. (Cohen and Wachter, 1989, p. 245)

Taken together, these theoretical approaches provide strong arguments to explain why, even from a strictly microeconomic perspective, stable, long-term employment relationships are by no means incompatible with, but rather in many instances conducive to (if not even a necessary prerequisite for), economic efficiency and societal welfare maximization. Moreover, theory also shows that long-term employment relationships between workers and firms necessarily involve a high degree of *uncertainty* for both sides. First, the existence of information asymmetries implies the risk of *opportunistic behaviour* by either party: thus firms, in order to reduce the risk of 'shirking' by workers, will be inclined to design special compensation schemes rewarding work effort and, at the same time, will want to retain the right to dismiss workers whose work effort fails to match expected standards. Workers, on the other hand, in the absence of any binding rules or fairness standards, cannot, for instance, rely on their employers not dismissing them before they reach seniority and get compensated for past work effort, thus reducing workers' willingness to accept efficiency-enhancing deferred compensation schemes in the first place. Second, the parties share a lack of information and face uncertainty with regard to the future into which their relationship must continue if the returns to their initial shared investments are to be reaped: for example, the firm may be forced to reduce its workforce and to relinquish its

promise of future compensation for past work effort in the event of an unforeseen major decline in the demand for its products. In the presence of information asymmetries and future uncertainty, the willingness of both parties to engage in long-term employment relationships involving sunk investments in relationship-specific capital, therefore, depends on whether they can devise and implement between them a governance structure that addresses the dual question of mutual trust and risk sharing.

Mutual trust requires safeguards against opportunistic behaviour by the other side. This involves firms not being able to fire workers (or unilaterally alter contract terms) 'at will' or arbitrarily, thereby depriving workers of their share in the rents from mutual sunk investments, and workers not being able to lower the firm's returns to such investments by withholding work effort ('shirking') without facing dismissal.

Risk sharing, on the other hand, requires enforceable rules that the parties, each within their relative capacity, equally share the potential losses resulting from unpredictable future events. This implies that firms cannot be forced to maintain the employment relationship if, owing to altered economic circumstances, this would involve a serious jeopardizing of the medium-term survival of the firm, but that they would have to compensate workers for unrewarded effort in the past.

Sources of market failures and rationales for public policy interventions

From the above one could conclude that 'workers' rights in jobs do exist as a result of voluntary exchanges' (Addison, 1989, p. 136) and that, consequently, there is no need for legislation. This has, indeed, been the view taken by the proponents of a radical *laissez-faire* in labour relations who have postulated that, given clearly defined and assigned property rights, unconstrained voluntary negotiations between private parties produce 'optimal' results and, therefore, have opposed any further interventions in the labour market (see, for instance, Epstein, 1984; Posner, 1984, pp. 990ff.). Such arguments rest on the assumption that freely negotiated, private contracts are also able to provide efficient solutions to the dual problem of opportunistic behaviour and future contingencies inherent in labour market transactions. Conversely, any plea for (or defence of) employment security legislation must provide evidence that mere private contracting in modern labour markets produces inefficiencies that can be avoided or minimized by third-party interventions.

Indeed, a large and growing body of research on labour law regulation (see, for example, Leslie, 1989; Schwab, 1989) and regulation in other fields (see Shavell, 1984; Cooter and Ulen, 1988; Spulber, 1989) has pointed out various situations in which market failures are likely to occur and in which third-party regulation (legislation) may produce superior results. Such a situation arises whenever the negotiation and monitoring of private contracts

would involve very high transaction costs that can be reduced by the establishment of general standards and rules by an external agency. With increasing sunk investments, increasing time horizons for reaping the rents thereof and an overall increasing volatility in the economic environment, the transaction costs of privately negotiating individual contracts between firms and workers can be expected to rise. Another situation for efficiency gains through legislation arises when private parties (in our case, firms and workers) are likely to agree on terms that produce external costs for third parties not involved in the original bargain ('externalities'), pointing to the fact that real markets are frequently not fully competitive. For the labour market, 'insider–outsider' theories provide many illustrations of how 'efficient' bargains between two parties ('insiders' and firms) produce costs for non-participating outsiders. In these cases, legal regulation forcing the parties to fully internalize the costs of their behaviour may be efficiency enhancing.

Binding legal standards may further increase overall efficiency if the *enforcement* of private contracts is very costly and legislation could provide enforcement at a lower cost. This is the case whenever contractual compliance is difficult to monitor, contract terms are vague and may give rise to conflicting interpretations, damages resulting from non-compliance are difficult to measure and causation of damages or harms is difficult to establish. In such instances, conflicts over contract terms are more likely to arise and to entail costly information gathering by external arbitrators in the course of lengthy case-by-case fact-finding investigations in which the (frequently implicit) actual terms of the contract and the behaviours of both parties during the employment relationship have to be retrospectively established, thus producing a high degree of (legal) uncertainty for both sides (see Craswell and Calfee, 1986; Kolstad, 1990). In fact, the aforementioned describes exactly those problems which have prompted several critics of the US common law system to recommend the introduction of European-type 'unjust dismissal' legislation in the United States (see Krueger, 1991; Gould, 1993, pp. 63ff.; Maltby, 1994): given the growing complexity and contingencies involved in 'modern' employment relationships,[7] the latter progressively cease to 'fit comfortably into the traditional common law scheme' (Leonard, 1988, p. 636).

Finally, Levine (1991) has pointed out that private contracting may fail altogether to rule out the dual problem of *adverse selection* and *moral hazard* inherent in open-ended employment relationships involving sunk investments. As long as (for example) the negotiation of 'just cause' requirements for employment terminations is left exclusively to the discretion of private parties, firms offering 'just cause' are prone to attract workers who in other firms, following an 'at-will' policy would face a higher risk of being dismissed; for the individual firm, such adverse selection of job candidates and the ensuing higher risk of shirking once the workers are hired may have an

incentive to discontinue offering 'just cause' standards for dismissals, thereby, of course, also affecting the lot of non-shirking workers. In such cases, legislation, by universalizing rules and standards for individual behaviour, may solve typical 'prisoner's dilemma' situations and thereby contribute to increasing overall efficiency (see also Buttler and Walwei, 1994).

Summing up, there appear to be many instances in which third-party interventions, and dismissal legislation in particular, by mitigating the risk of market failure, may indeed foster efficient contracting between firms and workers and thus enhance, rather than impede, overall labour market efficiency. In the present context, this implies that the direct economic costs imposed by dismissal legislation, which have been the focus of the current debate on employment security policies, need to be carefully weighed against the alternative costs of merely private contracting and – if the latter should fail – the socioeconomic welfare costs of forgone sunk investments.

Risks and sources of policy failure

The existence of market failures does not imply that dismissal protection regimes as we encounter them today are necessarily 'efficient' in the sense that the overall benefits emanating from them exceed their economic costs. Rather, as we have been able to identify several sources of *market failure* which, in the absence of regulatory interventions, lead to suboptimal results, we can also identify several sources of *policy failure* causing major inefficiencies and market distortions. Specifically, the following types of policy failure appear important in the present context.

Design failures

Policy design failures occur when policy programmes interfere with the basic logic underlying efficient private contracting in labour markets, with the result of discouraging instead of encouraging mutual sunk investments in relationship-specific capital, or of preventing labour market transactions altogether. In the area of dismissal protection legislation, such policy design failures arise, for instance, when dismissal protection regimes are overburdened by distributional or social policy objectives that are not aimed at merely preventing externalities from private contracting. An example would be absolute employment protection regulations for certain groups of workers (such as severely handicapped workers or older workers): in the absence of supporting measures (for example, job creation schemes for the hard-to-place workers or training subsidies), such regulations force firms to bear fully the costs for worker protection and thereby violate the risk-sharing principle underlying efficient employment contracts. This is particularly so when special protection for such groups is not seniority-graded, that is, tied to relationship-specific investments that create a genuine economic interest in or

rationale for stable employment relations, and does not include mechanisms (such as probationary periods) ruling out or at least reducing firms' risk of adverse selection. Absolute dismissal protection regulations of this kind, therefore, tend to induce unintended behaviour, such as discrimination against protected groups in firms' hiring policies, thus creating typical insider–outsider problems. Another instance of policy design failures is to be seen in vague and unclear standards and rules that fail to establish clearly defined property rights and/or to specify modes of recourse and penalties in the case of non-compliance and which therefore are prone to giving rise to conflicting interpretations and costly litigation and creating a high degree of legal uncertainty.[8] Finally, inefficiencies can also result from legal thresholds, such as employment-based firm-size thresholds that determine whether the firm is subject to statutory dismissal protection regulations or not. These thresholds produce inefficiencies (and frequently induce evasive behaviours by firms) in that hiring the one worker who raises the number of workers above the threshold causes the employment terms of all employed workers in the firm to change.

Implementation failures

Another potential source of policy failures is the way in which public policy programmes, laws and regulations are implemented. In the case of dismissal protection legislation, implementation rests primarily with the labour courts (or corresponding publicly mandated arbitration institutions or tribunals) and with special public authorization agencies (for example, local labour offices for major collective lay-offs or special boards responsible for deciding upon the justification of dismissals of special groups, such as disabled workers). Here, too, an important cause of inefficiencies may arise from judicial decisions (or precedents set) by labour courts (or equivalent bodies) which, in concretizing uncertain legal norms, basically change the substance of laws (as compared to preceding judicial interpretations), thereby retroactively affecting existing property right definitions and assignments and – as far as such court decisions establish general precedents – altering the terms of all permanent employment relationships within the specific jurisdiction (see Hamermesh, 1993). As in the case of legislative changes, the efficiency losses will be all the larger, the less judicial decisions are based on empirically grounded notions of 'reasonable' employer behaviour and the more they are explicitly or implicitly intended to shift the existing power relationship between workers and employers. In fact, the now vast body of (partly inconsistent) labour court decisions redefining legal standards and the ensuing high legal information costs and high degree of legal uncertainty have been at the very centre of recent criticism of legal dismissal protection by German employers. Another type of implementation failure results from long

authorization or arbitration procedures, which, apart from causing costly delays for the labour market parties, frequently preclude certain remedies (such as reinstatement of unjustly dismissed workers) because situations have changed when a decision or settlement is finally reached. In fact, ensuring speedy and 'professionally' conducted arbitration procedures in unjust dismissal litigation has frequently been proposed as a rationale for the establishment of a system of specialized labour courts or arbitration agencies for handling dismissal disputes, such as exist in several European countries.

Policy failures due to structural 'misfit' and institutional malcoordination

One of the most frequent sources of policy failures is malcoordination between institutional incentives and legally required behaviours, on the one hand, and the changing opportunity structures as well as shifting behavioural dispositions and repertoires of economic agents (workers and firms), on the other. That is, regulatory interventions that may have been efficiency enhancing in a particular historical situation and under specific economic conditions cease to be so when economic circumstances and prevailing forms of economic organization change. Moreover, the behavioural incentives and signals emanating from public policy interventions may be affected by changes in the wider regulatory and institutional environment in which they operate. For instance, enhanced market volatility and accelerated (exogenous) technological change may raise the costs of employment security to employers, thus reducing the latter's interest in continuing long-term employment relationships. In such an environment, in which adjustment speed ('time competition') becomes a core prerequisite for sustaining competitiveness, the institutionally stabilized logic of internal labour markets underlying most existing dismissal protection regimes may indeed come into conflict with the requirement of a faster reallocation of labour and declining returns to investments in relationship-specific capital (for example, specific skills acquired by workers).

The increasing 'misfit' between dismissal protection regimes and economic requirements and the ensuing inefficiencies are then seen as the result of a failure of public policy to adjust to changing economic conditions ('institutional inertia').[9] A different situation is seen when institutional or regulatory changes in other areas (such as rising legal minimum wages) discourage mutual sunk investments in relationship-specific capital or facilitate (and thereby encourage) the circumvention of dismissal protection regulations (for instance through fixed-term contracts). Given uncertainty and bounded rationality, employers may, for example, prefer hiring workers on fixed-term contracts, thereby, of course, not only reducing the (potential) cost of employment terminations, but also forgoing the rents to be reaped from

mutual sunk investments in the context of longer-term, open-ended employment relationships.

The above discussion of market and policy failures implies that assessing the impacts of dismissal protection legislation on labour market efficiency and overall societal welfare requires not only a detailed analysis of existing dismissal protection regimes and the characteristics of the various labour market settings in which they operate, but also a careful balancing of their costs and benefits. From a theoretical point of view, the latter must also include the development of well-grounded notions of the *opportunity costs* of existing dismissal protection regimes, that is, the likely costs and benefits of alternative approaches to managing the complexities of modern labour markets and modern employment relationships. It is clear that these are ideal standards and requirements which, as the result of operationalization difficulties and data limitations, most actual evaluation research cannot possibly fulfil. However, they define a set of theoretically grounded criteria on the basis of which the designs, methodologies and results of previous evaluation studies can be judged with regard to their relative capacity to yield reliable policy recommendations.

Elements of 'efficient' dismissal protection regimes

The conclusion that, under certain conditions, legislation may be superior in efficiency terms to mere private contracting does not yet answer the crucial question of *which type of legislation* is best suited to achieve these goals. There certainly is no abstract, universal answer to this question. Even if a model for an 'optimal' dismissal protection regime could be designed, it would still have to face the crucial test of merging into the diverse socioeconomic and wider institutional settings to be found in the real world and the specific behavioural rationalities or 'opportunity structures' prevailing therein. Nevertheless, the above theoretical discussion can provide some general clues or indications as to the basic attributes and elements of 'efficient' employment security policies and, thus, yield a set of analytical criteria by which the likely economic impacts of different existing dismissal protection regimes can be assessed.

First, in order to provide a supporting structure for investments in relationship-specific capital and the private returns resulting from them, efficient dismissal protection legislation will have to be fashioned close to the sunk cost rules underlying private contracting in modern labour market settings. More specifically, this has a number of implications. In order to mitigate the risk of opportunistic behaviour, efficient dismissal protection regimes will contain clauses restraining employers' freedom to dismiss workers 'at will' and tying individual employment terminations to the requirement of some objective and verifiable 'just cause', such as malperformance, shirking or

misconduct on the part of the worker. Moreover, the risk of adverse selection of malperforming or shirking workers would seem to require the existence of a waiting or probationary period for newly hired workers, during which legal dismissal restrictions do not apply, thus allowing firms a careful on-the-job screening of novices without facing major termination restraints. On the other hand, given that investments in relationship-specific capital tend to increase with tenure, the strictness of the criteria and documentation requirements applied to individual employment terminations will increase with the worker's seniority.

Risk sharing in the event of an economically induced decline in labour demand means that the firm retains the right unilaterally to terminate workers' employment if a change in economic conditions requires it to do so, provided that, prior to resorting to lay-offs, the firm has ostensibly taken all 'reasonable'[10] efforts to prevent lay-offs through alternative forms of workforce and hours reduction, such as a temporary employment ban and/or work sharing. When lay-offs become inevitable, the existence of deferred compensation schemes to elicit work effort and worker commitment (which can be assumed to be given in one form or another in all large-firm settings with internal labour market structures), would seem to require efficient dismissal protection regimes to contain clauses demanding that junior workers will be laid off first (lay-off by reverse seniority) and that lay-offs of tenured senior workers involve severance payments as (partial) compensation for past work effort.

Second, in order to prevent externalities arising from private maximizing behaviour, 'efficient' dismissal protection regimes will include regulations forcing the private labour market parties to *internalize* the full costs of their rent-generating transactions. Generally, such externalities are given when the external costs caused by dismissals and lay-offs exceed the costs that the labour market parties would have incurred to prevent dismissals or lay-offs. In order to reduce external costs caused by lay-offs, 'efficient' dismissal protection regimes will include *pre-notification requirements* for non-disciplinary dismissals, enabling workers to search for new employment while still on the job, thus reducing the burden to be carried by public unemployment insurance schemes caused by frictional unemployment. Since the likelihood of finding new employment can be expected to differ across workers according to the specificity of their acquired human capital, pre-notification periods will, furthermore, be seniority-graded: increasing with age and/or tenure in the current job.[11] Moreover, major externalities tend to result from large-scale lay-offs and mass dismissals which may produce a shock for local labour markets and impose various costs on local communities at large (for example, in the form of declining local product demand or declining real estate prices). To mimimize such externalities, efficient dismissal protection regimes will include clauses requiring advance notification of intended major

lay-offs to local labour market authorities, with the latter – in some instances – being entitled to demand a staggering of lay-offs to avert a local labour market shock.

Another type of externality may result from strategies by firms and the majority of incumbent workers ('insiders') to systematically 'weed out' individual workers who are less productive, thereby raising their own stream of revenue and contributing to an increasing segmentation of the labour market between highly productive 'insiders' and less productive 'outsiders'. In so far as such behaviour causes external costs exceeding the alternative (internal) costs of retaining such workers in the firm (as, for instance, in the case of pregnant women or older workers approaching retirement), legally imposed protective provisions raising the internal costs of dismissing such workers (for example in the form of special, tradable employment guarantees) may enhance overall efficiency.

Third and last, taking into account that the sunk cost rules characterizing modern employment relationships (as well as the adverse external effects of lay-offs) apply primarily to medium-sized and larger firms with internal labour markets, efficient dismissal protection regimes will further include selective exemptions for small firms which lack the scope and organizational structures to have recourse to alternative modes of workforce adjustment (such as attrition) and thus will face significantly higher costs due to dismissal protection than larger firms.

It is no coincidence that the above theoretically derived principles of 'efficient' dismissal protection regimes in modern labour market settings largely correspond to the basic structure and components of actual dismissal protection regulations found in many European countries (as well as to explicit or implicit practices *voluntarily* adopted by many large firms in countries without dismissal legislation, such as Japan and the United States): historically, the introduction of dismissal protection legislation has been, as a common feature across many European countries, closely associated with the spread of internal labour markets and their typical concomitants of capital-intensive production, economies of scale, a high degree of division of labour, high costs of labour turnover and continuing firm–worker attachments. In fact, in those countries for which historical evidence is available, many of the rules and procedures required by dismissal legislation had already become common practice in large parts of the economy *before* such legislation was first introduced or extended in the 1960s and 1970s. In these cases, legislation appears to have taken the form of a mere codification of widespread *de facto* practices that had evolved endogenously in dominant and advanced sectors of the economy (see Buechtemann, 1993).

Such a policy mode of legally codifying prevailing 'best practices' involves several advantages. First, by merely universalizing what the majority

of firms practise in the absence of exogenously imposed legal constraints, it produces few additional costs as its immediate impact is confined to a minority of (idiosyncratic) firms whose personnel policies deviate from the predominant pattern, encounters little political opposition and induces little circumvention. Second, through legally codifying the (implicit or explicit) terms of the majority of permanent employment relationships, it largely avoids the efficiency losses incurred by a redefinition and reassignment of property rights that interfere with permanent market exchanges. Third, by immediately drawing on 'efficient' practices in advanced sectors of the economy, it avoids the problems and risks of regulatory failure due to legislators' information deficits about the functioning of real labour markets and overcomes the crucial problem of the 'structural coupling' (Teubner) between public policy, on the one hand, and the behavioural dispositions and repertoires of economic agents, on the other.

Notes

1. Large parts of this entry are based on 'Employment and dismissal protection', by J. O'Reilly and K. Schömann in G. Schmid, J. O'Reilly and K. Schömann (eds) (1996), *International Handbook of Labour Market Policy and Evaluation*, Aldershot, UK and Brookfield, US: Edward Elgar.
2. *Insider–outsider theories* have shown, however, that – given market imperfections – such unconstrained agreements between firms and workers are not necessarily optimal in terms of overall economic efficiency, although they may be 'efficient' from the viewpoint of firms and employed 'insiders'.
3. An exception is civil servants with lifetime appointments.
4. An exception to this rule is *special* dismissal protection regulations that apply to certain categories of workers, such as members of firm-level worker representation bodies (for example, the German works councils), drafted persons on military service, women during pregnancy and on maternity leave, as well as severely disabled persons. However, in most of these cases, there is also the possibility of offering protected employees severance pay in order to induce them to quit. Such special protective regulations, which are not dealt with in this chapter (but see OECD, 1994, pp. 173ff.), thus, in effect merely raise the costs of employment terminations and do not provide any absolute protection against job loss, as is illustrated by the case of plant closures and firm bankruptcies.
5. 'Efficiency' here simply denotes maximizing the output value derived from a given set and quantity of inputs.
6. It should be noted that most of these models assume transactions between formally equal market participants and a clear definition and assignment of initial property rights. Such property rights 'establish the initial distribution of rights, the exclusivity of the rights, and the mechanisms under which transfers of property rights are effected and recognized' (Spulber, 1989, p. 48). Without the assignment of clearly defined, exclusive property rights, market transactions cannot take place.
7. In a unique historical analysis, Carter (1988) has shown for the United States (California) that the duration of employment relationships and the importance of near lifetime jobs have significantly increased between the late nineteenth and the late twentieth centuries.
8. An empirical indicator of such inefficiencies due to unclear standards would be the number of dismissals that are legally contested through litigation (see, for example, Barnard et al., 1995, pp. 36f.).
9. An empirical indicator of growing inefficiencies of this kind could be the overall amount paid by employers in order to buy out protected employees (severance pay).

10. 'Reasonable' in this context would refer to all those measures that do not jeopardize the economic viability of the undertaking at large. In this sense, Cooter and Rubinfeld (1989, p. 1068) observe that 'the "reasonable man" of the law is not very different from the "rational man" of economics'. Most legal dismissal protection regimes are based on the notion of a 'reasonable' employer.
11. Such 'externalities' resulting from private maximizing behaviour are frequently induced by the existence of external institutions, such as the public provision of unemployment or early retirement benefits without (any) (as in most European countries) or with incomplete (as in the United States) experience rating of employers' unemployment insurance contributions: in the absence of special legal safeguards, for instance, in the form of checks on the causes of job losses and the economic necessity and inevitability of dismissals, such institutions create incentives for the labour market parties to deliberately negotiate terms between them that involve costs for non-participating third parties, such as public unemployment insurance or retirement schemes.

References

Addison, J.T. (1989), 'Job rights and economic dislocation', in R. Drago and R. Perlman (eds), *Microeconomic Issues in Labor Economics: New Approaches*, New York: Harvester-Wheatsheaf.

Barnard, C., J. Clark and R. Lewis (1995), *The Exercise of Individual Employment Rights in the Member States of the European Community*, London: Department of Employment.

Buechtemann, C.F. (ed.) (1993), *Employment Security and Labor Market Behavior: Interdisciplinary Approaches and International Evidence*, Ithaca, NY: Cornell University Press.

Buttler, F. and U. Walwei (1994), 'The employment contract: cooperation requires stability', *IAB Labour Market Research Topics*, no. 7.

Carter, S.B. (1988), 'The changing importance of life-time jobs, 1892–1978', *Industrial Relations*, **27** (3), 287–300.

Cohen, G.M. and M.L. Wachter (1989), 'An internal labor market approach to labor law: does labor law promote efficient contracting?', in B.D. Dennis (ed.), *Proceedings of the 41st Annual Meeting of the IRRA*, Madison, WI: Industrial Relations Research Association.

Commission of the European Communities (1993), *Employment in Europe 1993*, DG V, Brussels.

Cooter, R.D. and D.L. Rubinfeld (1989), 'Economic analysis of legal disputes and their resolution', *Journal of Economic Literature*, **27**, 1067–97.

Cooter, R.D. and T. Ulen (1988), *Law and Economics*, Glenview, IL: Scott, Foresman & Co.

Craswell, R. and J.E. Calfee (1986), 'Deterrence and uncertain legal standards', *Journal of Law, Economics and Organization*, **2**, 279–303.

Epstein, R. (1984), 'In defense of the contract at will', *University of Chicago Law Review*, **51**, 947–82.

Gould, W.B. (1993), *Agenda for Reform: The Future of Employment Relationships and the Law*, Cambridge, MA: MIT Press.

Hamermesh, D.S. (1993), 'Employment protection: theoretical implications and some U.S. evidence', in Buechtemann (ed.).

Kolstad, C.D. (1990), '*Ex-post* liability for harm versus *ex-ante* safety regulation: substitutes or complements?', *American Economic Review*, **80** (4), 888–901.

Krueger, A.B. (1991), 'The evolution of unjust dismissal legislation in the United States', *Industrial and Labor Relations Review*, **44** (4), 644–60.

Lazear, E.P. (1990), 'Job security provisions and employment', *Quarterly Journal of Economics*, **105**, 699–726.

Leonard, A.S. (1988), 'A new common law of employment termination', *North Carolina Law Review*, **66**, 631–86.

Leslie, D.L. (1989), 'Economic analyses of labor law', in B.D. Dennis (ed.), *Proceedings of the 41st Annual Meeting of the IRRA*, Madison, WI: Industrial Relations Research Association.

Levine, D.J. (1991), 'Just-cause employment policies in the presence of workers' adverse selection', *Journal of Labor Economics*, **91**, 294–305.

Maltby, L.L. (1994), 'The projected economic impact of the Model Employment Termination Act', *Annals of the American Academy for Political and Social Science*, **536**, 103–18.
Organization for Economic Cooperation and Development (OECD) (1993), *Employment Outlook*, Paris: OECD.
Organization for Economic Cooperation and Development (OECD) (1994), *The OECD Jobs Study: Evidence and Explanations, Part 2: The Adjustment Potential of the Labour Market*, Paris: OECD.
Posner, R.A. (1984), 'Some economics of labor law', *University of Chicago Law Review*, **51**, 988–1011.
Schwab, S. (1989), 'The economics invasion of labor law scholarship', in B.D. Dennis (ed.), *Proceedings of the 41st Annual Meeting of the IRRA*, Madison, WI: Industrial Relations Research Association.
Shavell, S. (1984), 'Liability for harm versus regulation for safety', *Journal of Legal Studies*, **13**, 357–74.
Spulber, D.F. (1989), *Regulation and Markets*, Cambridge, MA: MIT Press.

PART V

REGULATION, TAXATION AND PUBLIC ENTERPRISE

21 Structures of public enterprise

Jürgen G. Backhaus

One of the most fascinating developments in law and economics analysis is the possibility of subjecting public institutions to a scrutiny that allows an assessment of their performance. This entry first gives an introduction to the basic approach and then details a law and economics-based performance analytic approach of public institutions such as public enterprises.

Introduction

As a consequence of the collapse of state socialism in Central and Eastern Europe, many countries found themselves with a large set of publicly owned entities which often imposed a considerable burden on the public budget. Hence governments designed different programmes to privatize these entities (which often could hardly be referred to as enterprises as they tended to display traits of bureaucracies) or at least to make them independent of public support. However, this task proved to be very difficult. For instance, the performance of the authority charged in Germany with the privatization of the state socialist industry inherited from the defunct German Democratic Republic gives a mixed picture of what could be accomplished under the most favourable of circumstances. As Table 21.1 indicates, out of a total of almost 14 000 enterprises in East Germany, a mere 354 remained to be privatized in the autumn of 1994. But all the assets had been sold for a trifling

Table 21.1 The Treuhand legacy

East German enterprises taken over	13 781
Number still to be privatized*	354
Enterprises in liquidation	3701
Private investment pledged	$133.2 billion
Jobs guaranteed	1.4 million
Revenues from sales of enterprises	$41 million
Management buyouts	2697
Enterprises acquired by non-German investors	855

Note: * As of September 1994.

Source: Treuhand, in Gumble (1994).

41 million dollars, many had been given away for a symbolic price only and the prime purpose of the privatization activity was clearly to save as many jobs as possible; indeed, a total of 1.4 million job guarantees had been attained.

However, these job guarantees often had to be paid for dearly. For instance, the Ecostar GmbH in Eisenhüttenstadt required a financial commitment of about 500 thousand dollars per job saved (Gumble, 1994). These figures show clearly that different purposes can be obtained by managing or privatizing public entities. In fact, next to generating employment and the obvious aim of reaping a profit, any number of policy objectives can be pursued by operating a public enterprise. These objectives may lie in competition, regional, foreign, defence, cultural or environmental policies or, indeed, any conceivable public purpose that may form part of a political agenda. It is for this reason that the assessment of the performance of such a public enterprise is by no means a straightforward task. The difficulty lies in identifying the objective against which the operation of the entity needs to be assessed. It is here that a strict application of a law and economics analysis can be of major importance.

Attribution of objectives

One of the most fascinating recent developments in law economics can be described as a conscious turning away from prescriptive propositions to research work intended to evaluate the actual performance of identifiable institutions in the light of contemporary public finance theory. One of these attempts is referred to by its supporters as the 'performance approach', where an attempt is made to investigate to what extent public enterprises actually achieve the goals attributed to them in the normative literature. The underlying idea is to determine the 'best practice' of a particular enterprise and assess other public enterprises in the same industry against that best practice. This entry is critical of the best practice approach and offers an alternative public choice-based theory. The best practice approach involves a performance analysis that unfolds in four stages. Henry Tulkens, one of the leading researchers supporting this approach, describes the stages as follows:

1. specifying explicitly the objective(s) attributed to the public firms;
2. providing a convincing justification for the selected objectives;
3. translating each one of these objectives in terms of indicators that are observable, and preferably measurable;
4. devising methods whereby the comparison can be established between the observed values of the said indicators, and those values that correspond to a complete fulfilment of the corresponding objective. (Tulkens, 1986, p. 430)

While the performance approach to evaluating public enterprise performance is undoubtedly an improvement on normative work, as Marchand et al. (1984) rightfully claim, the specific approach suggested may be characterized as a *conflict between technique and purpose*. If the purpose is to investigate to what extent normative public enterprise economics speaks to identifiable institutions in terms of detecting the normatively established priorities in the course of the actual performance of these institutions, that purpose of inquiry is an *epistemological* one and there is probably no point of disagreement between the present author and the researchers supporting the performance approach. If, on the other hand, the attributed criteria are used for the evaluation of the analysed enterprises instead of the theory from which they are derived, it is suggested that technique and purpose are at odds with each other.

To be sure, Tulkens for one is aware of the problem raised here. He writes: 'Indeed, a distinction ought probably to be made between the objectives attributed to public firms *by the economist*, when he develops his analysis, and those attributed to such firms *by society* when the latter establishes them' (ibid., p. 431, original emphasis). The possibility of one set of values held by the economics profession and another one held in society, although perhaps descriptive of the profession, seems to be at odds with received welfare economics.

The question then arises whether a mere *attribution* of policy objectives against which public enterprises are to be evaluated can be replaced by a theoretically based analysis leading to a determination of their (sets of) objectives. Again, the question is explicitly addressed by Tulkens himself:

> An implication for our purposes … is that objectives quite different from those above can conceivably be attributed to public firms, such as e.g. maximization of managers' compensation, or maximization of labor pay. Whether the term of performance still correctly applies in this perspective may be left to semantics; but a valid methodological question raised thereby is that of devising appropriate statistical methods for testing to what extent such 'positive' behavioral hypotheses can be supported by evidence. This needs to be done much along the same stages as those who outline the outset for the performance approach. (Ibid.)

Is the best practice really good enough?

Some of the problems arising in the course of performance analyses of the kind described above may be succinctly and somewhat provocatively summarized in the following five propositions, each of which will be followed by an explanation.

The objectives attributed to the operation of public enterprises (see also the 'family approach') do not reflect the specific purpose(s) *of public enterprises; they are rather general economic policy objectives*

In order to be able to analyse the performance of particular public enterprises such as railroads or post offices, one obviously has to define policy objectives of these public enterprises. In one particular case (Pestieau, 1987), this problem is solved by attributing the purported objectives of general public policy to the specific public enterprises. These objectives (in the case referred to here) are said to be efficiency, equity, financial balance and macroeconomic objectives. For the sake of this argument, it is not important which particular set of general policy objectives is attributed to the public enterprise. The principle of attribution, however, raises doubts. If all the different institutions responsible to a government are considered a *family*, Pestieau feels justified in suggesting that all members of the family have the same common objectives, and engage in the same activities (which is a *non sequitur*); that is, all government agencies are held essentially to subscribe to the same set of public policy objectives. This is a very curious assumption.

All the many institutions that comprise the state are, indeed, very different; they can achieve different purposes with different ease. One should therefore expect a *division of labour* among the different governmental institutions. Public enterprises are unique instruments of economic policy, since they allow governments to participate in markets and thus achieve specific results not attainable through non-market government behaviour. Whereas the typical instruments of economic policy may affect economic activity in general terms, public enterprises can be directed to specific purposes, specific geographic areas or even specific geographic points, specific demographic or ethnic or other social groups or otherwise specific political or social problems that, for one reason or another, are deemed worthy of political attention. Given this bewildering diversity, it is (almost) never possible to assign or attribute objectives to public enterprises in general, and certainly not a priori. For each particular case, the objectives of the public enterprise in question have to be empirically established by way of conducting a three-step procedure: (i) an analysis of the structure of the public enterprise, (ii) an analysis of its revealed performance, and (iii) an analysis of the interplay between the enterprise's own approach to account for its performance and the reaction of its principals to that account.

The structure (or form) of the institution needs to be analysed, because the constitution of the firm reveals the intentions and purposes of its founders. The performance assessments need to be analysed in terms of all the criteria the corporation itself and its principals deem relevant for describing it. The process of justifying this performance in terms of how the company reports on its performance and how the principals, openly or tacitly, react will reveal

and single out those criteria of performance assessment that can then be used by economists who want to undertake a performance study.

Efficiency is (almost) never in itself a policy objective of a public enterprise. It is unlikely (but not impossible) that we may find a public enterprise with essentially the same production function as an efficiently operated private enterprise

Efficiency is not an objective of private entrepreneurship. Why then should we expect efficiency to be a goal for public entrepreneurs? And what is 'efficiency' supposed to mean in such a context? Efficiency is not a goal, it is a condition for goal attainment. While private enterprises can generally be described as being managed in order to maximize the net present value of the firm, this is not meaningful for public enterprises. In democratic societies, ownership rights in public enterprises change hands frequently (and regularly) with the change in government, and owners are better described as *tenants*. Since public enterprises are fairly flexible tools, they tend to be used for all kinds of short-term objectives, of which at least some are not meant to be publicly discussed. For instance, a public railroad company, apart from its obvious mission of passenger and freight transportation, plays a role in regional and local planning and development, in technology policy, in European and international politics, in policies *vis-à-vis* specific sectors such as agriculture and mining, and so on. The list of possible objectives is inexhaustible. It is typical of a public enterprise that its basic mission, such as railroad or postal service, has to be fulfilled at an *acceptable* level, where acceptability is politically determined. The remaining leeway can then be filled in with political discretion. Therefore a public enterprise has to be called 'efficiently operated' if the remaining room for discretionary politics is efficiently used up. The definition of public enterprise efficiency can then be stated in the familiar Pareto form: *no further objective could be attained without compromising on any other (openly or tacitly) pursued objective*. The performance of such an enterprise, then, has to be judged in terms of the attainment of those objectives, subject to the satisficing constraint. As is readily apparent, the performance approach followed by Pestieau, Tulkens and others is only concerned with what has here been termed 'the satisficing constraint'.

The performance indicators suggested so far do not reflect the purpose(s) of running a public enterprise

In principle, any government has the choice of running a public enterprise as if it were a private one. This principle has exceptions. As has been pointed out before, the public owner of a public enterprise is the owner only for a definite period of time; he/she has the right of use, but not of complete ownership. However, the time horizon may well extend beyond the four or

five years we tend to associate with the terms of office available to democratically elected governments. Even at the federal level, a party may well expect to be in office for two terms. In local and state government, much longer terms are often the rule. When several public bodies share in the ownership, much longer time horizons may result from the need to agree on a common policy. Under these circumstances, governments have the option to use public enterprises as a revenue source. Then the difference between a public and a comparable private corporation could conceivably only consist in the length of the planning horizon. The longer the planning horizon of the democratic government, the smaller the difference. The smaller the difference, however, between the objectives of a public enterprise and the common objectives of a private one, the more likely we are to find little difficulty in estimating their production functions. Public enterprises conducted for profit as a revenue source are the best candidates for the performance approach suggested by Marchand et al. (1984).

Interestingly enough, we rarely observe governments pursuing this strategy, although there are important exceptions. This choice reveals a preference. The specific discretionary policy services rendered by public enterprises obviously are more important than the revenues which alternatively could be gained. The revenues forgone then yield an important hint for the assessment of the performance of a public enterprise. The policies pursued by spending the public enterprise rent (the rent equals the profits forgone) are off-budget and thus escape parliamentary scrutiny and public discussion. Reconstructing the policy function of a public enterprise by means of the revealed preference approach and, on the basis of this function, assessing the performance of the corporation is an important task for political economists. For estimating the public enterprise rent thus forgone, the public enterprise efficiently operated for a profit becomes an essential benchmark. Here lies another important application of the technique designed by Marchand et al.

Privatizations are not irrational

One of the most characteristic phenomena currently involving public enterprises is the almost universal wave of privatizations. Although privatizations assume many different forms, the phenomenon should clearly be seen as one particular form of conducting public business. Interestingly enough, the performance approach is silent when it comes to explaining privatizations. Pierre Pestieau (1987) creates the impression that the current wave of privatizations of public enterprises is somehow a sign of irrational politics. The property rights approach to the theory of the public firm, as sketched out above, does not agree with such a view (see Backhaus, 1989). As pointed out, the public owner of a public enterprise does not control the complete bundle of ownership rights; he/she is mainly restricted to the use of the firm. At the end of its

term, the present government has to turn over the corporation to its political opponent. There is only one way to prevent this from happening and to capture a substantial portion of the value of the firm at the same time: privatization. By selling a public enterprise, the present government can capture up to the entire stream of earnings from an efficiently managed firm. There are three aspects to this political move:

1. future revenues are transferred into the present; hence privatizations are equivalent to bond issues (see Wagner, 1977, for a full development of this argument);
2. the present owner is the last one who can use the public enterprise as a policy instrument; often the privatization process itself is extremely involved and loaded with political objectives;
3. the political power of the opponent is curtailed by blocking his/her access to the public enterprise in the future.

As seen in this perspective, there is nothing irrational about privatizing public enterprises in the context of a politicoeconomic view of public entrepreneurship.

Best practice performance assessments amount to measuring without a behaviourally based politicoeconomic theory

Unless there are identical political circumstances, the best practice of one (public) enterprise is not likely to be the best practice of another. Public enterprises are specific instruments of economic policy, and their objective functions can only be constructed after a careful analysis of their property rights structure and of the intentions of their political owners by means of the revealed preference approach. A performance analysis of a public enterprise is not possible without a theory explaining the political performance of these firms. Performance criteria, then, cannot be taken from economic discourse or from the observation of private firms operating in similar industries. Criticizing a public enterprise for not choosing the most advanced technique (best practice) amounts to ridiculing a hunter with bow and arrow for not using a rifle. The hunter will be amused, or worse, since he views with contempt those colleagues who gun down the deer at a distance instead of giving the animal a fair chance. Similarly, *the way public enterprises discharge their services is sometimes as important as the service itself*. Public enterprises serve multiple purposes, and the multiplicity of purposes certainly serves as a cover-up for a lot of slack. This is why performance analyses are important. They are meaningful, however, only if they are based on a realistic theory of public entrepreneurship. Such a theory is the precondition for deriving performance criteria, on which, in

turn, the performance analysis has to be based. The 'best practice' does not meet that standard.

Public enterprises produce public and private goods

About a century ago, public enterprises were still primarily launched and undertaken in order to contribute to public revenues in many European countries. Following a trend which started after the First World War, tax revenues have become increasingly more important than non-tax revenues. Today profits from public enterprises are a minor source of public revenues. From a strictly budgetary point of view, public enterprises today are often more important for the tax revenues they generate than for the non-tax revenues they provide in profits and sometimes also in fees. Today the charters of public enterprises often call for a whole range of different objectives other than the pursuit of profits. In Germany, there is even a school of experts in public enterprise economics who explicitly reject the idea of profit seeking for public enterprises and, instead, subscribe to the notion of '*Gemeinwohl*' as an appropriate description of their activities, thereby putting the whole public enterprise industry into the non-profit sector.

A point of theory

In trying to suggest an appropriate technique for analysing the performance of public enterprises thus described, we have to reckon with the many and different purposes governments try to pursue with them. As the preceding discussion may have suggested, the definition of the purpose of public enterprises is mainly negative. It is easier to characterize them in terms of the purposes they are explicitly *not* suggested as serving, instead of precisely establishing what the purposes are that they are supposed to fulfil. This may be a deplorable situation from the point of view of economic analysis, but we cannot reject the possibility out of hand that the apparent fuzziness of public enterprise purposes in most democratic societies is the very reason why they occupy such a prominent and increasing role in many economies, privatization programmes notwithstanding. The implication for attempts at evaluating the performance of public enterprises is that a technique of performance analysis has to be found that can serve as an *envelope* for the fuzzy purposes public enterprises are fulfilling. Instead of clearly establishing performance criteria drawn from applied welfare economics and *attributing* them to public enterprises, a technique may be preferable that starts from the public enterprise itself and attempts an increasingly more precise picture of the fuzzy set of objectives pursued. While such an evaluative procedure may be less straightforward than the Marchand et al. approach, the hope is that it will eventually be more enlightening in telling us what public enterprises really do, how they do it, and how well they do it.

The basic theoretical point of departure for this approach was introduced by Mancur Olson at a National Bureau of Economic Research (NBER) conference in 1973 devoted to the measurement of economic and social performance (Olson, 1973). Olson's point of departure is the suggestion that 'governments are in fact as well as by reputation usually inefficient, and that this is mainly because they deal with collective goods and externalities' (ibid., p. 359). It should be emphasized from the start that Olson explicitly refers to 'technical (in)efficiency'. His notion of efficiency has to be seen in the context of, for example, the literature on production functions, and not in the sense in which it is used by members of the Chicago school such as George Stigler. (See the interchange on X-inefficiency by Stigler, 1976, and Leibenstein, 1978.) This choice of efficiency concepts implies that public production, even where its methods have been shown to be inefficient for not reaching the production possibility frontier, may yet not be improvable in the Pareto sense because no better alternative is known that can readily be implemented. In terms of a transactions cost approach, this implies that public production, even where it is shown to be technically inefficient according to the performance analysis approaches, may still be efficient because unimprovable, since the transactions costs involved in this particular process of production are prohibitively high in rendering the internalization of externalities infeasible.

Which areas of public production is Olson referring to? His examples include law and order, defence, basic research and pollution control (Olson, 1973, p. 359), areas which form the new primary domain of public enterprises as opposed to the past smokestack industries. It's now some 40 years since Olson made this suggestion, public enterprises today are increasingly retreating from the production of merely private goods and thriving where they contribute to the production of positive externalities and public goods, a contribution often made as a byproduct of traditional production. Research laboratories with spin-offs for defence, passenger railroads to replace car transport and universities contributing to basic research are all cases in point and mark public entrepreneurship to which all Western states cling, irrespective of privatization programmes they may otherwise be engaged in.

The inherent inefficiency to which Olson pointed is unavoidable because it is inherent in the very definition of a public good. These inefficiencies are of two kinds. To the extent that no-one can be excluded from the consumption of the good, the true preferences for the provision of the public good are difficult to establish. Hence, in assessing the performance of public production, we never know whether too little or too much is being produced. The second source of inefficiency is more difficult to handle. Even where governments provide too little or too much of a particular public good, they may do so perfectly efficiently in terms of reaching the production possibility fron-

tier. The condition of jointness of consumption implies that the marginal calculus breaks down and that measuring of outputs becomes infeasible. As Olson puts it: 'The very characteristics of a collective good that make it a kind from which non-purchasers cannot be excluded, also make it a kind of which the output is not in the form of divisible units that can be readily counted' (ibid., p. 362).

Since evaluating the performance of public production requires the attribution of some values to the output, performance analysts face a grave dilemma:

> The customer [demanding a private good] is in a position to estimate the value to himself of an additional supply of service from a firm because he can experiment with different levels of purchase. His estimate of the value of the goods shows up in the marginal revenue the firm receives. By contrast, in the case of a collective good going simultaneously to many individuals, the individual consumer cannot take more or less to see how that effects his wellbeing. (Ibid., p. 363)

This second difficulty leaves us with the dilemma that the production functions for public goods and externalities cannot be established and production cannot be evaluated in terms of technical efficiency (ibid., p. 405). Since this will also be true when public enterprises simultaneously engage in the joint production of private and public goods, a different approach will have to be devised.

While Olson's point is extremely important and very basic, it does not lead immediately to ready-made solutions to improve the efficiency of public sector production. His own suggestions (selective and experimental pricing and spot controls) might help on occasion, but if systematically used would likely invite inefficient responses on the part of the public institution thus controlled (see Peirce, 1977). Not surprisingly, agreement could not be reached at the NBER conference, and Olson's important paper seems not to have had major consequences so far. Fortunately, a completely different strand of economic literature has developed right to the point that it can be linked up with Olson's approach.

Improving performance when objectives are vague

The economics of public enterprises and public administrations has traditionally received the attention of public finance theorists. Students used to be trained in methods of assessing the performance of public institutions and ensuring administrative efficiency. Public finance developed as a discipline separate from political economy because its subject matter, the conduct of state institutions, required a combination of economic theory and the art of public management and accounting. In the aftermath of the Keynesian revolution, applications of the Keynesian paradigm became for a while the main

focus of public finance research. The resulting decoupling between the discipline and its subject matter produced a gap which still remains to be filled satisfactorily in terms of modern political economy. Since a practical need was not addressed by the discipline, a substitute had to arise. In Germany, Ludwig Mülhaupt and his students played a major role in establishing the new sub-discipline of '*öffentliche Betriebswirtschaftslehre*' (public managerial economics). The essence of this approach is readily stated. In order to improve the performance of public institutions, modern management techniques are to be applied. The same techniques that have proved valuable in making corporations more efficient should also help public institutions more efficiently reach their objectives. This proposition raises the central question of work done by Günther E. Braun (1988).

What are the objectives of public institutions as compared to private corporations?

Braun's studies are organized in six chapters. A very short first chapter introduces the theme and summarizes the structure of the book. Chapter 2 discusses the research programme of public managerial economics. Chapter 3 introduces the methodology of comparative studies and Chapter 4 gives a general overview of the comparative analyses of objectives. Chapter 5 provides an in-depth description of the empirical study and Chapter 6 draws the conclusions for the possibility of using the planning methodologies developed for corporations in public institutions. The spectrum of public institutions discussed ranges from applied research organizations such as the Bundesgesundheitsamt (Federal Health Office), Bundesinstitut für Berufsbildung (a federal institute charged with researching vocational and professional education), the Gesellschaft für Mathematik und Datenverarbeitung and the Hahn-Meitner Institute (charged with research into computer science and nuclear physics, respectively), to the Federal Department of Agriculture, the Chancellery of the State of Rhineland Palatinate and several municipal welfare offices in North Rhine–Westphalia. As varied as these institutions are, they have two related characteristics in common: they are public and their objectives are vague. The vagueness of these objectives is throughout established and further detailed in 13 hypotheses, largely corroborated, with the exception of hypothesis 2. These hypotheses are as follows:

1. the more an agency is engaged in producing public goods, the less precise its objectives will be;
2. the more the objectives relate to technology and the natural sciences, the more precise they will be. This is the hypothesis that could not be corroborated; objectives relating to technology and the natural sciences are just as vague as those relating, for instance, to social policy;

3. the broader the menu of programmes at the top hierarchical level, the less precise its objectives will be;
4. the broader the service programme of a municipal office, the less precise its objectives will be;
5. the broader the service programme of a bureaucratic agency, the less precise its objectives will be;
6. the higher the external hierarchical level at which an institution operates, the less precise its objectives will be;
7. the higher the internal hierarchical level, the less precise its objectives will be;
8. the more professionalized an organization, the more precise its objectives will be;
9. the closer a public institution is to other actors in the political system, the less precise its objectives will be. The other actors may be various pressure groups, supervisory bodies, competing institutions and so on;
10. the more external participants are involved in decision making, the less precise its objectives will be;
11. the more long term and/or strategic the orientation of an organization, the less precise its objectives will be;
12. the more objectives serve as integrative or public relations devices, the less precise they will be; and
13. the less objectives are related to problems, programmes or projects, the less precise they will be.

Hypotheses 3, 6 and 8 could not be systematically tested with the available material.

Professor Braun uses these results to indicate where and to what extent corporate planning techniques can be tried in public sector organizations. Obviously, the more precise the objectives of public sector organizations are, the more applicable those planning techniques. On the other hand, softer planning techniques may be used where objectives are necessarily vague. The author is careful to point out that it would be wrong to insist on more precise definitions of objectives when the vagueness may be inherent and functional for public sector organizations. The very application of business techniques in public sector organizations can be a political ritual, and the benefits of the exercise should be considered in this light (Braun, 1988, ch. 6, § III). Performance studies commissioned by public authorities may sometimes serve a similar purpose.

The theoretical study by Olson and the empirical study by Braun both point in the same direction. The objectives of public institutions are typically vague and not readily expressed in terms of the received theory of the firm. The production of externalities and public goods is an important function of

public institutions, often pursued side-by-side with the production of private goods and services. As a rule, the performance of public institutions cannot be assessed with the same methods used for private firms. The exception to this rule is the case where a public enterprise is explicitly and exclusively run as a revenue source for the public purse. Although the objectives of public institutions seem vague and fuzzy, because they cannot be expressed in clear and simple terms, they can still be established with sufficient clarity to enable the analyst to assess the performance of the institution. While this claim can ultimately only be defended by pointing to a successful performance analysis on which agreement can be reached, the stepwise procedure developed in the following section may at least serve to indicate what can be expected from the alternative approach suggested here.

A performance analysis tries to establish what public institutions do and how well they do it

Before we can discuss the question of how well public institutions perform, we have to be certain what decision makers in these institutions are trying to achieve. The prime task of a performance analysis is therefore the establishment of the objectives of the public institution analysed; this task is difficult because, as we have seen, these objectives are typically left vague.

The point of departure for the analysis is the supposition that public institutions are efficiently operated in the Pareto sense. This supposition does not contradict Olson's assumption that public institutions are plagued by technical inefficiency. The supposition only implies that, given the objectives pursued, the mode of operation of that institution is not improvable because no better alternative is known that can readily be implemented. It follows from this central supposition that we can analyse the structure of a public institution in order to derive the objectives to be pursued therewith. Hence the first part of the performance analysis involves an analysis of the structure of the institution whose performance is to be researched.

Structure analysis

The central question to be answered with a structure analysis is this: which objectives can be efficiently realized given the prevailing structure of the institution or, alternatively, assignment of property rights? The analysis proceeds in four steps:

1. Who controls inputs and monitors input performance?
2. Who controls outputs and monitors output performance?
3. Who monitors the production process?
4. Who monitors the composition of the team of producers?

Since in a public institution the principle of residual clemency applies, not with respect to net profits, but with respect to (political) discretion, the assignment of property rights to particular actors whose interests can be established reveals the objectives of the institution. The objective will be the more multifaceted the more different actors exercise different property rights. These discretionary objectives have to be taken as the main objectives of the public institution, with the objective expressed in the charter of the institution serving as the satisficing constraint.

Analysing the revealed performance of an institution

While the structure analysis is exclusively devoted to establishing which objectives are to be reached, the analysis of the revealed performance yields insights about both aspects of the question; the revealed performance serves either to question or to corroborate the objectives established in the structure analysis; and it also offers hints as to what performance standards the actors in the institution wish to see applied themselves. The analysis proceeds in six steps:

5a. sample and define the objectives revealed in documents and pronouncements by authorized actors of the institution studied;
5b. sample and define the objectives found in non-authorized documents and pronouncements;
5c. compare the objectives established in steps (a) and (b) with those established by means of the structure analysis;
5d. compose a list of objectives differentiated as follows:
    d1. objectives established by both types of analyses,
    d2. objectives established by means of the structure analysis,
    d3. objectives established by means of the revealed performance analysis.

The complete list will be used for the following steps:

6. document the range of objectives;
7. document the extent to which these objectives have been achieved;
8. document the depth and continuity of achievement;
9. estimate the public enterprise rent and assess the achievement of the aforementioned objectives in terms of range, extent, depth and continuity against the public enterprise rent: that is, the profit forgone;
10. (optional) sketch a workable scenario of the privatization of all the activities leading to the full range of objectives established.

Analysis of the interplay between the institutions' own approach to account for its performance and the reactions of its principals to that account
The relative weights to be assigned to the different objectives can be derived from an analysis of the interplay between the efforts of the institution analysed to account for its performance and the reactions or non-reactions of the principals. The analysis again proceeds in five steps:

11. analyse the appointments to supervisory and management boards;
12. account for explicit policy changes and shifts in objectives between different institutions. Shifts will reveal a relative dissatisfaction with the performance of the institution which has to relinquish a particular task while at the same time expressing relative satisfaction with the performance of that institution which receives that assignment;
13. establish implicit policy changes and shifts not publicly discussed or documented;
14. establish discrepancies between language and behaviour in order to document ceremonial performance;
15. establish the range of activities not discussed at all but continued.

These activities can be safely assumed to be uncontroversial between principal and agent.

Summarizing the performance assessments in terms of inputs and outputs
Once, in the manner established, a full account of the performance of the institution has been given, a summary discussion should reveal the cost of achieving this range of objectives in terms of the public enterprise rent.

Concluding remarks
This entry has tried to grapple with the problem that public institutions serve a multitude of objectives that are difficult to establish and account for. The analysis is critical of studies which take the performance criteria from welfare economics. Instead, it is suggested that the criteria of performance assessment have to be derived, first, from an analysis of the structure of the institution and the objectives one can possibly achieve with this structure; second, from an analysis of the revealed performance as documented by the institution itself; and third, from an analysis of the interplay between principal and agent, that is, the public authority to which an institution is accountable and the institution itself. The entry concluded with the outline of a stepwise procedure for such a performance analysis.

References

Backhaus, Jürgen G. (1989), 'Privatization and nationalization: a suggested approach', *Annals of Public and Co-operative Economy*, **60** (3), 307–28.

Braun, Günther E. (1988), *Ziele in öffentlicher Verwaltung und Privatem Betrieb* [Goals in Public Administration and Private Enterprise] (Schriften zur öffentlichen Verwaltung und öffentlichen Wirtschaft, vol. 76) Baden-Baden: Nomos.

Gumble, Peter (1994), 'Expensive gamble, West-German program to salvage East shows signs of pay-off', *Wall Street Journal*, No. 113, 9 December.

Leibenstein, Harvey (1978), 'X-inefficiency Xists: reply to an Xorcist', *American Economic Review*, **68**, 203–11.

Marchand, M., P. Pestieau and H. Tulkens (eds) (1984), *The Performance of Public Enterprises: Concepts and Measurement*, Amsterdam: North-Holland.

Olson, Mancur (1973), 'Evaluating performance in the public sector', in Milton Moss (ed.), *The Measurement of Economic and Social Performance* (Studies in Income and Wealth, 38) New York and London: Columbia University Press for the National Bureau of Economic Research.

Peirce, William S. (1977), 'Bureaucratic failure; incremental drift from public goods', Working Paper No. 32, Cleveland: Case Western Reserve University, Department of Economics.

Pestieau, Pierre (1987), 'Measuring the performance of public enterprises: a must in times of privatization', contribution to the colloquium on Privatization in the European Community: A Comparative Perspective, Maastricht, 10–11 December.

Stigler, George J. (1976), 'The Xistence of X-efficiency', *American Economic Review*, **66**, 213–16.

Tulkens, Henry (1986), 'The performance approach in public enterprise economics', *Annalen der Gemeinwirtschaft*, **55** (4), 429–44.

Wagner, Richard E. (1977), 'Economic Manipulation for Political Profit: Macroeconomic Consequences and Constitutional Implications', *Kyklos*, **30**, 395–410.

22 Environmental law and economics

A. Allan Schmid

Law and economics has two contrasting approaches. One aspires to provide not only a positive theory of economic behaviour but also a guide to legal–economic policy choices. Efficiency and maximization of net economic product are claimed to be that guide (Faure, 1995). The second provides only a substantive description of the impacts of legal alternatives on different parties, with little attempt to aggregate them. These approaches will be contrasted with respect to several major legal questions concerning the environment, including the choice of injunction or liability as relief to injury, externalities and ownership, use of markets versus governmental regulation, and the takings issue. These will illustrate some of the theoretical and empirical findings in the environmental law and economics literature.

What is the appropriate relief when an actor causes environmental injury to another? The case of *Boomer* v. *Atlantic Cement Co.* (Court of Appeals of New York, 1970, 26 N.Y. 2nd 219, 257 N.E. 2nd 870) is often used to illustrate the problem (see Goetz, 1984). A large cement plant created smoke, dirt and vibration which damaged neighbours. The court found that the activity created a nuisance and awarded damages, but not an injunction. Cooter and Ulen (1988) argue that law and economics can provide an authoritative answer to the desirability of damages versus an injunction which legal analysis alone cannot. The landmark literature here is Calabresi and Melamed (1972) and Polinsky (1980). The authors argued that efficiency points to the awarding of damages to the many neighbours because otherwise there would be strategic bargaining and high transaction costs if an injunction caused the cement company to try to purchase the right to inflict these damages on the neighbouring homeowners. So cost minimization points to damages only since an injunction might leave the right in the hands of the homeowners even when the value was greater (net of transaction costs) in the hands of the cement company. (See also Ogus and Richardson, 1977.)

The efficiency literature assumes that the value to the cement company of use of the environment for noise and dirt deposit is greater than the use for the peace and quiet of home users. If efficiency is to be a guide to law, prices and values used to calculate efficiency must be independent of law and the assignment of rights. An alternative approach is provided by what Burrows and Veljanovski (1981) and Mercuro and Medema (1997) call ‘institutional law and economics’. While the efficiency approach uses the market price of

homes with and without exposure to cement factory pollution, imagine that some homeowners have unique tastes and would not sell at these prices. The institutional approach asks whether these unique tastes are to count (Calabresi, 1985). They would count if the homeowners own the right to an injunction, but not if they only have the right to court-computed damages based on easily observed market values. Legal choice is then a choice as to whether the unique tastes are reasonable, a matter which judges have long taken as their domain.

Once the cement factory has made its investment at a specific site, it has specific assets whose salvage value is much less if the site must be abandoned. The right of injunction then gives the homeowners a strategic bargaining advantage to obtain any rents the factory might otherwise have enjoyed. An empirical study of the behaviour of Florida phosphate mine operators and farmers illustrates the point (Crocker, 1971). Phosphate processing released dust which damaged adjoining pastures. Under Florida law the farmers had the equivalent of the right of injunction via the state pollution control authority. The miners purchased the pasture land at prices substantially above its market value. The value of the land resource is not independent of the legal alternatives under consideration and thus cannot be used as a guide to choosing between them (Samuels et al., 1997).

Economists have often served as expert witnesses in environmental damage cases. Among the most celebrated damage cases are those arising from oil spillage from tankers such as occurred in the wreck of the *Exxon Valdez* in Alaska and the *Amoco Cadiz* off the coast of France. To determine damage awards, economists have conducted contingent valuation studies asking people their willingness to buy the environmental products that were destroyed. But again, the economic value is not independent of rights. It is the rights determination that instructs whether the proper question for contingent valuation is willingness to buy (which assumes environmentalists do not own the resources) and willingness to sell (which assumes they are already owners).

Externalities and ownership

The assignment of liability is equivalent to the assignment of ownership among incompatible users of a resource. The touchstone of much law and economics scholarship here is the work of Ronald Coase. The efficiency school of law and economics has interpreted his work to mean that distribution of ownership is not important to achieve efficiency when transaction costs are zero. The market would achieve efficiency as long as the government was definitive about the placement of liability and no other government action was necessary. This is in contrast to the idea that a tax or regulation is necessary to correct externalities. If the person who places the highest value on the resource is given ownership, he or she will refuse all lesser market

bids and keep the resource. If the person with the lesser value is given ownership, he or she will sell to the person with the higher value and thus use of the resource is invariant with initial ownership. While the logic of this proposition is unassailable, Coase (1992) has now made it clear that most practical issues of resource conflicts involve transaction costs and thus initial ownership matters.

Coase and the institutionalists argue that conflicts over externalities are ubiquitous and reciprocal between alpha and beta, and the policy question is to decide which party to restrain. Coase argues that, in a world of significant transaction costs, it is not possible to deduce a general rule of taxation, liability or whatever (see Parisi, 1995, for a summary). He supports the calculation of net benefits on a case-by-case basis with the presumption that any legal innovation which reduces transaction costs is desirable. Institutional law and economics suggests that transaction costs are often the means by which the interests of third parties are protected and the reduction of such costs is functionally equivalent to a redistribution of rights, opportunities and exposures (Schmid, 1983).

Markets versus regulation

The efficiency approach to law and economics is the basis for a critique of many countries' current policies requiring certain standards for discharges or specific technologies to be used by all firms engaging in certain resource-using activities. Requiring all firms to use a particular wastewater treatment, for example, ignores the possibility of differences in costs among firms. Starting from some initial allocation of rights, if firms could buy waste discharge rights, those with cheap alternatives would sell and those with more expensive alternatives would buy the right to discharge rather than treat. Some countries are beginning to utilize pollution markets, but some people have objected to firms making profits from the right to pollute. As a positive proposition, groups striving to capture opportunities for profits do lobby for legal change to make it possible. At the same time, legal change is also driven by ideology. One powerful ideology is that firms should have only enough waste disposal rights to keep in business and generate employment. If technology improves, the environmental users should benefit from the achievement of higher quality, rather than the polluters having something to sell. Institutional law and economics suggests the interdependence between legal change and notions of fairness. The legitimacy of the market depends on the legitimacy of who has what to sell.

Takings or the extent of legal change

Legal systems in the English tradition make a distinction between legitimate change in the regulation of activity affecting the environment (police power)

and legal change which constitutes a 'taking' which must be compensated by the government. Whether or not the legal system has a constitutional provision that limits the uncompensated consequences of legal change, the issue of compensation is present in all legal systems. The efficiency approach to law and economics applies the test of Pareto improvement: does the change make someone better off without making someone else worse off? So it might be asked whether a proposed zoning which would prevent a real estate development to preserve a scenic view would meet the test (Knetsch, 1983; Stephen, 1988, ch. 6).

But the Pareto test requires some assumption of initial rights to calculate. If we ask about the developer's willingness to sell and forgo development, we assume that the developer has the right to create costs for the environmentalist. This presumes that the environmentalist is the buyer and has no land rights, but the rights situation could be reversed and the same test applied. If we ask about the environmentalist's willingness to sell and forgo the view, we assume that the environmentalist has the right to withhold opportunities for the developer. Willingness to pay and willingness to sell may be equal, but need not be if there is a wealth or endowment effect (Schmid, 1995, 2004).

A zoning law may shift use of the view to the environmentalist when those users would not have been able to buy the view from developers. Is this a loss of efficiency? But some environmentalists would not sell if they were the owners of the view. The environment is not divisible, allowing some to sell and others to keep. The institutional law and economics approach finds that the Pareto rule does not dispose of the issue. Since value depends on the distribution of rights, it cannot be a guide to the choice of those rights (Samuels and Mercuro, 1992).

The government is not the only source of change affecting the value of a person's opportunities. There are changes in technology, resource availability, population, preferences and so on. Many of these are subject to radical uncertainty and no law can fully anticipate them. When they occur, new interdependencies will arise and the inevitable question is who can seize them and who is subject to uncompensated losses. Samuels and Mercuro argue that the compensation problem is ubiquitous, that uncompensated takings are inevitable and that no analytically significant distinction can be made between legally and illegally imposed changes.

Human behaviour

The study of law and economics incorporates divergent views of human behaviour. Some find it sufficient to assume individual utility maximization involving largely unchangeable preferences. Others find it useful to acknowledge that people care about others and that preferences evolve and are subject to social interaction. Thus, for example, there are differences in predicting whether

people follow environmental laws when the costs of enforcement and monitoring are high. Some would argue that law is only effective if it can be enforced, and others that there is some role for law as a statement of ideals which some may follow even if sanctions cannot be made effective. For example, it is difficult to police people who litter. Nevertheless, many people do not litter, even when avoidance has a cost and policing is impossible. One area where the behavioural postulate is critical is with respect to judges. If judges followed a narrow self-maximization principle, they would always sell their services to the highest bidder when these acts could not be monitored. The chaos that follows from the fact that some do, and that in some countries most do, suggests the value of self-imposed learned behaviour and makes the study of institutions which affect this learning a priority for scholarship.

References

Burrows. P. and C.G. Veljanovski (eds) (1981), *The Economic Approach to Law*, London: Butterworths.

Calabresi, G. (1985), *Ideals, Beliefs, Attitudes and the Law*, Syracuse, NY: Syracuse University Press.

Calabresi, G. and A.D. Melamed (1972), 'Property rules, liability rules and inalienability: one view of the cathedral', *Harvard Law Review*, **85**, 1089–128.

Coase, R.H. (1992), 'The institutional structure of production', *American Economic Review*, **82**, 713–19.

Cooter, R.D. and Thomas Ulen (1988), *Law and Economics*, Glenview, IL: Scott, Foresman.

Crocker, T. (1971), 'Externalities, property rights and transactions costs: an empirical study', *Journal of Law and Economics*, **14**, 451–64.

Faure, M. (1995), 'Economic models of compensation for damages caused by nuclear accidents', *European Journal of Law and Economics*, **2**, 21–34.

Goetz, C.J. (1984), *Cases and Materials on Law and Economics*, St Paul, MN: West Publishing.

Knetsch, J.L. (1983), *Property Rights and Compensation*, Toronto: Butterworths.

Mercuro, N. and S. Medema (1997), *Economics and the Law: From Posner to Post-Modernism*, Princeton, NJ: Princeton University Press.

Ogus, A. and G. Richardson (1977), 'Economics and the environment: a study of private nuisance', *Cambridge Law Journal*, **36**, 284.

Parisi, F. (1995), 'Private property and social costs', *European Journal of Law and Economics*, **2**, 149–73.

Polinsky, A.M. (1980), 'On the choice between property rules and liability rules', *Economic Inquiry*, **18**, 233–46.

Samuels, W.J., S.G. Medema and A.A. Schmid (1997), *The Economy as a Process of Valuation*, Cheltenham, UK and Lyme, USA: Edward Elgar.

Samuels, W.J. and N. Mercuro (1992), 'The role and resolution of the compensation principle in society', in W.J. Samuels, *Essays on the Economic Role of Government*, Vol. 1, New York: New York University Press.

Schmid, A.A. (1983), 'Die Okonomische Theorie der Eigentumsrechte: Ein Besprechungsaufsatz', in J. Backhaus and H. Nutzinger (eds), *Eigentumsrechte und Partizipation*, Frankfurt: Haag & Herchen.

Schmid, A.A. (1995), 'The environment and property rights issues', in D. Bromley (ed.), *Handbook of Environmental Economics*, Oxford: Blackwell, pp. 45–60.

Schmid, A.A. (2004), *Conflict and Cooperation: Institutional and Behavioral Economics*, Oxford: Blackwell.

Stephen, F.H. (1988), *The Economics of Law*, Ames, IA: Iowa State University Press.

23 Environmental policies choice as an issue of informational efficiency

Donatella Porrini

Choosing between different environmental policy instruments

The problem of the choice of environmental policy instruments has been an issue since Pigou (1932) analysed the need for state intervention when private costs diverge from social costs, and suggested that the solution would be to internalize the externalities through taxation. Coase (1960) criticized the proposed state intervention, and affirmed that there is no reason to suppose that governmental regulation is called for simply because the problem is not very well handled by the market or the firm. The key feature is the presence of transaction costs that make one policy better than another.

The ensuing debate has been conducted along these two opposite views: on the one hand, the supporters of the idea that the choice of policy instruments to be applied following a market failure is a public matter and the state, as policy designer, should select the optimal instrument and take responsibility for its imposition in the public interest; versus, on the other, the supporters of market-based instruments, trying to fight a battle against a sort of 'anti-market' mentality based on a reluctance to apply market-oriented instruments (Lewis, 1996).

If we want to continue along these lines the problem would be to compare the efficiency of instruments that can be considered 'public oriented' and those that can be considered 'market oriented', where the first are characterized by a public agency with a public definition of conduct rule and a public enforcement system and the second are instruments based on market mechanisms stimulating the conduct of the firm indirectly and characterized by a private administration and enforcement system.[1]

Given this premise, the definition of a superior and not ideological line to be followed in the choice between different environmental policy instruments would appear to be a very difficult task.

But looking at this problem from a law and economics perspective, we can move from the theoretical definition of the efficiency of different instruments to their practical, and so direct, potential to achieve concrete objectives. In particular, three objectives emerge as relevant in judging the practical efficiency of environmental policies: the first is paying accident compensation to the victims; the second is prevention, in the sense of providing incentives for

firms to improve safety standards; and the third is connected with technological change in the sense of encouraging firms to adopt lower-risk technologies.

Given these three objectives, this chapter will analyse the efficiency of environmental policies, focusing on the imperfections that can emerge in their practical applications. In particular we shall concentrate on informational problems that can characterize the activities of the agencies and private firms in relation to the efficient implementation of the different environmental policy instruments.

In order to specify and precisely define the various instruments, Figure 23.1 presents a scheme of the different environmental policy instruments that will be examined. The first choice is between liability and regulation, where the latter includes a subdivision between, on one hand, a 'command-and-control' form of regulation based on the definition of standards and, on the other, market-based instruments as an indirect form of incentive for private firms. Another subdivision can be considered at this point: typically, command-and control regulation can be either technology or performance-based standards.

Market-based instruments can also be of different kinds: taxes or tradable permits. On the other side, liability is an instrument based on the judiciary system, and can be assigned on the basis of a negligence or a strict liability

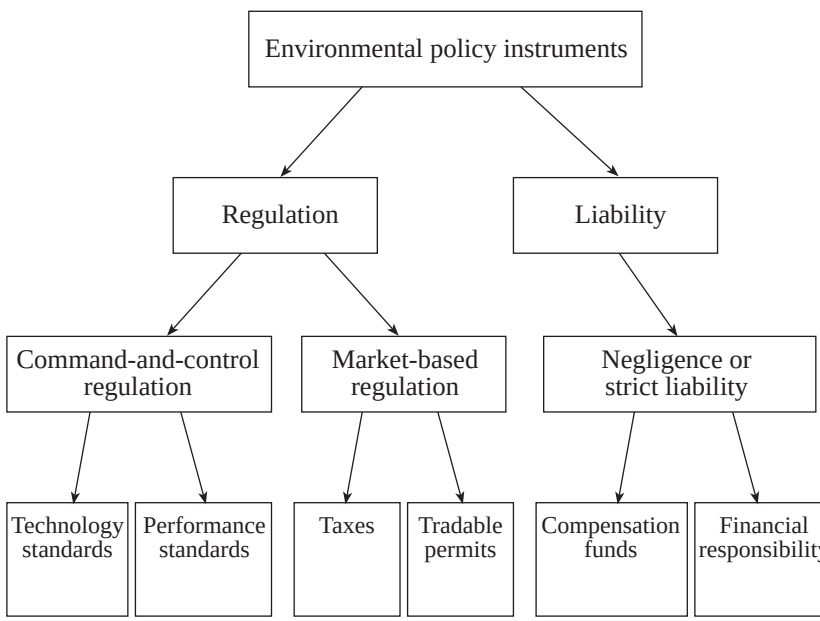


Figure 23.1 Scheme of different environmental policy instruments

regime. Finally the limit of 'judgment proofness', which can arise from the application of a liability system when the resources of the responsible party fall short of the damage amount, can be solved in two ways: a compensation fund and a financial responsibility system.

Before defining regulation, liability and the other instruments, it should be noted that in practice it is difficult to refer to a specific policy instrument, given that the environmental policy choice usually involves a mix of them.

Regulation versus liability

> Regulation and tort law are alternative methods (though often used in combination) for preventing accidents. The former requires a potential injurer to take measures to prevent the accident from occurring. The latter seeks to deter the accident by making the potential injurer liable for the costs of accident should it occur. (Landes and Posner, 1984, p. 417)

However, before dealing with the problem of the comparison between the two instruments we shall first define their main features.

A regulation system is typically characterized by a centralized structure in the sense that it is a system based on an authority or an agency that uses a number of tools to control environmental damages. By 'regulation' we mean 'a directive to *individual* decision-makers requiring them to set one or more output or input quantities at some specified levels or prohibiting them from exceeding (or falling short of) some specified levels' (Baumol and Oates, 1975, p. 126). Usually the regulation system takes the form of the setting of standards: in this case under a mandatory technology or abatement standard, the regulator can order the firms to reduce their emissions by a certain percentage, to emit no more than a specified amount of a pollutant, and/or to install a particular abatement technology. Alternatively, there are incentive market-based instruments, such as marketable permits and taxes (Backhaus, 1999).

First, we can affirm that in a world with perfect or at least complete information, following a law and economics approach, the policy instrument consisting in a regulatory system is efficient in solving the problem of internalizing the potential effects of environmental accidents. In fact, using this instrument, *ex ante* the firm has an incentive to take adequate precautions. But 'problems of measurement and the breakdown of second-order conditions (among other things) constitute formidable obstacles to the determination of truly first-best environmental policy' (Cropper and Oates, 1992, p. 685).

In fact, in an incomplete information context many problems can arise. These will be analysed in the next section in connection with the different forms of regulation.

A liability system for environmental damages can be considered to be a policy instrument in the sense that such a system provides protection for the

victims against the consequences of an environmental accident and gives incentives to the actors in a potential accident to take the necessary preventive measures (Calabresi, 1970; Shavell, 1987). Consider the typical liability system applied to risks created by hazardous activities: in this case the victim files an action against defendants for all injuries caused by their conduct, claiming a causal link between the defendants' conduct and the plaintiff's injury.

A liability system can be applied using either a negligence or a strict liability regime. The law and economics literature generally concludes that both regimes provide a potential polluter with incentives to take adequate preventive measures. But problems arise if we consider the practical application of the regimes and the presence of informational issues. Regarding the specific case of environmental accidents, it is particularly difficult to determine the standard to assign liability on the basis of negligence: for example, pollution has many sources and many victims and it is a hard task to prescribe efficient pollution standards based on a calculus of the abatement cost and the external harm of every source of pollution.

In fact, in the case of assignment of liability for an environmental accident, a strict liability regime is applied more often. In the United States, the CERCLA (Comprehensive Environmental Response, Compensation and Liability Act 1980, 1985, 1996) type of liability is typically a strict, joint and several liability, on the owners and operators of the firm that is responsible for an environmental accident. The liability system in the European Union, as set out by the Commission, in the *White Paper on Environmental Liability*,[2] is essentially a strict (no-fault) and non-retroactive liability system for damage caused by inherently dangerous activities.[3]

The two policy instruments, regulation and liability, may present different informational issues, as in Rose-Ackerman (1991, p. 54),

> Statutory regulation, unlike tort law, uses agency officials to decide individual cases instead of judges and juries; resolves some generic issues in rulemakings not linked to individual cases, uses nonjudicialized procedures to evaluate technocratic information, affects behavior *ex ante* without waiting for harm to occur, and minimizes the inconsistent and unequal coverage arising from individual adjudication. In short, the differences involve who decides, at what time, with what information, under what procedures, and with what scope.

In a comparison of regulation and liability, Shavell (1984a) considers as a first determinant the difference in know-how between private parties and the regulatory authority related to the benefits of activities, the cost of reducing risks, and the probability and the severity of accidents. It is evident that the nature of the activities carried out by firms is such that private parties could have a better knowledge of the benefits, the risks involved and the cost of

reducing risks. In such a case a liability system is better because it makes private parties the residual claimants of risk control. However, in some cases the regulator is better informed because of the possibility of centralizing information and decisions, in particular when a knowledge of risks requires special replicable and reusable expertise. In such a case, direct regulation is likely to be the better system.

A second determinant is the limited capacity of private parties to pay the full costs of an accident, either because of limited liability or because of insufficient assets. Shavell claims that a traditional liability regime does not provide private parties with proper incentives for taking precautions. A regulatory system can impose decisions on firms either directly or indirectly. So, the greater the probability or the severity of an accident, and the smaller the assets of the firm relative to the potential damages, then the greater the appeal of regulation. But, as we shall see, public funds and financial responsibility can be applied.

The third determinant is the likelihood that the responsible parties would face a legal suit for harm caused. This is a particular problem in environmental risks: in many cases the victims are widely dispersed; they may not be motivated to initiate a legal action; harm may be evident only after a long delay; and specifically responsible polluters may be difficult to identify. Compared with a regulatory system, the liability system is more uncertain and provides less incentive for risk control.

The fourth determinant is the amount of administrative expense incurred by the private parties and the public. The cost of a liability system includes the cost of compliance, the legal expenses and the public expenses for maintaining legal institutions. The cost of the regulatory system includes the public expenses for maintaining the regulatory agencies and the private costs of compliance. In this case the advantage of the liability system is that legal costs are incurred only if a suit is brought and, if the system works well, in the sense that there are incentives for choosing the efficient level of care, the suits are few and therefore the costs are low. On the other hand, under regulation, the administrative costs are incurred whether or not the harm occurs because the process of regulation is costly by itself and the regulator needs to collect information about the parties, their activities and the risks.

Considering the four determinants, Shavell concludes that administrative costs and differences in knowledge favour liability, while inability to pay and avoiding lawsuits favour regulation. In general, a liability system is more efficient when private parties possess better information and when there is a low probability that an accident will occur. Regulation is usually better when harm is great, is spread among many victims or takes a long time to become apparent, when accidents are not very rare events, and when standards or requirements are easy to establish and control.

Shavell (1984b) deals with the characterization of the relationship between the regulation systems, as complements or substitutes in providing incentives to reduce the level of risk, showing that no single regulation system leads the parties to exercise the socially desirable level of care. This is due to different factors: on the one hand, the agency suffers from a lack of information and, on the other, a liability system presents the possibility that the parties would not pay fully for harm and might not even be sued. Shavell first considers the case where only the *ex post* regulation system or the *ex ante* one can be used and then the case where both systems can be used jointly so that the parties must satisfy *ex ante* standards and are also subject to *ex post* liability. The conclusion is that it is generally socially advantageous to use both *ex ante* and *ex post* regulation systems.

In another contribution, Kolstad et al. (1990) stress that the two policies may be complementary. They claim that even if the phenomenon of complementary use of *ex ante* and *ex post* systems is widespread, the economic literature has generally studied the two separately, characterizing each of them by different inefficiencies. In addition to Shavell's (1984a) four determinants, Kolstad et al. consider the imperfection in the definition of legal standards which may lead firms to choose a level of precaution different from the socially optimal one. They conclude that the liability system, applied jointly with *ex ante* regulation, can correct the above inefficiencies, at least in part.

Other contributions try to include informational issues in the analysis of the comparison between *ex ante* and *ex post* regulation systems. Schmitz (2000) presents a critical evaluation of the above papers suggesting the use of the two systems as complementary instruments to overcome the limited efficiency of liability due to enforcement errors and to injurers avoiding lawsuits. Schmitz proposes the comparison between *ex ante* and *ex post* systems as imperfect instruments through a formal model of how the imperfections affect the outcome: the extension of liability to private financiers is imperfect in so far as the private financier maximizes his/her own profit rather than social welfare; the regulatory agency may be captured by the parties who may cause environmental accidents; an asymmetric information framework is considered where the level of precautionary activities is the private information of the firm. The author shows that if injurers cannot avoid a lawsuit and if the magnitude of liability is set at the optimal level, it can never be socially advantageous to employ both the systems as complementary instruments if all injurers face the same wealth constraints. But joint use can be valuable if wealth varies among injurers and in the latter case, the regulatory standard can be set at a level that is lower than the one corresponding to the social optimum obtained when only *ex ante* regulation is used.[4]

This analysis of the contributions on the comparison between *ex ante* and *ex post* regulation systems has shown some results in terms of the choice of

environmental policies, including informational issues. In the next section we shall analyse in turn different instruments included in the general definition of regulation and liability.

Analysing different forms of regulation

Generally the implementation of any form of environmental regulation requires that the quantity of polluting emissions and the monetary costs of the damage caused by an eventual accident should be determined. This implies setting up a monitoring procedure and then using regulatory tools to set standards or distribute the cost to firms through a tax or through permits based on their polluting emissions.

In each case, informational problems can derive from closely monitoring the firm's conduct, for example, its emission levels, and these problems can lead to an inefficient level of enforcement and to overdeterrence. But these issues need to be analysed in connection with each regulatory instrument.

There are different forms of regulation: public-oriented command-and-control instruments that require the use of a particular technology or the observation of a performance standard, prescribing the maximum amount of pollution that a source can emit; and market-oriented instruments that are essentially pollution taxes or a system of tradable permits. Supporters of command-and-control technology requirements have clashed with devotees of incentive-based approaches advocating taxes and tradable allowances (Wiener, 1999).

Under highly restrictive conditions, it can be shown that both environmental policy instruments share the desirable feature that any gains in environmental quality are obtained at the lowest possible cost (Baumol and Oates, 1975). Hahn and Stavins (1991, p. 6) commented:

> Theoretically, the government could achieve such a cost-effective allocation of the pollution control burden among sources if it could ensure by some means that all sources controlled at the same marginal cost. However, such an approach would require the government to have detailed information about the cost functions of individual firms and sources – information that the government clearly lacks and could obtain only at great cost, if at all.

With regard to command-and-control instruments, regulatory measures are generally defined and imposed by agencies that prescribe what measures a firm should take to prevent harm. Therefore, the existence of an agency charged with meeting these objectives is assumed. The essence of the agency activity is to control the actions of many individuals and independent actors (firms, households, other government units), and to induce them to take constraining actions contrary to their narrow self-interests (Bohm and Russell, 1985). These measures can be imposed by general rules or individual licences,

taking the form of emissions standards based on a particular quality or quantity of emissions in the environment. Non-compliance with such standards is usually enforced by administrative or criminal sanctions.

Command-and-control instruments set uniform standards for firms: on one hand, such instruments force all firms to shoulder an equal share of the mitigation burden, regardless of the relative costs of this burden to them even if the same firms can adopt preventive measure at much less cost than others; and on the other, the command-and-control instruments directly and effectively limit dangerous emissions.

In its application, this regulation system has proved to be well suited to setting policies regarding the definition and implementation of standards. The centralized search facilities, the continual oversight of problems and a broad array of regulatory tools can make the regulation system capable of systematically assessing environmental risks and of implementing a comprehensive set of policies. But, regulatory agencies may not be very flexible in adapting to changing conditions, and a centralized command structure relying on expert advice may be subject to political pressure as well as to collusion and capture by the regulated firms.

We can distinguish between different kinds of standards: technology or performance. In the former, the firm is not free to choose the measures by which it will achieve a certain environmental quality; this is literally a command-and-control instrument which imposes a certain technology that has to be used by the firm. In the latter, there is some degree of freedom for the firm in the sense that the standard determines the amount and the quality of substances that the firm can emit but then the firm can choose the technology to achieve it.

The performance-based standard does not stipulate any particular equipment to be used to comply with a regulation to achieve a specific ecological goal, thus giving private parties a certain amount of flexibility. This feature can be an advantage in relation to the informational problems connected in general with the use of command-and-control instruments: the agency activities are not costless – checking the behaviour of the actors against applicable regulatory orders, or determining what is owed by way of emission charges implies some expense; the activity of monitoring is another cost for the agencies; there is also the so-called 'capture' problem – public agencies can be motivated by financial rewards and promises of promotion or there can be a connection between their own and their firm's interests given that they are vulnerable to bribery from third parties or from the offenders they are supposed to monitor.

We can now analyse other kinds of regulatory instruments, which, in contrast to command and control, do not directly prescribe what the behaviour of potentially polluting firms should be. A Pigouvian tax, for example, is a way to attribute a price to pollution that will be incorporated by the firm in

the price of its products, but the incentive for the adoption of abatement techniques relies on the market mechanism because if a firm does not apply the optimal techniques, it will produce more pollution, pay more taxes and sell its products at a higher price than its competitors. Market-based instruments as regulatory devices that shape behaviour through price signals rather than explicit instructions on pollution-control levels or methods, are often described as 'harnessing market forces' because they can encourage firms and individuals to undertake actions that serve both their own financial interest and public policy goals (Stavins, 1998).

Using these instruments, rather than traditional command-and-control ones, provides a dynamic incentive for technology innovation. This is accomplished by allowing firms to share the burden of pollution control more efficiently through encouraging them to achieve reductions in pollution more cheaply. So market-based instruments such as taxes and tradable permits should generally be preferred to technology requirements and fixed emissions standards because the incentive-based instruments are typically far more cost-effective and innovation generating than their alternatives (Keohane et al., 1997). In particular, these instruments could provide continuous dynamic incentives for the adoption of superior technology, since it is always in the interest of firms to clean up more if sufficiently inexpensive clean-up technologies can be identified (Jaffe and Stavins, 1995).

Compensation funds versus financial responsibility

A very important problem connected with liability policy instrument is the limit of a firm's financial resources compared with the amount of the damages that could derive from an environmental accident. This problem arises when identified polluters are 'judgment proof' and so not able to pay for the total cost of the environmental damage. Moreover, given that they do not pay the full cost, then they are not motivated to adopt an adequate preventive measure (Summers, 1983; Shavell, 1986).

In the United States this problem arose in cases of smaller firms involved in risky production activities (Ringleb and Wiggins, 1990). From an economic point of view, this is a problem of internalization in the sense that some of the losses of the victims may go unclaimed under conventional strict liability; moreover in some cases, firms facing considerable liability risks may reduce their capital using 'judgment proofness' as an evasion strategy (Van't Veld et al., 1997).

Among internalization instruments, there is one that uses a compensation fund. Usually funds are created in connection with a regulatory system to cover environmental damage, contaminated site costs and victim compensation amounts. The fund can be financed by a taxation system or by a firms' contribution system.

The most important application of this instrument is the one by CERCLA, which enabled the Environmental Protection Agency (EPA) to clean up contaminated waste sites directly by utilizing funds from the Hazardous Substances Response Trust Fund, commonly known as the 'Superfund'. The Superfund was created to provide the federal government with the financial resources necessary for cleaning up contaminated sites and facilities. The fund is financed through a combination of federal appropriations, industry taxes and judgments entered against responsible parties. CERCLA authorized the EPA to target specific contaminated sites across the country and to rank those sites through a national priority list (NPL), which generally determines the order in which the various sites will be cleaned up.

So on the one hand, as in the US experience, compensation funds as environmental policy instruments prove to be an efficient kind of emergency tool when a quick intervention is necessary. On the other hand, in the implementation of these public-oriented instruments many problems can arise from a distributional point of view, because the existence of a public fund generates social costs connected with taxation sources that make income distribution problems relevant (Lewis, 1996). Moreover from a law and economics point of view, the literature shows that this system can result in a lack of motivation by firms to adopt preventive measures (Porrini, 2001).

We can also analyse another kind of instrument as an alternative to solve the judgment-proof problem: financial responsibility. By this, we consider all the tools that require polluters to demonstrate *ex ante* sufficient financial resources to correct and compensate for environmental damage that may arise through the activities of a firm. In its common application, financial responsibility implies that the operation of hazardous plants and other business is authorized only if firms can prove that future liability claims will be financially covered, for example, letters of credit and surety bonds; cash accounts and certificates of deposit; self-insurance and corporate guarantees. Letters of credit and surety bonds are purchased from banks or insurance companies: they are paid to a third-party beneficiary, often the government, under certain circumstances such as the failure of the purchaser to perform certain obligations. Cash accounts and certificates of deposit place cash or some other forms of interest-bearing security into accounts that are made payable or assigned to a regulatory authority. Self-insurance is purchased by companies with relatively deep pockets to satisfy coverage requirements by demonstrating sufficient financial strength. A corporate guarantee allows another firm, such as a parent corporation, to satisfy the coverage requirement and financial guarantors must themselves agree to cover the liabilities of the firm (Boyd, 2002). Since the 1980s, financial responsibility has been widely applied in the United States within the framework of the liability assignment system for environmental damage.[5]

While a market for assurance coverage has developed in the United States to provide a wide variety of financial instruments tailored to individual firms and regulatory needs, in the European Union this kind of instrument has a corresponding importance but relatively little diffusion.[6] However, this does not exclude the possibility that financial responsibility instruments have already been provided within the individual member states, and in fact some national enforcement has occurred.[7]

These experiences show that financial responsibility may be complementary, sometimes mandatory, to the legislation on liability assignment of environmental damage. It is usually required as an integral part of some kind of *ex ante* regulation, to ensure that the damaged natural resources are made good. In its different applications, it has a common motivation: to ensure the future internalization of the costs caused by the polluter in order to indemnify the victims and discourage different forms of environmental deterioration.

In the presence of informational issues, financial responsibility can also be seen as a solution to asymmetric information problems that can arise in the relationship between firms and the financiers (Porrini, 2002). First, given the contractual relationship between the financial institutions and the firms, there is a strong incentive for the financial institutions, insurance companies or banks, to check that the firm is taking adequate preventive measures (Feess and Hege, 2000). Second, the firm itself is motivated to take precautions because financial responsibility ensures that the expected costs of environmental risks appear on the firm's balance sheet and in its business calculation.

Concluding remarks

In this contribution we have analysed from a law and economics point of view the efficiency of different environmental policy instruments on the basis of the achieved targets and taking informational problems into account.

The literature on the choice between regulation and liability has been reviewed, showing evidence of an increasing interest in informational issues.

Concerning other forms of environmental policies inside the two main categories of regulation and liability, as a general rule whenever the nature of the activities carried out by the firms is such that the private parties have better information about the benefits and costs of reducing risks, then the market-based system is to be preferred. The advantage of making the private parties directly responsible for risk control is clear but an indirect form of involvement, such as through a tax system or through financial responsibility, could also have positive effects.

Moreover, when there are large differentials among firms in the abatement cost of pollution, relying on market-based instruments provides the advantage of economies on the need for public agencies to acquire information. But it is also possible that a public agency is better informed about those risks

because information and decisions can be centralized, particularly when a better awareness of the risk factors requires special expertise to be shared in different cases and situations.

Finally, an important advantage emerges in the analysis of the enforcement of financial responsibility because through this instrument financial institutions, such as banks or insurance companies, can play an important role by using the capital market to create guarantees in favour of companies operating in risky sectors.

Notes

1. For the contrast between private and public enforcement, see Cooter (1984).
2. Commission of the European Communities, *White Paper on Environmental Liability*, COM (2000), 66 final, Brussels, 9 February 2000.
3. In the words of the European Commission: 'Strict liability means that fault of the actor need not be established, only the fact that the act (or the omission) caused the damage. At first sight, fault-based liability may seem more economically efficient than strict liability, since incentives towards abatement costs do not exceed the benefits from reduced emissions. However, recent national and international environmental liability regimes tend to be based on the principle of strict liability, because of the assumption that environmental objectives are better reached that way. One reason for this is that it is very difficult for plaintiffs to establish fault of the defendant in environmental liability cases. Another reason is the view that someone who is carrying out an inherently hazardous activity should bear the risk if damage is caused by it, rather than the victim or society at large. These reasons argue in favour of an EC regime based, as a general rule, on strict liability' (see note 2: para. 4.3, under the title 'The type of liability, the defences to be allowed and the burden of proof').
4. The comparison between liability and regulation can also be modelled using a fomal economic approach based on a principal–agent kind of representation. See Boyer and Porrini (2002a, 2002b).
5. Financial responsibility is provided for by CERCLA, by the Safe Drinking Water Act (SDWA), by the Outer Continental Shelf Lands Act (OCSLA), and by the Surface Mining Control and Reclamation Act (SMCRA). Also in the Resource Conservation and Recovery Act (RCRA) and in the Oil Pollution Act (33 U.S.C. §2716 of 1990).
6. In fact, in §4.9 of the White Paper, on 'Financial security', we can find the statement: 'When looking at the insurance market – insurance being one of the possible ways of having financial security, alongside, among others, bank guarantees, internal reserves or sector-wise pooling systems – it appears that coverage of environmental damage risks is still relatively undeveloped, but there is clear progress being made in parts of the financial markets specialising in this area'. And the enforcement of such an instrument seems to be delayed in time, according to the statement that 'Moreover, the EC regime should not impose an obligation to have financial security, in order to allow the necessary flexibility as long as experience with the new regime still has to be gathered. The provision of financial security by the insurance and banking sectors for the risks resulting from the regime should take place on a voluntary basis'.
7. For example, in Italy, the Ministero dell'Ambiente [Ministry of the Environment], in a decree of 8 October 1996, defined the method for granting financial guarantees in favour of the state by companies that carry out waste transportation activities related to reclaiming, restoration of site conditions, waste transportation and disposal, as well as the reimbursement of any further damage caused to the environment. Another example is the Flemish experience, and more particularly the proposals of the Interuniversity Commission for the revision of environmental law in the Flemish region, which made elaborate provisions concerning financial guarantees (Faure and Grimeaud, 2000).

References

Backhaus, J. (1999), 'Regulatory taxation', in J. Backhaus (ed.), *Elgar Companion to Law and Economics*, ch. 16, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Baumol, W.J. and W.E. Oates (1975), *The Theory of Environmental Policy*, Englewood Cliffs, NJ: Prentice-Hall.

Bohm, P. and C.S. Russell (1985), 'Comparative analysis of alternative policy instruments', in A.V. Kneese and J.L. Sweeney (eds), *Handbook of Natural Resource and Energy Economics*, vol. 1, ch. 10, Amsterdam: North-Holland.

Boyd, J. (2002), 'Financial responsibility for environmental obligations: are bonding and assurance rules fulfilling their promise?', in T. Swanson and R. Zerbe (eds), in *An Introduction to the Law and Economics of Environmental Policy: Issues in Institutional Design*, Oxford: JAI-Elsevier, pp. 417–85.

Boyer, M. and D. Porrini (2002a), 'Law versus regulation: a political economy model of instrument choice in environmental policy', in Anthony Heyes (ed.), *Law and Economics of the Environment*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 249–79.

Boyer, M. and D. Porrini (2002b), 'The choice of instruments for environmental policy: liability or regulation?, in T. Swanson and R. Zerbe (eds), *An Introduction to the Law and Economics of Environmental Policy: Issues in Institutional Design, Research in Law and Economics*, 20, Oxford: JAI-Elsevier, pp. 247–69.

Calabresi, G. (1970), *The Cost of Accident*, New Haven, CT: Yale University Press.

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

Cooter, R. (1984), 'Prices and sanctions', *Columbia Law Review*, **84**, 1523–60.

Cropper, M.L. and W.E. Oates (1992), 'Environmental economics: a survey', *Journal of Economic Literature*, **30**, pp. 675–740.

Faure, M. and D. Grimeaud (2000), 'Financial assurance issues of environmental liability', Report for the European Commission, (www.europa.eu.int/comm/environment/liability/insurance_gen.htm).

Feess, E. and U. Hege (2000), 'Environmental harm, and financial responsibility', *Geneva Papers of Risk and Insurance: Issue and Practice*, 25, 220–34.

Hahn, R. and R. Stavins (1991), 'Incentive-based environmental regulation: a new era from an old idea?', *Ecology Law Quarterly*, **18**, 1–42.

Jaffe, A.B. and R.N. Stavins (1995), 'Dynamic incentives of environmental regulation: the effects of alternative policy instruments on technology diffusion', *Journal of Environmental Economics and Management*, **29**, 43–63.

Keohane, N.O., R.L. Revesz and R.N. Stavins (1997), 'The political economy of instrument choice in environmental policy', *Resources for the Future*, Discussion Paper 97–25, February.

Kolstad, C.D., T.S. Ulen and G.V. Johnson (1990), '*Ex post* liability for harm vs. *ex ante* safety regulation: substitutes or complements', *American Economic Review*, **80**, 888–901.

Landes, W. and R. Posner (1984), 'Tort law as a regulatory regime for catastrophic personal injuries', *Journal of Legal Studies*, **13**, 417–34.

Lewis, T. (1996), 'Protecting the environment when costs and benefits are privately known', *Rand Journal of Economics*, **27** (4), 819–47.

Pigou, A.C. (1932), *The Economics of Welfare*, 4th edn, London: McMillan & Co.

Porrini, D. (2001), 'Economic analysis of liability for environmental accidents', *Rivista Internazionale di Scienze Economiche e Commerciali*, no. 2, 189–218.

Porrini, D. (2002), 'Effetti economici della lender's liability e della financial responsibility per danni ambientali' [Economic Effects of Lender's Liability and Financial Responsibility in Case of Environmental Damages], *Rivista Italiana degli Economisti*, no. 1, April, 101–26.

Ringleb, A.H. and S.N. Wiggins (1990), 'Liability and large-scale, long-term hazards', *Journal of Political Economy*, **98**, June, 574–95.

Rose-Ackerman, S. (1991), 'Regulation and the law of torts', *American Economic Review, Papers and Proceedings*, **81**, 54–8.

Schmitz, P.W. (2000), 'On the joint use of liability and safety regulation', *International Review of Law and Economics*, **20** (3), 371–82.

Shavell, S. (1984a), 'Liability for harms versus regulation of safety', *Journal of Legal Studies*, **13**, 357–74.
Shavell, S. (1984b), 'A model of the optimal use of liability and safety regulation', *Rand Journal of Economics*, **15**, 271–80.
Shavell, S. (1986), 'The judgment proof problem', *International Review of Law and Economics*, **6**, June, 45–58.
Shavell, S. (1987), *Economic Analysis of Accident Law*, Cambridge, MA: Harvard University Press.
Stavins, R. (1998), 'Market-based environmental policies', Resources for the Future, Discussion Paper 98–26.
Summers, J.S. (1983), 'The case of disappearing defendant', *University of Pennsylvania Law Review*, **132**, 145–85.
Van't Veld, K., G. Rausser and L. Simon (1997), 'The judgment proof opportunity', Fondazione Eni Enrico Mattei Working Paper, no. 83.
Wiener, J.B. (1999), 'Global environmental regulation: instrument choice in legal context', *Yale Law Journal*, **108**, 677–800.

24 Tradable emission rights

Edwin Woerdman

Introduction

Traditional instruments for environmental regulation are emission standards and taxes. An emission standard, also referred to as 'command and control', defines the maximum amount of emissions of a certain pollutant or the maximum amount of emissions per unit of output or energy. An emission tax sets a tax rate on emissions to achieve a predefined emission level. Although both types of legal instruments to control environmental pollution are still being used in various – but certainly not all – areas of environmental policy, they have also met with increasing opposition from policy makers, managers and scholars, with economists at the fore. Taxes and energy efficiency standards, for instance, have been criticized for being not only too costly and rigid for firms, but also too ineffective for governments, because it is not clear how high the tax or energy efficiency rate must be set to achieve the national emission target. Moreover, high costs and low effectiveness endanger the compliance of governments with internationally agreed emission levels. To facilitate cost savings, increase flexibility and strengthen compliance, a relatively new instrument is beginning to find its way into environmental law: tradable emission rights.

Emissions trading, as it is also referred to, originates from the United States (US), but the instrument is now also implemented in other parts of the world such as the European Union (EU). The idea is, simply, that it is cheaper for one polluter to reduce its emissions as required by some environmental standard than for another, and that it is profitable for both if the former reduces its emissions below the standard and sells its surplus to the latter which is then allowed to emit more than the standard prescribes. The latter is allowed to emit more, but the former must emit less. This implies that the overall emission ceiling is met, with or without trading. Trading makes it cheaper, though, making compliance easier.

In this chapter, a law and economics approach will be taken to analyse tradable emission rights as a relatively new instrument in environmental law. The aim is, first, to analyse this legal instrument in terms of effectiveness and efficiency, and second, to analyse what role efficiency plays in a number of legal issues associated with it. To that end, different emissions trading design variants will be discussed against the background of the path-dependence approach, according to which a suboptimal design might eventually lock-in

institutionally rather than evolve into an optimal one. Moreover, the legal relevance will be demonstrated of the (often overlooked) economic concept of opportunity costs when allocating tradable emission rights, in particular with respect to state aid regulation and the polluter-pays principle in European law. The issues of harmonization and competitive distortion will be fully addressed.

Tradable emission rights: history, use and legal nature

Dales (1968) is usually seen as the founding father of the tradable emission rights concept, Montgomery (1972) as the one who provided formal proof of its efficiency, and Tietenberg (1980) as the one who firmly advocated and established it in environmental economics. Emissions trading can be traced back to the property rights school in economics, according to which externalities should be internalized (for example, Demsetz, 1967). This means that negative external costs such as environmental pollution, which are not reflected in the market price, should be included in this price by allocating property rights.[1]

In 1975 the US Environment Protection Agency (EPA) began experimenting with emissions trading to control air pollution. Since then the concept and variants thereof have been used in various other US programmes, for instance to reduce ozone-depleting substances under the Montreal Protocol (since 1988) and to reduce SO_2 emissions under the 1990 Clean Air Act Amendments (CAAA) (where such emissions have been traded since 1993). Outside the US, some experience was gained mainly with tradable quota systems, such as the tradable ammonia quota in the Netherlands (since 1994), but the definitive breakthrough of emissions trading outside the US is expected to occur in the context of climate policy in the course of the new millennium's first decade.

Several countries are in the process of planning or starting to implement trading schemes to reduce greenhouse gas (GHG) emissions. The Kyoto Protocol (Article 17) allows international emissions trading between 2008 and 2012. The annex on emissions trading in the Marrakech Accords allows governments to authorize legal entities to transfer and/or acquire emissions under Article 17. The Kyoto Protocol of 1997 entered into force on 16 February, 2005, after the Russian Federation had ratified it, bringing the number of ratifications over the legally required threshold. In 2003, when entry into force of the Kyoto Protocol was still uncertain, the governments of the European Community (EC), for instance, had already approved a directive that enables CO_2 emissions trading for power generators, steelmakers as well as cement, paper and glass manufacturers to start in 2005. More or less similar emissions trading schemes are already up and running in Denmark (since 2001) and the United Kingdom (UK) (since 2002). Also outside the EU, various countries such as Switzerland, Norway, Japan and Canada, intend to build national tradable emission rights

systems, which could eventually be linked to the EC scheme provided that they mutually recognize their transferable units. Moreover, the fact that the US withdrew from the Kyoto Protocol in 2001 did not prevent some North American companies from designing voluntary pilot emissions trading schemes of their own, such as the so-called Chicago Climate Exchange.

Most economists see tradable emission rights as property rights, because of their exclusive use, market value and incentive effects. In the trading scheme for SO_2 emissions in the US, however, a legal provision was adopted that an emission right, called an 'allowance', does not constitute a property right (in section 403(f) of the CAAA). The legislator chose this formulation so that the government would not have to compensate polluters for 'taking' allowances when the authorities lower the annual emission caps. Both in this scheme and in the European CO_2 emissions trading system, an emission right is basically defined, in legal terms, as an allowance that authorizes a legal entity to emit a certain amount of pollution during a specified period. This is not so much a permanent, private property right, but rather an authorization that can be terminated or limited by the government.

Although some then conclude that emission rights are, and should be, temporary 'rights of use' (for example, Convery et al., 2003), the law and economics literature prefers to characterize allowances as mixed, hybrid or regulatory property rights (for example, Rose, 1999; Yandle, 1999). Emission rights contain elements of both public and private property rights: instead of common law private rights and liability rules that form over time when conflicts over resource use arise, allowances are non-permanent, government-mandated rights that combine state control over the emission quotas with private freedom for polluters to choose how to comply (which could be referred to as 'command without control'). Moreover, although allowances in the American SO_2 emissions trading scheme are not property rights *themselves*, property rights *in* allowances are in fact recognized as emitters can receive, hold and transfer them, while excluding all others, besides the government, from interfering with their possession, use and disposition (Cole, 1999: 113–14).

Different types of tradable emission rights

On a domestic level, there are two general options to shape a tradable emission rights system, namely, on the basis of (a) permit trading or (b) credit trading. Under permit trading, a government allocates emission ceilings to private parties, allowing them to trade with each other. This is also referred to as private trading, firm-to-firm trading, allowance trading, inter-source trading or cap and trade. Under credit trading, however, one private party can sell credits to another by reducing its own emissions below a baseline, laid down in (energy-efficiency) environmental standards and possibly enforced by cov-

enant. Credit trading is sometimes also referred to as the unilateral approach to project-based emissions trading or, more recently, as performance standard rate trading. The distinction between these two basic types of legal instruments is a crucial one, because according to neoclassical economic theory, permit trading is the superior alternative (for example, Tietenberg et al., 1999: 106).

Permit trading, which incorporates emission ceilings, is efficient and effective. Newcomers and growing firms have to buy permits, also referred to as 'allowances', from other firms (or from a government reserve) to cover the additional polluting activities. Those who leave the industry keep their allowances, which they can sell. The system is efficient, because every emission allowance that is used to cover the emissions has a price: either the purchase price of new allowances or the revenues that the polluter forgoes by not selling the allowances it already possesses (which are opportunity costs as we shall explain in more detail later on). Each unit of emissions therefore has a price, since each unit could be sold. Moreover, if the economy grows, the demand for allowances increases, but the supply remains constant as a result of the emission ceiling. This means not just that the emission target will be achieved, but also that the scarcity of environmental space is reflected in a higher price for carbon-intensive products, thus encouraging technological innovation and an efficient restructuring of the economy in the direction of sustainable energy use.

Credit trading, which does not incorporate emission ceilings, is less efficient and its effectiveness is uncertain. A firm can create credits voluntarily by reducing its emissions below the emission level required by the applicable voluntary or regulatory policies and measures. For instance, if the policy is a performance (or: relative) standard which requires a certain quantity of CO_2 per unit of output or energy, a firm should multiply this standard with its production volume to obtain its total emission figure. If this firm emits less CO_2 than this baseline (or: benchmark) figure by initiating a certain abatement project, it can sell these credits to another firm.

Although companies can achieve cost savings by selling credits, the environmental scarcity under credit trading is not reflected in a price for each unit of emissions. If the economy grows, the supply of credits also increases because companies do not have an emission ceiling but have to observe an energy-efficiency standard. If an energy-intensive company wants to expand production, or if a newcomer enters the industry, it thus has a right to new emissions. These do not have to be purchased from existing polluters, or from a government reserve, within an environmental consumption space as in the permit trading system. Instead, the company receives its emission credits above and beyond the existing quantity. The consequence is that the social costs of the extra emissions are not fully reflected in the

costs per unit of product and thus not in the product price. Carbon-intensive products are therefore priced too cheaply, leading to an inefficient restructuring of production.

The market transaction costs of credit trading would not differ much from permit trading, because both types make use of the information advantages of the private sector and do not require advance approval of every entitlement transfer. Nevertheless, the determination of the allowed emissions for a given year is more difficult under credit trading, because these are not given (as under permit trading), but have to be calculated on the basis of existing climate policy, for example by multiplying the performance standard with the energy use in that year, which can be done accurately only *ex post*.[2]

The political economy of instrument choice

Despite the advantages to society of permit trading, seen from the viewpoint of neoclassical economics, certain politicians are inclined to opt for credit trading, either or not based on voluntary agreements. The usual political economy explanation for this is that credit trading has advantages for certain interest groups, such as the industry which does not have to purchase extra emission rights if companies seek to expand their production (for example, Dijkstra, 1999). However, there are also advantages of credit trading for politicians themselves (for example, Woerdman, 2004). Permit trading sets emission ceilings by (re-)distributing property or user rights, while credit trading uses existing environmental policy as a baseline for the calculation of the tradable emission reductions.

An institutional economics explanation for the political attraction of credit trading is that the start-up 'capital' or political transaction costs of permit trading are relatively high since it comes to replace existing environmental policy, while credit trading builds increasingly on extant policy (ineffective and inefficient as it may be). Another explanation for the political attraction of credit trading is that under permit trading a choice must be made between auctioning emission allowances or give them away free (for example, 'grandfathering' based on historical emissions). Under credit trading, emissions are always given away free, thus lowering the political visibility of the (re-)distribution issue.

Path dependence and the risk of institutional lock-in

In the emissions trading literature some contend that is it not problematic to start with credit trading, assuming that such a scheme can later be transformed into a permit trading system (for example, Tietenberg and Victor, 1994). However, this comes at a risk. The political choice of credit trading – a suboptimal type of emissions trading – can result in an institutional lock-in from which it may be difficult to escape in the future (Woerdman, 2004).

According to some versions of evolutionary theory, history moves in the direction of a superior alternative. According to the lock-in concept, however, history is path dependent and this evolution can come to a standstill, temporarily or not (for example, North, 1990; Arthur, 1994). An institutional lock-in then refers to the dominance of suboptimal regulation, such as a (set of) inefficient and/or ineffective legal instrument(s), in the presence of a superior alternative. Regulation is thought to be dominant when it is (formally adopted and) effectively implemented, while its alternative is not.

Essential components in the analysis are the distinction between the set-up costs of establishing a legal institution and the running costs of continuing it. Set-up costs can be subdivided into sunk costs (of the existing arrangement) and switching costs (of a new arrangement). The former are not relevant for the decision whether or not to continue and extend the existing arrangement because they were made in the past ('bygones are forever bygones', according to economic theory), but switching costs are relevant when establishing a new one because they still have to be made. Four factors can then be identified that contribute to a possible institutional lock-in of credit trading.

First, credit trading profits from the learning effects associated with building on existing environmental policy. Learning effects lower the average costs of running the established system. Moreover, it is possible that policy makers may expand the existing legal institution with credit trading because they are unacquainted, or are not sufficiently well acquainted, with permit trading.

Second, credit trading builds on the sunk costs of existing environmental policy. These start-up costs that have already been incurred play no role in the decision to continue current environmental policy without emission ceilings, whether or not modified to take account of credit trading. Although permit trading reduces running costs, it involves relatively high start-up costs because it implies crossing over to a new legal arrangement. Resistance by vested interests contributes to these switching costs.

Third, policy makers will be more persuaded to opt for credit trading if there is a predominant perception that the problem-solving capacity of the existing environmental laws is growing or stable. If the effort of policy makers is directed to 'satisficing' rather than 'optimizing', they are less receptive to theoretically superior alternatives such as permit trading.

Fourth and finally, credit trading can profit from network or coordination benefits by building on extant policy. If regulation is seen as the government's 'product', administrative costs are its 'production costs'. Contrary to what Pierson (2000) suggests, there is an incomplete (but not absent) analogy with increasing returns to scale. In short: the advantage for the government of building upon existing legal arrangements does not originate from producing larger quantities of (similar) rules, but what matters is that the differential

administrative costs (the extra costs of adding another collection of units) decline as the institutional scale increases. This can be done by expanding an existing environmental instrument (horizontally) to cover extra target groups, such as more segments of industry, or the government itself can expand the instrument (vertically) by adding an element such as credit trading. These processes are driven by increasing returns to scope, rather than increasing returns to scale.

The preconditions for an institutional break-out are the opposite of those set out above for an institutional lock-in. This puts us in a position to analyse the path-dependent evolution of environmental law. The chances for permit trading then improve, for instance, as the information on this instrument ameliorates, as the costs of switching to a permit trading system decline, and as the problem-solving capacity of existing environmental regulation deteriorates.

Case study: did the Netherlands break out?

Environmental regulation in the Netherlands is an interesting case in point (for example, Woerdman et al., 2002). Various ministries, as well as the Vogtländer Committee, which was investigating emissions trading, initially pleaded in favour of credit trading by building upon existing standards for those sectors of industry that were energy-intensive and competing internationally. The four factors mentioned above explain their position. However, a number of scientists (in particular economists) and ultimately also the Social-Economic Council, in its capacity as adviser to the Dutch government, in fact pleaded in favour of the non-incremental option of permit trading. Thanks to decreasing political transaction costs of permit trading, the balance now seems to tilt in favour of the latter.

As the information on permit trading improved, the problem-solving capacity of extant environmental policy deteriorated and cultural barriers concerning 'pollution rights' crumbled, there was an increasing preparedness to tackle the legal impediments. The largely successful experiments with permit trading in Denmark and the UK, coupled with the will of the Europeans to display effective leadership in international climate policy after the Americans had left the Kyoto Protocol in 2001, also contributed to a lowering of the switching costs of permit trading. Moreover, the European Commission adopted a pioneering role, exerting internal pressure by preparing a directive that would enable permit trading in the EU from 2005 for installations of large industrial sectors. On an overarching European level, as against individual member state level, there was hardly any existing climate policy, let alone well-established legal instruments, to build upon. In 2003, after some amendments, the European Parliament agreed to this directive in a co-decision procedure and member states now have to put it into effect. This also

means that Dutch climate policy, after years of uncertainty, is en route to an institutional break-out.

Nevertheless, some companies and policy makers still try to steer the national allocation of emission rights in the direction of credit trading, by linking the height of ceilings for individual companies, within the ceiling for an industry as a whole, to the size of their production. This sort of linkage, which is advocated on fairness grounds by many (energy-intensive) companies, and even by some scientists (for example, Groenenberg and Blok, 2002), is not fully efficient. On the level of the individual firm, it is then signalled that production growth implies free emission space. Economists know, however, that there is no such thing as a 'free lunch'. The price of the extra emission space should make clear that an expansion of carbon-intensive production can lead to destroying economic value, because it would necessitate relatively expensive, additional emission reduction measures elsewhere in the economy.

A bleaker picture must be painted for acidification policy in the Netherlands. Here, no international developments in the direction of permit trading are present and a credit trading system for NO_x emissions is being prepared to make the system of strict emission standards for large combustion plants more flexible. As of yet, there are no plans to make a switch to permit trading here. The path-dependent history is illustrative. The government and the chemical industry could not agree upon a NO_x emissions covenant, mainly due to scale differences between the companies in this heterogeneous sector, after which they decided to make current standards less rigid by means of credit trading. It is not likely that this 'moving train' will now suddenly change track to follow the destination of the more efficient and effective alternative of cap and trade, not least because of the vested interests that push hard to avoid emission ceilings. This underlines that the Netherlands has not yet escaped from an institutional lock-in in all corners of environmental regulation.

Competitive distortions and state aid

Contrary to the general design of the EU directive on CO_2 emissions trading by choosing in favour of permit trading, the specific harmonization of the allocation of emission rights, as foreseen under this directive, is not compatible with the neoclassical economic textbook model. The European Commission feared, however, that competitive distortions and state aid could arise under EC Article 87 when member states would be left free to choose their allocation method (COM, 2000a). This could imply that a company in one member state would have to buy its emission allowances at an auction, whereas its competitor in another member state would get them for free, for instance by means of 'grandfathering' based on historical emission figures. Therefore, Article 10 of the directive requires that every member state allo-

cates at least 95 per cent of its allowances free of charge during the 2005–07 period and that at least 90 per cent of the allowances is allocated for free in the 2008–12 period. Such differences regarding the method of allocation are impossible under credit trading, because in that system emissions are always allocated for free. There is *de facto* no choice between allocation free of charge and auctioning under credit trading.

The concept of 'competitive distortion' (and thus not something like fair competition or equal treatment) is at the centre of the legal definition of state aid under EC Article 87. Legal research has shown that this concept has been used in European environmental law, sometimes as a distortion of efficiency and sometimes as a distortion of fair competition (Van der Laan and Nentjes, 2001). Many authors are inclined to think that the allocation of emission rights across member states must be harmonized to avoid competitive distortions (for example, Lefevere and Yamin, 1999; Cozijnsen, 2002). This depends, however, on what law and economics perspective decision makers and lawyers take (Woerdman, 2004).

Everyone understands that a company has to spend money for its emission allowances at an auction and that it saves these costs when these rights are allocated free of charge. Some lawyers then refer to a gratis versus a 'financial' distribution of emission rights and refer to the first option as less market oriented (for example, Peeters, 2002: 191). But many authors do not seem to realize that allowances that have been allocated free of charge also involve costs for firms. According to a neoclassical economic approach, allowances handed out free of charge have an 'opportunity cost' when they are used for covering the emissions of the allowance owner (for example, Nentjes et al., 1995; Grafton and Devlin, 1996). This alternative cost, which is equal to the price for which the allowances could have been sold, must be included in the product price if the firm does not want to go bankrupt in the longer term. The opportunity cost is the revenue forgone by not selling the allowances, but employing them in producing output: instead of using the allowances, the firm could have sold them.[3]

On the basis of this aspect, which has been neglected by various authors (for example, Peeters, 2003: 90), there is no competitive distortion in the aforementioned example in which one firm has to buy its emission rights, while its competitor gets them for free. The latter company does not have a cost advantage *vis-à-vis* the former (*ceteris paribus*). Because the firm with gratis emission rights has to pass on the alternative cost, it cannot ask lower product prices (in a well-functioning market) than a firm with auctioned emission rights.[4] This means that harmonization on the basis of gratis allocation is not necessary according to this efficiency-oriented approach.

The decision to harmonize is therefore not grounded in efficiency, but in equity concerns. A neo-institutional economic approach learns that allocation

free of charge distorts fair (not efficient) competition if the emission rights are auctioned in another country, because companies with free permits are financially favoured above their competitors which had to buy their emission rights. A firm with gratis allowances therefore has more financial resources than a comparable firm abroad with auctioned allowances, because the former type of allocation implies a capital gift. The allocation of emission rights itself then leads to unequal changes in the financial positions of and thus the competitive relations (the 'level playing field') between comparable firms.

According to EC Article 87(1), we speak of state aid, for example if the aid, granted by a member state or through state resources in any form whatsoever, distorts or threatens to distort competition by favouring certain undertakings or the production of certain goods, in so far as it affects trade between member states. The European Commission uses additional criteria to determine whether or not a measure is to be regarded as state aid, like the 'state origin' criterion which is also satisfied if the government forgoes revenue that would otherwise have been collected (COM, 2000b).

Allocating emission rights free of charge is not state aid in the efficiency approach of competitive distortions with a view to their opportunity costs, but it is state aid according to the equity approach because of the financial advantage that gratis allocation implies and because the government would have received income (either to be recycled in the economy or not) in the alternative option of auctioning. Free allocation, like grandfathering, in fact means that the (hypothetical) revenues from an auction are given to the polluters (for example, Welch, 1983: 168). But even then there are legal possibilities to allow an allocation free of charge on the basis of EC Article 87(3). The European Council, for instance, can 'simply' decide that the state aid under consideration is compatible with the common market. According to a law and economics approach, such legal ambiguities and complexities add to the perceived set-up costs (by decision makers) of permit trading.

Interestingly, in its decisions of April 2000 and November 2001, the European Commission indeed considered grandfathering to be state aid in the experimental permit trading schemes of Denmark and the UK by using the state origin criterion: the state forgoes revenue which could derive from auctioning the valuable permits (COM, 2000c; COM, 2001). Nevertheless, the Commission exempted the aid in both cases by using environmental, economic, legal as well as political arguments. The grandfathering was allowed as state aid, among other things, by following EC Article 87(3)(c) on developing certain activities or areas and surprisingly enough not by pointing at aspects of the allocation *per se*, but by highlighting the more general characteristics of the emissions trading scheme itself, like the positive contribution it makes to environmental protection, and by stating a political desire to gain experience with and prepare for emissions trading.

Opportunity costs and the polluter-pays principle

By choosing in favour of harmonization under the directive on emissions trading, the EU has acknowledged, as explained above, that it incorporates not only efficiency, but also equity aspects of the allocation of emission allowances into its legal decision regarding EC Article 87. It would be inconsistent if this equity aspect suddenly ceases to play a role in other legal matters related to the same allocation of emission allowances. An example is the question whether allocation of emissions free of charge, as foreseen in the European directive, is compatible with the polluter-pays principle established under EC Article 174. The answer to this question is that there is compatibility between gratis allocation and the polluter-pays principle if emphasis is placed on the efficiency aspect of the distribution of emission rights: a consistent application of the equity approach, however, leads to a potential legal conflict between free distribution, like grandfathering, and EC Article 174.

The existing legal literature has hardly considered this issue. And those authors who have considered the issue basically believe that a polluter does not pay for his/her pollution if emission rights are obtained for free, contrary to a polluter that has to buy the required emission rights at an auction (for example, Nash, 2000). In his most recent book, however, Woerdman (2004) makes a distinction between two possible approaches. According to the efficiency approach, a polluter also pays when he/she has received his/her allowances free of charge, namely in the form of the opportunity cost of using the allowances (next to his/her, direct, emission reduction costs). Instead of using the allowances, he/she could have sold them. According to the equity approach, in the case of gratis allocation, it is not the polluter who pays, but the public, because allocating emission rights free of charge can be interpreted as a wealth transfer from the public (or government) to the polluters. In the form of emission rights that have economic value, the polluters, one could argue, receive the revenues that the government would have obtained in the case of an auction.

The interesting thing now is that the member states of the EU have opted, as we have seen, for the harmonization of the allocation of emission allowances on a free-of-charge basis so as to avoid competitive distortions. This harmonization, however, is not necessary on efficiency grounds: not only companies with auctioned allowances, but also companies with allowances allocated for free make (opportunity) costs. The fact that a firm got its emission rights free of charge does not mean that it can ask lower product prices than a comparable firm that had to buy such rights. The former has no competitive advantage over the latter: allowances distributed free of charge are a so-called 'lump-sum subsidy' that does not distort efficiency. A lump-sum subsidy does not distort efficiency in the product market, since it does not affect marginal emission reduction costs and it does not alter the output

and price decisions of firms. Nevertheless, by choosing to harmonize, the EU apparently has interpreted the gratis allocation of emission rights not in terms of opportunity costs, but in terms of a financial advantage or welfare transfer. This advantage plays a role, as we have noted before, in the equity approach.

A consistent application of this interpretation then means, however, that allocating emission rights free of charge is not compatible with the polluter-pays principle under EC Article 174. Only in terms of opportunity costs can this allocation method be seen as compatible with this principle. If the EU were to suddenly stress the opportunity cost of gratis emission allowances in the case of the polluter-pays principle, while it disregarded this characteristic when considering the case of harmonization, this would be an inconsistent view on the allocation aspect of emissions trading.

This complex issue could remain a legal-theoretic debate behind the scenes, but it could, in principle, also grow into a political-legal conflict that would slow down (or even block) the implementation of permit trading in the member states of the EU. It seems, however, that the political will of the European institutions to implement emissions trading based on an allocation of emission rights free of charge carries more weight than the consistency of the legal and economic argumentations that are at the root of this system.

In addition, if there is an incompatibility, it still does not mean that the emissions trading system could not function. It is not the first time that legal–economic inconsistencies and imperfections have been observed, for instance in political and judicial decisions regarding European competition law and individual cases of state aid in general (for example, Hildebrand, 1998: 413; Cini and McGowan, 1998: 143). Furthermore, in the emissions trading directive, EU policy makers have chosen an, in principle, economically and legally solid design of trading pollution under an emission ceiling in the form of permit trading.

This does mean, however, that the decision making on this directive has been in two minds about the legal consequences of permit allocation, namely an equity view on the harmonization of the allocation of emission rights regarding EC Article 87 and an efficiency view on the polluter-pays principle applied to the same allocation of emission rights regarding EC Article 174. This is inconsistent from a law and economics point of view, but more pragmatic lawyers are likely to perceive it as a handy and politically acceptable way of getting the emissions trading system up and running.

Moreover, to some extent, one could argue that at least the European Parliament has been consistent, namely in using the equity view (although they could be accused, in principle, of neglecting the efficiency view). Members of the parliament, who initially pleaded for more auctioning, said that allocating first 90 per cent and then 95 per cent of the allowances free of charge would ensure the ‘progressive’ application of the polluter-pays principle

and cause 'less' distortion of competition, insisting that further harmonization should be considered, including auctioning, for the time after 2012 (Worsley and Freedman, 2003: 15). On the one hand, it seems that a harmonized scheme that prescribes 100 per cent auctioning would ensure the 'full' application of the polluter-pays principle for them, which is consistent with the equity view. On the other hand, it illustrates that in politics an allowance allocation method can be a 'bit' (or 'more' or 'less') consistent with the polluter-pays principle, whereas in a conceptual approach a particular allocation method is either consistent with this principle or not.

On such a theoretical level, the analysis in this chapter also shows that the approach of various lawyers is incomplete if not incorrect. According to Nash (2000: 3, 13), for instance, the weak form of the polluter-pays principle is a rule against subsidizing pollution and the strong form is a requirement to internalize the costs of pollution. Nash argues that distributing allowances free of charge as under grandfathering is inconsistent with the weak form of this normative principle, because allowances are allocated at no cost which implies a subsidy. Moreover, because inconsistency is 'demonstrated' with the weak form, Nash does not check compatibility with its strong version. These arguments would appear to be inaccurate. First, although grandfathered permits are indeed a subsidy, which plays a role in the equity approach, the point is that they are a lump-sum subsidy which does not affect efficiency. The question of consistency with the weak form of the polluter-pays principle thus depends on the (equity or efficiency) approach taken. Second, grandfathered permits do internalize the cost of pollution, as demonstrated above, because of their opportunity costs. The question of consistency with the strong form thus depends on whether one recognizes and acknowledges the aspect of opportunity costs when allocating emission rights free of charge. This indicates that a law and economics approach is, at least, capable of supplementing the existing legal literature on this issue.

Conclusion

In economic terms, tradable emission rights originate from the property rights school and attach a price to environmental pollution. In legal terms, an emission right is not a permanent, private property right, but rather an authorization, or hybrid property right, which can be terminated or limited by the government. A tradable emission rights system can be designed on the basis of permit trading or credit trading. Permit trading is superior according to neoclassical economic theory. Credit trading, also referred to as performance standard rate trading, is inefficient and its effectiveness is uncertain. The environmental scarcity is not reflected in a price for each unit of emissions: when the economy grows, the supply of credits increases as well, because polluters do not have an emission ceiling. Under permit trading, also called

allowance trading or cap and trade, polluters do have an emission ceiling. This design option is both efficient and effective: when the economy grows, the demand for emission rights increases, but the supply of such rights remains constant because of the emission ceiling. The US gained experience with both design options: their SO_2 allowance trading scheme is well known. After years of scepticism, the EU agreed upon a directive that allows CO_2 permit trading, to start in 2005, for various segments of industry.

However, some companies and policy makers in the EU still try to steer the national allocation of emission rights in the inefficient direction of credit trading, for instance by linking the height of ceilings for individual companies, within the ceiling for an industry as a whole, to the size of their production. An institutional economics perspective can explain such attempts. Credit trading is politically attractive, not only for firms, which do not have to purchase new emission rights when they expand, but also for policy makers: the set-up costs (or political transaction costs) of permit trading are relatively high, since it comes to replace existing environmental policy, while credit trading builds incrementally on extant policy by using it as a baseline to calculate the transferable emission reductions. In addition, some have argued that is it not problematic to start with (elements of) credit trading, assuming that such a scheme can later evolve into a permit trading system. A path-dependence approach, however, suggests that this comes at a risk, because starting with a suboptimal type of emissions trading can result in an institutional lock-in from which it may be difficult to escape in the future. Sunk costs, coordination benefits, learning effects and vested interests are some of the primary self-reinforcing mechanisms that contribute to such a lock-in.

Under the aforementioned EU directive, every member state is required to allocate its allowances free of charge to avoid competitive distortions and state aid under EC Article 87. The desirability of such harmonization depends, however, on what law and economics perspective one takes. Often overlooked is that not only the alternative of auctioning, but also allocation free of charge involves costs for firms: emission rights allocated for free have opportunity costs when they are used to cover the emissions of the allowance owner. A firm with gratis emission rights has to include these costs (equal to the allowance price) in the product price if it does not want to go bankrupt in the longer term. This means that it cannot ask lower product prices than its competitor with auctioned rights. According to this efficiency approach, therefore, allowance allocation differences do not lead to competitive distortions or state aid, making harmonization unnecessary. By choosing in favour of harmonization in the directive, however, the EU stresses the legal relevance of the equity approach concerning EC Article 87: a company with gratis allowances has more financial resources than a comparable foreign firm with auctioned allowances, because allocation free of charge implies a capital gift.

It would be inconsistent if this equity aspect were suddenly to cease to play a role in other legal matters related to the same allocation of emission allowances. An example is the question whether gratis allocation is compatible with the polluter-pays principle established under EC Article 174. According to the efficiency approach, a polluter also pays when allowances are allocated for free, because he/she does have costs, namely the opportunity costs of employing (and thus not selling) the emission rights in producing output. However, according to the equity approach (which does not focus on the efficiency similarities between auctioning and free allocation, but on their financial differences), it is not the polluter who pays, but the public, because gratis allocation can be interpreted as a welfare transfer from the public to the polluters. A consistent application of the equity approach therefore leads to a potential legal conflict with the polluter-pays principle in the EU. Politics, however, is not always consistent from a law and economics point of view: it seems that the political will of the European institutions to allocate emission rights for free carries more weight than the conceptual consistency of the emissions trading system.

Notes

1. Some authors argue that more or less similar ideas can be traced back to as far as John Stuart Mill, who wrote in 1848 about the possibility of giving air a market price, or Aristotle, who wrote more than 2000 years ago that which is common to the greatest number has the least care bestowed upon it (see Yandle, 1999: 17; Cole, 1999: 105).
2. Project-based credit trading, like joint implementation (JI) and the clean development mechanism (CDM) as defined under Articles 6 and 12 of the Kyoto Protocol, respectively, is a different story. In that case, an investor receives credits for achieved emission reductions in a (usually foreign) host country. These emission reductions are measured from a baseline that estimates future emissions at the project location if the project had not taken place. These baselines have to be approved before the transaction is allowed, which increases market transaction costs.
3. To clarify the concept of opportunity (or alternative) costs it may be useful to compare (a) the difference between a firm with gratis allowances and a firm with auctioned allowances to (b) a young farmer who has inherited land from his father and a young farmer who has no family in the agrarian sector and thus has to buy land. The first farmer must pass on the opportunity cost of the land in the price of, say, the milk or corn he produces if he does not want to end up bankrupt: he could have sold the land instead of using it. He does have a financial advantage over his young competitor who had to buy the land from someone else, but as a result of the opportunity cost of the land he cannot ask lower prices for his products than his competitor, so that competition is not distorted.
4. In an imperfect market, competition is only distorted in a number of exceptional cases. An example is a price war in which the company with gratis emission rights can survive longer, financially, than a company with auctioned rights. A price war is risky, however, not only from an economic, but also from a legal perspective: the price fighter can be prosecuted by the authorities that enforce antitrust law.

References

Arthur, W.B. (1994), *Increasing Returns and Path Dependence in the Economy*, Ann Arbor, MI: University of Michigan Press.

Cini, M. and L. McGowan (1998), *Competition Policy in the European Union*, London: Macmillan.
Cole, D.H. (1999), 'Clearing the air: four propositions about property rights and environmental protection', *Duke Environmental Law and Policy Forum* **10** (103), 103–30.
COM (2000a), *Green Paper on Greenhouse Gas Emissions Trading Within the European Union*, Green Paper presented by the Commission, 8 March, Brussels: European Commission.
COM (2000b), *XXIXth Report on Competition Policy 1999*, document SEC(2000)720 final, 5 May, Brussels/Luxembourg: European Commission.
COM (2000c), *State Aid No. N 653/99 – CO_2 Quotas (Statsstøttesag nr. N 653/99 – CO_2 kvoter)*, Letter by Mario Monti to the Danish government, English version (draft), 12 April, Brussels/Luxembourg: European Commission.
COM (2001), *State Aid No. N 416/2001 – United Kingdom Emission Trading Scheme*, Letter by Mario Monti to the UK government, C(2001)3739fin, 28 April, Brussels/Luxembourg: European Commission.
Convery, F.J., L. Redmond, Louise Dunne and L.B. Ryan (2003), 'Assessing the European Union Emissions Trading Directive', Paper presented at the 12th Annual Conference of the European Association of Environmental and Resource Economists (EAERE), 28–30 June, Bilbao, Spain.
Cozijnsen, J. (2002), 'Internationale handel in emissierechten: handhavingsinstrument ten behoeve van het Kyoto Protocol en de Europese en nationale uitwerking' [International trade in emission rights: enforcement instruments under the Kyoto Protocol and the European and national implentation], *Milieu & Recht* 7/8, 197–201.
Dales, J.H. (1968), *Pollution, Property and Prices: An Essay in Policy-Making and Economics*, Toronto: Toronto University Press.
Demsetz, H. (1967), 'Toward a theory of property rights', *American Economic Review* **57**, 347–59.
Dijkstra, B.R. (1999), *The Political Economy of Environmental Policy: A Public Choice Approach to Market Instruments*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Grafton, R.Q. and R.A. Devlin (1996), 'Paying for pollution: permits and charges', *Scandinavian Journal of Economics* **98** (2), 275–88.
Groenenberg, H. and K. Blok (2002), 'Benchmark-based emission allocation in a cap-and-trade system', *Climate Policy* **2**, 105–09.
Hildebrand, D. (1998), *The Role of Economic Analysis in the EC Competition Rules*, The Hague: Kluwer Law International.
Lefevere, J. and F. Yamin (1999), *EC Trade and Competition Law Issues Raised by the Design of an EC Emissions Trading System*, Scoping Paper No.1, Study for the European Commission DG XI, Designing Options for Implementing an Emissions Trading Regime for Greenhouse Gases in the EC, London: Foundation for International Environmental Law and Development (FIELD).
Montgomery, W.D. (1972), 'Markets in licences and efficient pollution control programs', *Journal of Economic Theory* **5** (3), 395–418.
Nash, J.R. (2000), 'Too much market? Conflict between tradable pollution allowances and the "polluter pays" principle', *Harvard Environmental Law Review* **465**, 1–59.
Nentjes, A., P. Koutstaal and G. Klaassen (1995), *Tradeable Carbon Permits: Feasibility, Experiences, Bottlenecks*, Dutch National Research Programme on Global Air Pollution and Climate Change (NRP), NRP Report no. 410 100 114, Groningen/Bilthoven: RuG/NRP.
North, D.C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
Peeters, M. (2002), 'Een EU-markt voor broeikasgasemissierechten' [An EU market for greenhouse gas emission rights], *Milieu & Recht* 7/8, 189–96.
Peeters, M. (2003), 'Emissions trading as a new dimension to European environmental law: the political agreement of the European Council on greenhouse gas allowance trading', *European Environmental Law Review* (March), 82–92.
Pierson, P. (2000), 'Increasing returns, path dependence, and the study of politics', *American Political Science Review* **94** (2), 251–67.

Rose, C.M. (1999), 'Expanding the choices for the global commons: comparing newfangled tradable allowance schemes to old-fashioned common property regimes', *Duke Environmental Law and Policy Forum* **10** (45), 45–72.
Tietenberg, T. (1980), 'Transferable discharge permits and the control of stationary source air pollution: a survey and synthesis', *Land Economics* **56** (4), 391–416.
Tietenberg, T., M. Grubb, A. Michaelowa, B. Swift and Z.X. Zhang (1999), *International Rules for Greenhouse Gas Emissions Trading: Defining the Principles, Modalities, Rules and Guidelines for Verification, Reporting and Accountability*, UNCTAD/GDS/GFSB/Misc.6, Geneva: United Nations Conference on Trade and Development (UNCTAD).
Tietenberg, T. and D.G. Victor (1994), 'Possible administrative structures and procedures for implementing a tradeable entitlement approach to controlling global warming', in *Combating Global Warming: Possible Rules, Regulations and Administrative Arrangements for a Global Market in CO_2 Emission Entitlements*, Geneva: United Nations Conference on Trade and Development (UNCTAD).
Van der Laan, R. and A. Nentjes (2001), 'Competitive distortions in EU environmental legislation: inefficiency versus inequity', *European Journal of Law and Economics* **11** (2), 131–52.
Welch, W.P. (1983), 'The political feasibility of full ownership property rights: the cases of pollution and fisheries', *Policy Sciences* **16**, 165–80.
Woerdman, E. (2004), *The Institutional Economics of Market-based Climate Policy*, Amsterdam: Elsevier.
Woerdman, E., J.T. Boom and A. Nentjes (2002), 'Economy versus environment? design alternatives for emissions trading from a lock-in perspective', in M.T.J. Kok, W.J.V. Vermeulen, A.P.C. Faaij and D. de Jager (eds), *Global Warming and Social Innovation: The Challenge of a Climate-Neutral Society*, London: Earthscan, pp. 160–78.
Worsley, R. and R. Freedman (2003), *Europarl Daily Notebook: 02-07-2003*, Brussels: European Parliament.
Yandle, B. (1999), 'Grasping for the heavens: 3-D property rights and the global commons', *Duke Environmental Law and Policy Forum* **10** (13), 13–44.

25 Regulatory taxation

Jürgen G. Backhaus[1]

Introduction

This entry serves to emphasize the difference between three distinct types of taxation: classical or traditional taxes serve to generate revenues with a minimum excess burden or welfare loss; regulatory taxes, on the other hand, can be of two kinds. Pigouvian taxes are regulatory taxes in the sense of accomplishing slight changes in individual or firm behaviour; the production or consumption functions are assumed as given. The point of environmental taxation and of designing an environmentally sound tax system, on the other hand, is to accomplish deep and structural changes in the economic and ecological behaviour of individuals, households and firms, that is, changes of patterns and not changes of degree. When these basic patterns of behaviour and processes are to be influenced, it is important to identify those basic aspects that are environmentally and ecologically important and susceptible to change as a consequence of a tax intervention. Consequently, one has to identify those instances where basic choices can be taken and where decision makers face alternatives among which they can choose.

Generally speaking, environmental taxes of this kind belong to a type of regulatory taxes that is designed to affect not only choices within structures and constraints, but also choices *about* such structures and constraints. Since Pigouvian taxes, the first type of regulatory taxes, are well documented in the literature, the emphasis here is on the other type of regulatory taxes which aim at structural change. Throughout, this is done with reference to environmental regulatory taxes.

This approach is different from the standard approach to discussing environmental taxation. The law and economics approach pursued below looks at structural changes and the possibility of effecting such changes through tax instruments. The standard approach to ecological taxation is to impose regulatory taxations so as to curtail environmentally and ecologically undesirable effects and thereby even generate a double dividend consisting of, on the one hand, the improvement of the environment and the ecological system and, on the other hand, a revenue which can be used for different policy ends, such as those relating to the environment and the ecological system. Taxes which promise to generate such a double dividend are, of course, politically very attractive. However, it has repeatedly been pointed out that the double divi-

dend approach is basically flawed (Bovenberg and de Mooij, 1994; Backhaus, 1995; Schneider and Volkert, 1997).

With so much ambiguity surrounding environmental taxation, a fresh approach may well be tried. This entry explores a Coasean perspective in order then to proceed to an analysis based on a rigorous application of Buchanan's 'cost and choice' concept. A third section contains a simple illustration. The entry concludes with a summary and an emphasis on unresolved issues.

A Coasean view

Ronald Coase has consistently argued with respect to the special case of externalities that such externalities merely reflect the existence of incompatible uses of the same resources by different agents. In the presence of well-established property rights, and in the absence of significant transactions costs, externalities cannot persist, since the right of use will be appropriated by that actor with the highest willingness to pay; or, put another way, the person who experiences the highest cost from an externality will seek to contain those costs either by establishing his/her property rights over the entire range of resource uses or else by trying to recover as large a compensation as possible from the sale of his/her rights of usage.

By way of example, consider a coal-based electricity generation plant located in a valley close to a river from which it can take cold water for its cooling tower, re-emitting the water partly into the river and partly into the air. Assume further that, beyond the warming effect, there is no further contamination of the water which is being reintroduced into the local ecological system. In addition, electricity generation, by burning coal, leads to potentially toxic fumes which can be reduced to safe levels by building large smokestacks and thereby distributing the substance over a large area. The local economy in the ecological system is further assumed to consist of agriculture-based production: for the sake of simplicity we assume wine growing on the slopes of the mountains surrounding the river and fishery in the river.

From a physical point of view, there is emission of fumes into the air, which can be spread over a large area so as to avoid safety problems: there will be some warming of the water and a certain amount of (clean) steam being released into the air. A combination of a warmer river and steam in the air are likely to lead to the formation of mist which inhibits exposure of the grapes to the sun. From the point of view of the fishery industry, the warming of the river will affect the existing supply of fish. Finally, if additional sources of toxic smoke are located outside the ecological system in question, with similar techniques of dispersion, the operation of the carbon-based power plant within the system may well lead to unsafe levels of toxic exposure, meaning that halting the operation of the plant would reduce the levels of toxic exposure again to safe levels.

On the basis of these assumptions, let us look at the externalities in question from a Coasean point of view. The intra-system externalities crucially depend on the number of different uses we allow to be made of the ecosystem. We have incompatibilities between the power plant and the wine-growing industry, between the power plant and the fishing industry, but no incompatibility between wine growing and fishing – small wonder, since they have probably coexisted for a long period of time with well-established property rights governing the use of the river and the surrounding mountain slopes. The negative effects from warming of the river can probably be contained on the part of the fishing industry if there is certainty about the temperature of the water and fish used to higher temperatures can be bred and harvested in that stretch of the river. That would imply, however, that the inflow of colder water will now have to be regarded as a negative externality from the point of view of using the river for warmer water fish. Hence any arrangement between the fishing industry and the power plant would have to involve guarantees and compensation schemes with respect to the temperature of the river. Both the right to breed and harvest fish and the right to emit warm water would have to be curtailed so as to make the two uses of the river compatible. Fishermen would not be allowed to breed and harvest those species of fish that are particularly vulnerable to large fluctuations in water temperature, and in particular vulnerable to excessively low levels of water temperature, because that is most difficult for the power plant to control during the spring season when the ice is melting in the mountains. With respect to these two parties, then, a partitioning of the traditional rights of uses would allow trade and contractual arrangements governing this trade.

With respect to the relationship between the power plant and the wine-growing industry, the formation of mist is crucial, which implies that, under those weather conditions when mist has a crucial influence (in particular during the early autumn when sun exposure is necessary for the grapes to ripen and mist can lead to contamination by insects and diseases) the two uses of wine growing and cooling with river water have to be reconciled. Hence a contractual arrangement needs to be designed that governs the ejection of steam during these sensitive periods: alternative forms of cooling or additional forms of recooling the steam before re-emitting it into the river or ejecting it into the air may have to be considered, the timing of ejection or re-emission may also be crucial during the day, as, for instance, the formation of late afternoon mist on sunny days is unlikely. Again a partitioning of property rights with respect to the water use, including the possibility of mist formation and with particular emphasis on that possibility during crucial periods of growth and harvest, will in all likelihood allow for the internalization of the externality through an effective definition of property rights.

The analysis has, so far, been substantially facilitated by limiting the number of alternative and partly incompatible uses of the common resource (the valley with the river) for the three different purposes of energy generation, wine growing and fishing. In addition, the plant emits potentially toxic fumes, which, however, become toxic only if additional sources of emission exist outside the system. Obviously, here a different ecological regime needs to be defined for that particular resource use, which is independent from the other resource uses. It is here that the principle of subsidiarity has to be invoked (see Backhaus, 1995). In order to draw conclusions from the Coasean analysis, it is interesting to note which decisions need to be taken in order to allow compatible resource use by means of an optimal partitioning of property rights. With respect to the power plant, it is important to recognize the alternative uses in choosing appropriate technologies and modes of operation. On the one hand, with respect to agriculture (wine growing), plant operation needs to take into account weather conditions during particular periods of the year with respect to providing for a containment of the emission of the water as a cooling substance into both air and river. This may require a different definition of peak times with respect to production, alternative uses of cooling equipment, larger cooling systems so as to have more standby cooling power and close monitoring of weather conditions. These decisions, in all likelihood, can be taken cost effectively at the time of plant construction and design. Later, it will likely be more expensive to make alterations or to take these conditions into account as constraints on operation. This aspect of the scenario is consistent with the assumptions in the global warming debate, which also contain the notion that an early intervention will lead to lower total costs than a later one.

With respect to the relationship between the power plant and the fishermen, again the management of the ejection and re-emission of the cooling substance is crucial. Interestingly, here at times a larger inflow of warm water into the river may be required in order to protect the fishing stock from a sudden drop in water temperature. As obviously the pattern of breeding and harvesting fish can largely contain the cost of sudden temperature drops, and during the winter different fish can populate the river from those during the summer, the requirement may only involve relatively small peak periods and, in this case, require the re-emission of uncooled water into the river. Again a large water reservoir and the ability to use and manage different forms of cooling will be required for this approach to be successful. Very similar to the first case, the cost of operation will again largely be affected by the timing of the decisions, and, when the decisions can be taken at the time of design and construction of the plant, the costs will in all likelihood be much lower than if later alterations of plant design, process of production or management need to be implemented.

Basing environmental taxation on opportunity costs

'The tax which each individual is bound to pay ought to be certain, and not arbitrary. The time of payment, the manner of payment, the quantity to be paid, ought all to be clear and plain to the contributor, and to every other person' (Smith, 1971, bk 5, ch. 2). When we deal with effective environmental taxation, every single aspect of this principle, which is part of the traditional canons of taxation, stands to be violated. To see this, we have to invoke the difference between objective costs and subjective or choice influencing costs that has been effectively introduced into public finance analysis by James M. Buchanan (1969). Pigou's theory of regulatory taxation, which underlies the current literature on environmental taxation, is based on the notion that the external costs of a particular activity can be internalized into the decision calculus by means of imposing a tax, thereby reducing the externality.[2] This requires, however, that the taxing authority be aware of the costs. This in turn requires the notion that costs can be objectively measured. Costs that can be objectively measured, on the other hand, need not be those that influence decisions. As we have seen in the introductory survey, carbon taxes are supposed to meet an inelastic demand for carbon fuel, only in the long run leading to the substitution of more energy-efficient carbon fuels for less energy-efficient ones. Speaking of an inelastic demand subject to taxation means that the tax does not influence the decisions of the taxpayer to any large degree.

In order to appreciate the breadth of this statement, consider the example used before in the following way. After the electric power plant has been installed and power supplies have been improved for the region, the wine growers suddenly and unexpectedly experience heavy losses due to the formation of mist in the valley. They intervene with the government, which proceeds to impose a 'power penny', that is, a certain percentage charge on each currency unit billed. The '*x*' percentage charge, called a 'power penny' in order to emphasize its environmental appeal, is calculated to equal the losses of the wine growers and is implemented as an ecological regulatory charge. No doubt this power penny satisfies Smith's criterion of a certain tax, and it can be readily implemented, since the charge can just be added to the normal bill. Yet the power penny is unlikely to have any significant effect on the activity of the power plant that causes the externality. Although the vintners may be compensated for the damages they suffer, the economy as a whole still suffers loss, with either less wine production or wine of lower quality, or both, while energy will be more expensive for all. Hence the bundle of products available to consumers in the economy will be smaller with the same money income; the power penny only spreads the externality, it does not internalize it. The power penny, in this sense, is no different from the smokestack. It just spreads the externality, like a financial equivalent to the chimney.

An environmentally and ecologically relevant tax needs to be a tax that affects the choice-influencing costs. In designing such a choice-influencing environmental tax, we have to look for the time and locus at which environmentally relevant choices are made. Since these choices are not made on any basis of periodicity as a normal tax would require, an incentive needs to be created which triggers the tax. This may sound paradoxical, since a tax is by definition an involuntary payment that for this very reason tends not to be initiated by the entity to be taxed. In the case of environmentally relevant taxes, however, the situation must be different. Neither choice-influencing costs, which are subjective, nor locus and time of decision making are directly or indirectly observable by an outside agency, such as a tax authority. In our example, the appropriate decisions with respect to the design of the plant, the production process and the management of production with a view to allowing different uses of the same resource can, but need not, be made before production activity actually starts or when new investments are due.

The objective of the tax is to encourage environmentally and ecologically preferable plant design, production processes and production management, but what this design, these processes and the management will be is known only to the decision makers involved, and only after an appropriate decision-making process has been launched. This implies that the tax involved can only be of the type of an incentive. If that is the case, however, the question arises as to what might constitute the tax base.

As we noted above, partitioning of property rights so as to make different resource use compatible in fact moves the production possibility frontier in a north-eastern direction. The resulting production potential constitutes a source of earnings for the participants, which in itself constitutes a tax base. When a tax authority tries to encourage the creation of a stream of income in order to broaden the tax base, it can grant deductions as an incentive. Such deductions can then form an appropriate lever to encourage environmentally or ecologically desired changes, which cannot be prescribed but can only be triggered by way of incentive. The system then takes on the following form. When a particular project is to be added to an existing ecosystem, as in our case of the power plant, building and operation permits need to be secured as usual. Similarly, financing will be dependent on these permits. Under the present regime, no particular attempt at redefining property rights through multiple negotiations can be expected. The transactions costs required to lead to an optimal repartitioning of property rights in these cases can be assumed to be substantial. It is here that the ecological tax instrument must provide a lever.

An opportunity for renegotiating a partitioning of property rights in order to make different uses of the same resource compatible can be created by providing for a tax incentive on the grounds of ecologically desired improvements. If a scheme such as the one outlined above can be presented and

shown to lead to substantial savings and/or extended production possibilities, an x year tax credit for the income generated through the scheme can be granted upon certification by an environmental regulatory authority. The tax credit can, in turn, be used as collateral for the outlays needed in order to facilitate implementation of the agreement arrived at. This solution implies that the information needed for the decision of the tax authority is totally forthcoming from the parties involved and, as it needs to be certified by the environmental regulatory authority and does not require any additional information gathering on the part of the tax authority there is little room for error or misjudgement. The tax credit can easily be implemented within the context of current procedures, and the tax base has been broadened by the volume of the income generated (I) through repartitioning of property rights multiplied by $(n - x)$, with n being the time during which taxes can be expected to be levied. The tax base is:

$$\sum_{1}^{n} I\left(\frac{n-x}{r}\right),$$

with r being the relevant discount rate.

As we look at Adam Smith's criteria, we notice that this scheme violates every single aspect of his principle of certainty. The tax incentive is not the same for each taxpayer, as it crucially depends on the specifics of the repartitioning scheme negotiated. Although the tax incentive is not arbitrary, it is certainly not well known in advance. The time at which it becomes effective depends only on the activities of the taxpayers, as does the manner of payment, as even does the quantity to be paid, and none of these is clear and plain in advance. What becomes clear and plain is the tax credit, as this instrument needs to be used as collateral. The tax credit, in turn, drives the implementation of the scheme. If the authorities err in setting the tax credit too low, the scheme will not be launched and the additional increase of the tax base will not be forthcoming. If they set it too high, they still increase the tax base but forgo some tax revenue.

Interestingly enough, the environmental tax as a tax credit satisfies the double dividend criterion. The double dividend even has an institutional realization. The environmental dividend is being certified by the environmental authority and the fiscal dividend is being certified by the tax authority granting the tax credit.

A simple illustration

At the heart of the proposal made in this entry lies Buchanan's notion of choice-influencing costs and, specifically, choice-influencing taxes. The proposal is driven by the need to define a choice-influencing tax, since the point

of introducing ecological taxes is to influence decision makers in their ecologically relevant behaviour. For purposes of illustration, we have defined the stylized situation of a river valley with three partly incompatible uses of the ecosystem: wine growing, fishing and power generation. Of course, this is only a stylized model; any degree of complexity can be introduced without affecting the basic issue, which is to define property rights so as to make the different uses of the resource compatible. By repartitioning property rights, the production possibilities of the economy (here in the river valley) are being broadened, but these benefits are partly offset by the transactions costs that need to be incurred in order to bring about the new definition of property rights.

If taxes are to influence choices, it is important to define or identify those decision points where choices can actually be made. The difficulty with identifying such decision points is that a tax administration cannot by itself accomplish this task. Only a general rule broadly describing the decisions and their consequences will allow the identification of a particular decision point, a point in time which is subject to decisions by the taxpayer. This is in itself not an anomaly. In the presence of capital gains taxes, for instance, taxpayers will make their decision so as to minimize the liability, and they will therefore choose the point of a transaction that leads to a capital gain with a view to minimizing the tax liability. In the case of ecological taxes, we want to stimulate neither capital accumulation nor tax avoidance but a positive ecological decision, which means that an incentive needs to be defined so as to induce parties to change their behaviour in an ecologically positive way. In our stylized example, the decision requires cooperation of the three parties involved, with respect to (i) the choice of technology and process management of power production; (ii) the choice of fish to breed and harvest; and in all likelihood (iii) also the choice of grapes. The crucial decision involves the consensual repartitioning of property rights in order to make a resource use compatible, and on this repartitioning decision rest the three decisions identified above.

An important issue to be raised with respect to any proposal for an ecological tax concerns the informational requirements to be demanded from the different parties involved. In general, it cannot be assumed that a tax authority can develop expertise in complicated environmental issues. In our stylized example, no such information is required of the tax authority. The information rests exclusively with the parties involved, and it needs to be presented in a form that is amenable to control and certification by a specialized public authority typically dealing with environmental issues. The burden of documentation and proof lies entirely on the party or parties claiming the tax credit. Hence the informational requirements with respect to the administrative feasibility of the ecological tax proposed are minimal.

An important issue in this case is technology choice. The stylized example involves two different types of technology choice. On the one hand, inside the ecological system as defined above, the relevant technology choice refers to the cooling system, which in our example must be flexible enough to provide for either warm water directly emitted into the river, or water being retained within the cooling system long enough to prevent the formation of mist at crucial moments. However, another technology choice is relevant with respect to the ecological system itself and interests outside this system. This is the choice of fuel to be used for the generation of electrical power, such as carbon or nuclear energy.

An important concern among economists always revolves around the practicability of the solution proposed and the possibility of strategic behaviour and bargaining. With respect to Pigouvian taxes, it has often been pointed out that they create an incentive to increase the externality prior to imposition of the tax. The tax credit, which is the form the ecological tax takes in our example, creates no such incentive. The tax credit is granted to the extent that an improvement can be demonstrated. The tax credit can only be claimed against income actually generated after and due to the improvement. There is no possibility for the parties to increase the externalities in question before entering into the repartitioning agreement, since that would harm the interests of at least two of our three parties assumed.

Related to the bargaining issue is the problem of whether the tax credit granted can be too high or too low. In this context, it is important to establish upper and lower limits. The lower limit obviously is zero, which means that no incentive is present. The upper limit depends on the income claimed to be generated as a result of the ecological improvement. Here the tax authority relies on data provided by the taxpayers, who have no incentive to overestimate their projected income, as this will eventually be subject to taxation. Error on the part of the tax authority can be readily contained by being conservative with respect to the number of years over which the tax credit is granted. On the other hand, being too conservative means standing in the way of stimulating ecological improvements. Obviously, there are error possibilities of the first and of the second kind which lie within the scope of discretion of the tax authority, as in any other case of taxation.

Finally, we should point out that this particular proposal is also generalizable to the problem left outside the discussion so far: the choice of fuel for electric power generation. The important aspect here is that the ecological unit needs to be defined differently, and consequently also in all likelihood a different ecological tax unit with a different set of tax and environmental authorities involved will have to make the respective decisions. In principle, however, incompatible resource use can also in this case be resolved by attempting a repartitioning of property rights. Whether the transactions costs involved in

repartitioning these property rights are outweighed by the possible benefits will only become apparent once the ecological tax credit is in place.

Concluding remarks

Regulatory taxes can be of two kinds. Alongside the well-known Pigouvian taxes, law and economics analysis points to a second, more radical type. In the context of environmental taxation, an effective ecological tax is a tax credit on personal or corporate income taxes upon certifications of an ecological improvement scheme. The tax credit kicks in when the scheme becomes effective. The operation of the scheme depends solely on the cooperation of taxpayers; it requires no additional information on the part of necessarily underinformed public authorities. Interestingly, there is a double dividend in the sense that, along with an ecological improvement, there is a new revenue stream for the government.

Notes

1. I should like to thank Antony Dnes, Santos Pastor, Peter Senn, Richard E. Wagner and Wolfgang Weigel for helpful comments.
2. In mentioning Pigou's name, I feel compelled to add that Pigou himself felt corrective taxes to be impossible, as the taxing authority would lack the requisite information: 'But the practical difficulty of determining the right rates of bounty and of duty would be extraordinarily great. The data necessary for scientific decision are almost wholly lacking' (Pigou, 1947, pp. 42–3).

References

Backhaus, J. (1995), 'Subsidiarity and ecologically based taxation: aspirations and options', in Sabine Urban (ed.), *Europe in Progress*, Wiesbaden: Gabler.

Bovenberg, L. and R. de Mooij (1994), 'Environmental levies and distortionary taxation', *American Economic Review*, **84** (4), September, 1085–9.

Buchanan, J.M. (1969), *Cost and Choice: An Inquiry in Economic Theory*, Chicago: Markham.

Pigou, A.C. (1947), *Socialism vs. Capitalism*, London: Macmillan.

Schneider, F. and J. Volkert (1997), 'Die Realisierung ökologisch-orientierter Wirtschaftspolitik – eine Unmöglichkeit? Überlegungen aus Sicht der Neuen Politischen Ökonomie' [The realization of an ecological-oriented economic policy – an impossibility? Observations from the vantage point of the new political economy], in Sylke Behrends (ed.), *Ordnungskonforme Wirtschaftspolitik in der Marktwirtschaft*, Berlin: Duncker & Humblot.

Smith, Adam (1971), *The Wealth of Nations*, London: Dent.

PART VI

DISPUTE RESOLUTION

26 Dispute resolution

*Thomas J. Miceli**

Introduction

Economists are interested in the resolution of legal disputes for several reasons. First, the manner in which parties resolve their disputes (whether by settlement or trial) has an effect on the cost of operating the legal system. Second, economic theory can explain how parties will resolve their disputes (positive analysis), and can therefore prescribe procedural rules for lowering the cost of dispute resolution (normative analysis). Finally, legal disputes provide the raw material from which courts make new legal rules or reaffirm old ones. These rules in turn serve as guides for decisions that affect the rate at which disputes will arise in the future (for example, investment in precaution against accidents).

This entry examines the economic analysis of the way legal disputes are resolved. It begins by examining the settlement–trial decision. The analysis shows that settlement will be the predominant outcome given the mutual gain to the parties from avoiding litigation costs. In order to explain the existence of trials, researchers have proposed the existence of differing perceptions or asymmetric information by the parties regarding the outcome of a trial.

The resulting models offer insights into various methods for controlling the costs of litigation. In this context, we examine the practice of pre-trial discovery and cost shifting under the English Rule and Rule 68 of the Federal Rules of Civil Procedure. The discussion then turns to the question of whether there is too much litigation from a social perspective, and to various explanations for the success of frivolous lawsuits, or suits with little or no social value. Finally, the entry considers the question of the selection of cases that go to trial. This issue is relevant for researchers seeking to draw inferences about the population of disputes from trial data, and also for the question of whether the law evolves in the direction of efficiency without the conscious help of judges and litigants.

Modelling the settlement–trial decision

Consider the following simple model of a legal dispute. A plaintiff files a lawsuit in an effort to recover monetary damages against a defendant.[1] Prior to trial, the parties engage in bargaining in an effort to arrive at a settlement. In fact, the vast majority of civil actions result in settlements and therefore never go to trial. An economic model of dispute resolution seeks to explain

this fact by showing the mutual gains to the parties from settling outside of court. This turns out to be easy; the more difficult task is to explain why a small percentage of cases actually go to trial. A simple model illustrates both the gains from settlement and possible reasons for trial.

Let $P_pJ - C_p$ be the plaintiff's expected value of trial, where P_p is the plaintiff's assessment of her probability of victory at trial, J is the expected award in the event of a victory, and C_p is the cost of trial. Assume (for now) that $P_pJ - C_p$ is positive. Similarly, let $P_dJ + C_d$ be the defendant's expected cost of trial, where P_d is the defendant's assessment of the plaintiff's probability of victory and C_d is the defendant's cost of trial. (Although we have assumed that the parties agree on J, in general they could hold different assessments of this variable as well.)

A pre-trial settlement involves a payment $S \geq 0$ that the defendant makes to the plaintiff to resolve the dispute. (Assume that there is no cost of reaching such a settlement.[2]) The plaintiff will prefer to accept S rather than go to trial if $S \geq P_pJ - C_p$, and the defendant will prefer to pay S rather than go to trial if $S \leq P_dJ + C_d$. Combining these two conditions yields a necessary condition for a settlement:

$$P_dJ + C_d \geq P_pJ - C_p,$$

or

$$C_p + C_d \geq (P_p - P_d)J. \tag{26.1}$$

If this condition holds, there exists a value of S that is mutually acceptable to the parties compared to trial. That is, a 'settlement range' exists.

Note that this condition necessarily holds if $P_p = P_d$, or if the parties agree on the expected outcome of a trial. This makes sense since, if the parties agree on how the court will settle the dispute, they can implement the same outcome in a settlement while saving the joint cost of trial, $C_p + C_d$. An additional reason for settlement not captured in this model is that risk-averse disputants will prefer the certain outcome of a settlement as compared to the uncertain outcome of a trial.

The simple bargaining model of litigation thus provides two reasons for out-of-court settlements: avoiding the costs of a trial and avoiding the uncertain outcome of a trial. However, it remains to explain why, despite these gains, some cases still end up at trial. One explanation for a trial is that the inequality in (26.1) is reversed so that a range for S does not exist. Note that this requires $P_p - P_d$ to be positive and larger than the ratio of the joint costs of a trial to the expected judgment. One way $P_p - P_d$ can be positive is if the parties hold *differing perceptions* about the outcome of a trial.[3] In particular,

$P_p > P_d$ reflects *optimism* by both parties about their prospects at trial. Optimism about the outcome of a trial can arise from different views by the parties about the strengths of their cases, about the relevant law, or it may simply arise from irrationality (Cooter and Ulen, 1988, p. 487).

Another reason why (26.1) may not hold in a given case is that the parties may possess *asymmetric information* about the value of trial.[4] For example, suppose that a particular plaintiff knows her probability of victory at trial, P_p, but the defendant only knows the distribution of P_p across the population of all plaintiffs, where the variation in P_p could reflect, in the case of an accident, differences in the extent to which plaintiffs were contributorily negligent.[5] Suppose that, in this setting, the defendant makes a single, take-it-or-leave-it settlement offer, S, to minimize his expected costs.[6] As long as the cost-minimizing choice of S (denoted S^*) is less than the expected value of trial for the plaintiff with the highest expected value in the population, some cases will go to trial. In particular, plaintiffs with $P_pJ - C_p \leq S^*$ will accept the offer and settle, while plaintiffs with $P_pJ - C_p > S^*$ will reject the offer and go to trial.

Even when inequality (26.1) is satisfied, implying the existence of a settlement range, a trial can still occur if the parties are unable to decide on the division of the gains from settlement. If they cannot agree on the division, bargaining may break down and a trial will ensue (Cooter et al., 1982).

Implications of the model for the cost of litigation

The preceding model yields several implications regarding the structure of various procedural and cost-shifting rules aimed at promoting settlement and reducing the cost of litigation. This section examines pre-trial discovery and cost shifting under the English Rule and Rule 68.

Pre-trial discovery

The practice of pre-trial discovery allows the parties to a legal dispute to obtain information about their opponent's case prior to trial.[7] Condition (26.1) shows that discovery, if it is truly informative, will indeed increase the likelihood of settlement by reducing differences in the parties' perceptions about the outcome of a trial. In particular, discovery will tend to cause P_p and P_d to converge, thereby increasing the size of the settlement range.

In addition to promoting settlement, which lowers the cost of dispute resolution, discovery can also increase the fairness and accuracy of dispute resolution by causing settlements to better reflect the merits of a case, and it can reduce the likelihood that meritless claims will succeed (Cooter and Rubinfeld, 1994). At the same time, however, discovery can be abused by parties to a dispute merely to gain a strategic advantage in bargaining. For example, one of the parties can make excessive information requests in an

effort to impose heavy compliance costs on the other party, thereby inducing the latter to drop the claim or settle for a lesser amount.

The English Rule for allocating legal costs

The above model assumes that parties to a legal dispute bear their own litigation costs. This is referred to as the 'American Rule'. Several researchers have argued that the costs of litigation may be lowered by imposing different cost allocation rules. The most commonly mentioned alternative is the 'English Rule', which requires that the loser pay the winner's litigation costs. Advocates of the English Rule argue that it saves on litigation costs by discouraging plaintiffs from filing claims with little merit.[8]

To examine this claim in the context of the differing perceptions model, note that the plaintiff's expected value of trial under the English Rule is given by $P_pJ - (1 - P_p)(C_p + C_d)$, reflecting the fact that, if she wins, she pays no litigation costs, but if she loses she pays her own costs plus those of the defendant. Similarly, the expected cost of trial for the defendant under the English Rule is $P_d(J + C_p + C_d)$. Using the above approach for determining the feasibility of a mutually acceptable settlement yields the following condition for existence of a settlement range under the English Rule:

$$C_p + C_d \geq (P_p - P_d)(J + C_p + C_d). \tag{26.2}$$

Note that the right-hand side of (26.2) is larger than the right-hand side of (26.1), while the left-hand sides are the same. Thus a settlement is *less* likely under the English Rule than under the American Rule. The reason is that the English Rule increases the stakes of a trial relative to the American Rule since, in addition to J, the assignment of legal costs hinges on the outcome of the trial.[9] Counteracting this conclusion is the fact that risk-averse litigants will be *more* likely to settle under the English Rule because it makes trials riskier affairs (Coursey and Stanley, 1988; Donohue, 1991).

The preceding analysis implies that the English Rule may lead to higher litigation costs than the American Rule if it results in a higher trial rate among those cases that are filed. As noted, however, the purported advantage of the English Rule is that it discourages claims with little merit from being filed in the first place. The above model implies that a suit has a lower expected value to a plaintiff under the English Rule as compared to the American Rule if $P_pJ - (1 - P_p)(C_p + C_d) < P_pJ - C_p$, or if

$$P_p < C_d/(C_p + C_d). \tag{26.3}$$

Thus plaintiffs with low assessments of their probability of victory will indeed be less likely to file suit under the English Rule.[10] Empirical compari-

sons of the English and American rules have generally supported the above theoretical predictions (Snyder and Hughes, 1990; Hughes and Snyder, 1995).

Cost shifting under Rule 68

Another rule aimed at reducing the cost of litigation is Rule 68 of the Federal Rules of Civil Procedure, which requires a plaintiff who refuses a settlement offer to pay the defendant's post-offer legal costs if the plaintiff receives a judgment at trial less than the rejected offer.[11] (No sanction is imposed if the plaintiff loses at trial (Miller, 1986).) To examine this rule in the context of the above model, suppose that the defendant offers an amount S which the plaintiff refuses. Let $G(S)$ be the probability that the judgment at trial will be less than S, where G is increasing in S. Thus the plaintiff's expected value at trial under the threat of a Rule 68 sanction is $P_pJ - C_p - P_pG(S)C_d$ and the defendant's expected cost of trial is $P_dJ + C_d - P_dG(S)C_d$. Note first that both of these expressions are *lower* than the corresponding expressions with no fee shifting (that is, under the American Rule). Thus Rule 68, in contrast to the English Rule, is 'pro-defendant'. This will tend to make lawsuits less valuable for plaintiffs regardless of the merits of their cases, and defendants will therefore be able to settle cases for less. Although this pro-defendant bias may lower litigation costs, it is unfair to plaintiffs with meritorious claims and may tend to discourage defendants from taking precaution to avoid disputes.

As for the impact on the settlement rate, note that the above procedure implies that the condition for existence of a settlement range under Rule 68 is:

$$(C_p + C_d) + (P_p - P_d)G(S)C_d \geq (P_p - P_d)J. \tag{26.4}$$

Compared to (26.1), this shows that if $P_p - P_d > 0$, settlement is more likely under Rule 68 as compared to the American Rule. Intuitively, the expected sanction reduces the value of trial more for plaintiffs than it reduces the cost of trial for defendants (given $P_p > P_d$), thereby increasing the settlement range. However, if $P_p - P_d < 0$, the reverse is true; that is, a settlement may be less likely under Rule 68 because the settlement range contracts. Thus the model does not provide a clear prediction about whether Rule 68 promotes settlement of legal disputes.

Is there too much litigation?

Given the high cost of the legal system for resolving disputes, an important normative question is whether there is too much litigation from a social perspective. This section addresses this question by examining two issues: the social versus private incentive to bring suit, and economic explanations for the success of frivolous suits. It concludes by discussing alternative approaches to dispute resolution as a response to the high cost of litigation.

The social versus private incentive to sue

The private incentive to sue is based on a plaintiff's comparison of her cost of bringing suit and the expected settlement amount or award of damages at trial. In contrast, the social value of suit is based on a comparison of the social cost of litigation, including the defendant's costs plus any public costs, and the social value of a suit, which consists of the incentives it creates for injurers to take reasonable care to reduce the risk of harm. In general, these comparisons will differ, resulting in an actual level of litigation that diverges from the socially optimal level. The amount and direction of the divergence depends on the interaction of two effects. First, plaintiffs ignore the cost of a suit to others; this effect will tend to cause too much litigation. Second, plaintiffs ignore the impact of litigation on the incentives of injurers to take cost-justified precaution against risk.[12] This effect can lead to too little or too much litigation. On net, therefore, it is impossible to determine theoretically whether there is too much or too little litigation (Shavell, 1982b, 1997).[13]

Frivolous suits

Although the previous section noted that lawsuits can provide a social benefit by deterring unreasonably risky behaviour, there is a common perception that a significant fraction of lawsuits have little or no social merit and are filed solely in the hopes of obtaining a settlement. Since these so-called 'frivolous' suits impose a significant burden on the legal system, an analysis of how they succeed and how they can be discouraged is an important topic in the economic analysis of litigation.

It is first necessary to define a frivolous suit. One definition is that it is a suit without value at trial – either the plaintiff has sustained no damages ($J = 0$) or the claim is without merit ($P_p = 0$). A broader definition includes suits that may have some merit ($P_pJ \geq 0$), but whose value at trial is less than the cost of a trial, that is, $P_pJ - C_p < 0$. The latter are referred to as 'negative expected value' (NEV) suits (Bebchuk, 1996). Henceforth we shall employ this definition.

The key question regarding NEVs is, why would a defendant agree to settle such a suit, knowing that the plaintiff would never rationally take the case to trial? As an illustration, suppose that a plaintiff files a NEV suit. The efendant then has the option to offer a positive settlement amount or zero. learly, the plaintiff will accept any $S > 0$, but, if faced with $S = 0$, she will rop the suit rather than go to trial since $P_pJ - C_p < 0$. Anticipating this, the efendant will refuse to make a positive settlement offer. A rational plaintiff ill therefore not find it profitable to file a NEV suit in the first place. As this xample illustrates, the success of NEVs turns on the *credibility* of the laintiff's threat to go to trial if the defendant refuses to settle.

One explanation for the success of NEVs is based on the asymmetric information model described above (Bebchuk, 1988; Katz, 1990). Suppose that there are two types of potential plaintiff: those who are truly injured and those who are not. Perhaps both types of plaintiff were involved in accidents, but only the former suffered injuries. Though defendants know the fraction of uninjured plaintiffs in the population of accident victims, they cannot distinguish the injured from the uninjured in individual cases. In this setting, uninjured victims can succeed in obtaining positive settlement amounts if the fraction of uninjured plaintiffs in the population is small enough. Intuitively, defendants are willing to pay a small number of meritless claims in order to avoid the cost of taking those cases with merit to trial. The success of NEVs in this case therefore results from their ability to masquerade as truly injured plaintiffs.

A weakness of the asymmetric information model for explaining the success of NEVs, however, is that it cannot account for settlements in which the defendant *knows* that the suit is NEV. Bebchuk (1996) has developed a model that overcomes this weakness by appealing to the sequential nature of the litigation process.[14] In particular, he notes that actual pre-trial bargaining unfolds over a sequence of periods, during each of which some of the total costs of litigation are spent. Thus, although $P_pJ - C_p < 0$ for a NEV, as long as $P_pJ > 0$, it may pay for the plaintiff to incur the litigation costs in any single stage if a settlement has not yet been reached.

To illustrate, suppose there are n stages over which the total costs of litigation are divided evenly (a stronger assumption than necessary). Then $P_pJ - (C_p/n) > 0$ is possible even though $P_pJ - C_p < 0$. Thus, beginning in the final stage, backwards induction shows that the plaintiff has a credible threat to proceed with the case through each of the n stages of litigation. Consequently, it pays the defendant to settle the case in stage one for a positive amount.

Alternative dispute resolution

The high cost of litigation has caused disputants to rely increasingly on alternative methods of dispute resolution, such as arbitration and mediation. The principal advantage of these alternatives is the cost savings resulting from the reduced complexity and length of proceedings compared to full-length trials. However, an evaluation of the social desirability of these methods must also consider the accuracy of their results and the incentive effects that they create for parties to avoid disputes in the first place (Shavell, 1995).

The selection of disputes for litigation

As noted above, most disputes settle before reaching trial, and since settlements are private, very little is known about the characteristics of cases that

settle compared to those that go to trial. And, since all available data on legal disputes come from those cases that proceed to trial, theoretical models of dispute resolution are needed to provide insight into the nature of the bias, if any, that is reflected in trial data.[15]

Using the differing perceptions model, Priest and Klein (1984) advanced the hypothesis that the plaintiff win rate in cases that go to trial will tend to be 50 per cent. This hypothesis is based on the claim that the set of cases that go to trial is not a random sample of all cases, but rather consists disproportionately of cases that are a 'toss-up'. This claim follows from the argument that cases in which one side is a clear winner will result in little disagreement between litigants and will therefore be settled, while cases with no clear winner will result in greater disagreement and hence a higher trial rate. Several analysts have since re-examined the Priest–Klein hypotheses from theoretical, empirical and experimental perspectives and have found evidence for the existence of a bias in trial data but little evidence for the '50 per cent rule'.[16] As an example, recall from above that the asymmetric information model predicts that trials will consist primarily of plaintiffs with high probabilities of victory (Baird et al., 1994, ch. 8).

Both the differing perceptions and asymmetric information models predict that a trial is more likely the larger are the plaintiff's damages, all else equal (that is, (26.1) is less likely as J increases). This, combined with the fact that inefficient legal rules by definition produce larger damages than do efficient rules, implies that inefficient rules will be litigated at a higher rate than efficient rules. As a result, judges will have a disproportionate opportunity to reconsider inefficient rules. And, as long as judges do not have a strong bias against efficiency, it follows that the number of efficient laws should grow relative to the number of inefficient laws. This conclusion has led to the hypothesis that the common law will tend to evolve in the direction of efficiency without the conscious help of judges or litigants (Priest, 1977; Rubin, 1977; Posner, 2003).

Several generalizations of the simple litigation model affect the strength of this hypothesis. First, if the costs of inefficient laws are dispersed rather than concentrated, they may not be litigated more frequently than efficient laws because individual litigants will not fully internalize the social benefits of overturning them. Second, if judicial decision by precedent is important, selective litigation may result in greater entrenchment of inefficient laws over time (Landes and Posner, 1979). Finally, if the parties' litigation expenditures are endogenous rather than fixed, they will tend to spend more on cases involving higher stakes (Goodman, 1978; Katz, 1988). If judges are responsive to litigant expenditures, this effect will tend to magnify the favoured party's position. Thus, if the relative private returns to the litigants reflect the social benefits, and if the court is not strongly biased by inefficient

precedents, this effect will accentuate the tendency of the law to evolve towards efficiency.

Until recently, theoretical models of litigation have largely ignored the role of judges in shaping the evolution of the law (aside from the impact of precedent). This was primarily due to the lack of a good theory of judicial decision making. However, this represents an important gap in our understanding of the common law process, so researchers have begun to examine the motivation of judges, both theoretically (Miceli and Cosgel, 1994; Posner, 1995, ch. 3), and empirically (Higgins and Rubin, 1980; Cohen, 1991).

Notes

* I wish to acknowledge the helpful comments of Steve Shavell.

1. The analysis is similar for disputes involving non-monetary judgments. See Shavell (1993).
2. This is a simplification, since pre-trial bargaining is clearly not costless. However, as long as bargaining is less costly than a trial, the results of the analysis continue to hold.
3. The differing perceptions model of litigation is based on the work of Landes (1971) and Gould (1973). Also see Posner (2003, ch. 21) and Cooter and Rubinfeld (1989).
4. See, for example, Png (1983), Bebchuk (1984, 1988), Reinganum and Wilde (1986) and Nalebuff (1987). Schweizer (1989) considers the case of two-sided incomplete information, and Spier (1992) explicitly accounts for the dynamics of pre-trial bargaining in the presence of asymmetric information. Baird et al. (1994, ch. 8) compare the differing perceptions and asymmetric information models.
5. Other versions of the asymmetric information model assume that the plaintiff's costs, C_p, vary, or that the defendant has private information about his cost of trial. See Daughety and Reinganum (1993) for a general analysis of these models. Hylton (1993) shows how differences in P_p and P_d can arise from errors by the court.
6. When the uninformed party makes the offer, the model is one of 'screening'. In contrast, when the informed party (the plaintiff in this case) makes the offer, the model is one of 'signalling' (Daughety and Reinganum, 1993; Baird et al., 1994, ch. 8).
7. For economic models of discovery, see Shavell (1989), Baird et al. (1994; ch. 8), Cooter and Rubinfeld (1994), Hay (1994) and Miceli (1997, ch. 8).
8. See, for example, Shavell (1982a), Katz (1987) and Donohue (1991).
9. The asymmetric information model similarly leads to the prediction that the English Rule leads to fewer settlements (Bebchuk, 1984).
10. Since most cases settle, the value of a claim will depend on the expected settlement amount rather than the expected return at trial. The result in the text holds as long as the settlement amount on average reflects the expected outcome of a trial.
11. See Miller (1986), Anderson (1994), Spier (1994) and Chung (1996).
12. Whether a lawsuit is settled or goes to trial can have different effects on deterrence. See Polinsky and Rubinfeld (1988).
13. Also see Menell (1983), Kaplow (1986) and Rose-Ackerman and Geistfeld (1987).
14. Also see Rosenberg and Shavell (1985).
15. Experimental evidence can also provide insights (Coursey and Stanley, 1988; Stanley and Coursey, 1990; Thomas, 1995), though this evidence ultimately relies on the underlying theoretical model of bargaining.
16. See, for example, Priest (1985), Wittman (1985), Eisenberg (1990), Stanley and Coursey (1990), Hylton (1993), Thomas (1995), Waldfogel (1995) and Shavell (1996).

References

Anderson, David (1994), 'Improving settlement devices: Rule 68 and beyond', *Journal of Legal Studies*, **23** (1), 225–46.

Baird, Douglas, Robert Gertner and Randal Picker (1994), *Game Theory and the Law*, Cambridge, MA: Harvard University Press.
Bebchuk, Lucian (1984), 'Litigation and settlement under imperfect information', *Rand Journal of Economics*, **15** (3), 404–15.
Bebchuk, Lucian (1988), 'Suing solely to extract a settlement offer', *Journal of Legal Studies*, **17** (2), 437–50.
Bebchuk, Lucian (1996), 'A new theory concerning the credibility and success of threats to sue', *Journal of Legal Studies*, **25** (1), 1–25.
Chung, Tai-Yeong (1996), 'Settlement of litigation under Rule 68: an economic analysis', *Journal of Legal Studies*, **25** (1), 261–86.
Cohen, M. (1991), 'Explaining judicial behavior, or what's constitutional about the sentencing commission', *Journal of Law, Economics and Organization*, **7**, 183–99.
Cooter, Robert, Stephen Marks and Robert Mnookin (1982), 'Bargaining in the shadow of the law: a testable model of strategic behavior', *Journal of Legal Studies*, **11** (2), 225–51.
Cooter, Robert and Daniel Rubinfeld (1989), 'Economic analysis of legal disputes and their resolution', *Journal of Economic Literature*, **27** (3), 1067–97.
Cooter, Robert and Daniel Rubinfeld (1994), 'An economic model of legal discovery', *Journal of Legal Studies*, **23** (1), 435–63.
Cooter, Robert and Thomas Ulen (1988), *Law and Economics*, New York: Harper-Collins.
Coursey, Don and Linda Stanley (1988), 'Pretrial bargaining behavior within the shadow of the law: theory and experimental evidence', *International Review of Law and Economics*, **8** (2), 161–79.
Daughety, Andrew and Jennifer Reinganum (1993), 'Endogenous sequencing in models of settlement and litigation', *Journal of Law, Economics and Organization*, **9** (2), 314–48.
Donohue, John (1991), 'Opting for the British rule, or if Posner and Shavell can't remember the Coase theorem, who will?', *Harvard Law Review*, **104**, 1093–119.
Eisenberg, Theodore (1990), 'Testing the selection effect: a new theoretical framework with empirical tests', *Journal of Legal Studies*, **19** (2), 337–58.
Goodman, John (1978), 'An economic theory of the evolution of the common law', *Journal of Legal Studies*, **7** (2), 393–406.
Gould, John (1973), 'The economics of legal conflicts', *Journal of Legal Studies*, **2** (2), 279–300.
Hay, Bruce (1994), 'Civil discovery: its effects and optimal scope', *Journal of Legal Studies*, **23** (1), 481–515.
Higgins, R. and P. Rubin (1980), 'Judicial discretion', *Journal of Legal Studies*, **9**, 129–38.
Hughes, James and Edward Snyder (1995), 'Litigation and settlement under the English and American rules: theory and evidence', *Journal of Law and Economics*, **38** (1), 225–50.
Hylton, Keith (1993), 'Asymmetric information and the selection of disputes for litigation', *Journal of Legal Studies*, **22** (1), 187–210.
Kaplow, Louis (1986), 'Private versus social costs in bringing suit', *Journal of Legal Studies*, **15** (2), 371–86.
Katz, Avery (1987), 'Measuring the demand for litigation: is the English rule really cheaper?', *Journal of Law, Economics and Organization*, **3** (2), 143–76.
Katz, Avery (1988), 'Judicial decisionmaking and litigation expenditure', *International Review of Law and Economics*, **8** (2), 127–43.
Katz, Avery (1990), 'The effect of frivolous litigation on the settlement of legal disputes', *International Review of Law and Economics*, **10** (1), 3–27.
Landes, William (1971), 'An economic analysis of the courts', *Journal of Law and Economics*, **14** (1), 61–107.
Landes, William and Richard Posner (1979), 'Adjudication as a private good', *Journal of Legal Studies*, **8** (2), 235–84.
Menell, Peter (1983), 'A note on private versus social incentives to sue in a costly legal system', *Journal of Legal Studies*, **12** (1), 41–52.
Miceli, Thomas (1997), *Economics of the Law: Torts, Contracts, Property, Litigation*, New York: Oxford University Press.

Miceli, T. and M. Cosgel (1994), 'Reputation and judicial decision making', *Journal of Economic Behavior and Organization*, **23**, 31–51.
Miller, Geoffrey (1986), 'An economic analysis of Rule 68', *Journal of Legal Studies*, **15** (1), 93–125.
Nalebuff, Barry (1987), 'Credible pretrial negotiation', *Rand Journal of Economics*, **18** (2), 198–210.
Png, Ivan (1983) 'Strategic behavior in suit, settlement and trial', *Bell Journal of Economics*, **14**, 539–50.
Polinsky, A. Mitchell and Daniel L. Rubinfeld (1988), 'The deterrent effects of settlements and trials', *International Review of Law and Economics*, **8** (1), 109–16.
Posner, Richard (1995), *Overcoming Law*, Cambridge, MA: Harvard University Press.
Posner, Richard (2003), *Economic Analysis of Law*, 6th edn, New York: Aspen.
Priest, George (1977), 'The common law process and the selection of efficient rules', *Journal of Legal Studies*, **6** (1), 65–82.
Priest, George (1985), 'Reexamining the selection hypothesis: learning from Wittman's mistakes', *Journal of Legal Studies*, **14** (1), 215–43.
Priest, George and Benjamin Klein (1984), 'The selection of disputes for litigation', *Journal of Legal Studies*, **13** (1), 1–55.
Reinganum, Jennifer and Louis Wilde (1986), 'Settlement, litigation and the allocation of litigation costs', *Rand Journal of Economics*, **17**, 557–66.
Rose-Ackerman, Susan and Mark Geistfeld (1987), 'The divergence between social and private incentives to sue: a comment on Shavell, Menell and Kaplow', *Journal of Legal Studies*, **16** (2), 483–91.
Rosenberg, David and Steven Shavell (1985), 'A model in which suits are brought for their nuisance value', *International Review of Law and Economics*, **5**, 3–13.
Rubin, Paul (1977), 'Why is the common law efficient?', *Journal of Legal Studies*, **6** (1), 51–63.
Schweizer, Urs (1989), 'Litigation and settlement under two-sided incomplete information', *Review of Economic Studies*, **56**, 163–78.
Shavell, Steven (1982a), 'Suit, settlement and trial: a theoretical analysis under alternative methods for the allocation of legal costs', *Journal of Legal Studies*, **11** (1), 55–81.
Shavell, Steven (1982b), 'The social versus private incentive to bring suit in a costly legal system', *Journal of Legal Studies*, **11** (2), 333–40.
Shavell, Steven (1989), 'Sharing of information prior to trial', *Rand Journal of Economics*, **20** (2), 183–95.
Shavell, Steven (1993), 'Suit versus settlement when parties seek nonmonetary judgments', *Journal of Legal Studies*, **22** (1), 1–13.
Shavell, Steven (1995), 'Alternative dispute resolution: an economic analysis', *Journal of Legal Studies*, **24** (1), 1–28.
Shavell, Steven (1996), 'Any frequency of plaintiff victory at trial is possible', *Journal of Legal Studies*, **25** (2), 493–501.
Shavell, Steven (1997), 'The fundamental divergence between the private and social motive to use the legal system', *Journal of Legal Studies*, **26** (2), 1–38.
Snyder, Edward and James Hughes (1990), 'The English rule for allocating legal costs: evidence confronts theory', *Journal of Law, Economics and Organization*, **6** (2), 345–80.
Spier, Katherine (1992), 'The dynamics of pretrial negotiation', *Review of Economic Studies*, **59**, 93–108.
Spier, Katherine (1994), 'Pretrial bargaining and the design of fee-shifting rules', *Rand Journal of Economics*, **25** (2), 197–214.
Stanley, Linda and Don Coursey (1990), 'Empirical evidence on the selection hypothesis and the decision to litigate or settle', *Journal of Legal Studies*, **19** (1), 145–72.
Thomas, Robert (1995), 'The trial selection hypothesis without the 50 percent rule: some experimental evidence', *Journal of Legal Studies*, **24** (1), 209–28.
Waldfogel, Joel (1995), 'The selection hypothesis and the relationship between trial and plaintiff victory', *Journal of Political Economy*, **103** (2), 229–60.
Wittman, Donald (1985), 'Is the selection of cases for trial biased?', *Journal of Legal Studies*, **14** (1), 185–214.

PART VII

DIFFERENT SOURCES OF THE LAW

27 Judicial independence

Sophie Harnay

Although the idea that the judiciary should be independent is relatively new and the mixing of judicial, legislative and executive duties commonplace for most of history, an independent judiciary is now seen both as a salient feature of a government under the rule of law and as the core of many constitutional arrangements in most developed countries as well as in the legal literature. A number of constitutional and legal writings by political philosophers and constitutional lawyers, including John Locke, Charles-Louis de Secondat Montesquieu, James Madison and Thomas Jefferson have defined judicial independence as an essential aspect of the separation of powers, central to the conception of the judiciary as the third branch of government. In this view, the separation of power into different functions, to be exercised by distinct branches of government that mutually check and balance each other, is aimed to prevent the government bodies from abusing the power that the constitution assigns to them. Within this framework, an independent judiciary is required to look after the balance of powers and to sanction possible unlawful abuses of power by the other branches. This requires that the decisions of the courts will not be altered or ignored by the other branches of government when they are in charge of their enforcement.

Threats to judicial independence are issued not only by the political branches of government but also by the particular interests at work in the judicial process. Judicial independence therefore also includes freedom from any external pressure, such as freedom from economic factors (judges must be free from having their financial well-being depend on the outcome of the cases they hear) and freedom from physical compulsion (judges cannot be coerced into rendering particular decisions). Thus, in accordance with the most basic precepts of due process, independence from private interests prevents a judge from being a judge in his/her own case. It also provides specific disqualification standards requiring judges to decline to give an opinion whenever they have pecuniary interest in a case. In addition, in reference to the notion of impartiality, the principle of judicial independence makes it necessary to sanction bribes and corruption situations that would make the pecuniary well-being of a judge directly dependent on the sense of the decision. In other words, independence requires the compensation of judges being determined so as to allow them to act upon their convictions without consideration of the personal consequences of their decision. From this perspective,

constitutional provisions organizing judicial independence can be expected to protect judges not only against external economic and social forces but also against themselves and their own temptations.

Consistently with the standard rationality and self-interest assumptions, the economic literature develops an analysis of judicial independence that is mostly based on the examination of the goals and motivations of agents. It points out both the political and litigation dimensions of the principle. This chapter will first present the variety of institutional arrangements aimed at organizing judicial independence and their various outcomes. Second, judicial independence from the political decision makers will be analysed through the lens of the economic analyses in both normative and positive dimensions. Independence from the litigants will then be discussed in relation to the rent-seeking problem. Finally, concluding comments will briefly present some controversies that shape the debate around judicial independence, in relation to the judicial accountability debate and the influence of peers on judicial decision making.

The variety of institutional arrangements

Beyond a general agreement on principles, the institutional structures aimed at guaranteeing judicial independence take very different forms. For instance, a host of institutional arrangements can be observed in the field of designation and recruitment procedures. Judges can be designated through partisan or non-partisan elections, as is the case in some US states, where judges are elected by the general public. They can also be politically appointed, like most of the judges in constitutional courts, or appointed through competitive recruitment procedures, like the French judicial bureaucracy. Furthermore, any one country can resort to the whole range of appointment procedures, depending on the type of court and level of jurisdiction. Similarly, various arrangements concerning judicial immunity from removal from office can be observed. For instance, while some systems choose to grant judges a life tenure, which is expected to enable them to make independent decisions without fear of resignation or transfer or of any other professional sanction, other countries or courts do not allow judges to be reappointed when their term of office comes to an end. This variety is backed up with a wide array of compensation and incentive mechanisms that regulate judicial salaries, promotions and career prospects.

Thus the diversity of institutional arrangements aimed at protecting judicial independence appears to be matched only by the variety of the means to infringe upon it. For instance, the judicial dependence on political decision makers for the budget, which exists in most countries, may give birth to manipulations of judicial rewards by the political sponsor in order to induce courts to issue particular decisions. Discrimination at the point of entry into

the profession and screening could also be expected to occur. As a consequence, judges could be expected to be more responsive to external pressures than is usually observed. However, quite surprisingly, political decision makers themselves exercise self-restraint in their use of the structural tools. The distinction between substantive and structural independence helps to account for this at first sight paradoxical situation in which judges are able to make independent decisions in spite of their institutionally dependent situation. More precisely, substantive independence refers to the courts' ability to issue decisions that diverge from the preferences of the political decision makers, whereas structural independence refers to the institutional arrangements that aim at enabling the existence of substantive independence (Salzberger, 1993). Consequently, degrees of substantive and structural independence may not coincide systematically and weak structural independence may not always have a detrimental effect on substantive independence. It might even allow courts to issue independent decisions, including in cases in which the government is a litigant, as long as the political branch agrees not to take advantage of the weakness of structural provisions to encroach on substantive independence. As a consequence, identical outcomes may be achieved under very different structural arrangements, whereas apparently similar constitutional provisions may lead to various outcomes in practice, contingent upon the behaviour of the other actors in the judicial field. Judicial independence may thus not be primarily a matter of constitutional protection or statutes and consequently restricted to positive law. It should rather be seen as the result of practical political and judicial experience as well as the outcome of the incentives of the other agents to respect independence.

Independence from political decision makers

A great deal of literature addresses the puzzle of an independent judiciary: on one hand, the so-called 'demonopolization theory' suggests a normative explanation based on an analogy with antitrust law; on the other, the law and economics and political economy literature favours a positive approach along the lines of the rational choice paradigm.

The demonopolization theory

Directly related to the view initially developed by constitutional framers and historically embodied in constitutional writings, the theory of the demonopolization of power considers the government as an initial monopolist in the political market whose possible abuses of power must be thwarted and counterbalanced by specific mechanisms. Consequently, a system of checks and balances must be put in place in order to allow a competitive situation to emerge. From this perspective, the separation of powers is aimed at balancing and controlling the different branches of government. Accordingly, an institu-

tionally distinct and independent judiciary stands as the guardian of public interest and fundamental rights against possible attacks by other branches of government. Insulated from the political process, judges are expected to act as impartial agents making the right decisions according to the democratic values that underlie the constitution. Thus, independence and the power to declare legislative and executive decisions unconstitutional allow them to limit rent-seeking activities by interest groups (Downs, 1957; Buchanan and Tullock, 1962; Buchanan, 1975; Epstein, 1987; Macey, 1988; Mashaw, 1990).

Law and economics models

Since it relies on a somewhat idealistic conception of the judiciary as a protection against private interests, the previous theory may run afoul of the standard assumptions of economic theory. By contrast, the law and economist's analyses try to reconcile judicial independence with the rational choice paradigm. According to the very title of Posner's article (1994), judges simply maximize 'the same thing everyone else does'. From this perspective, judges are explicitly depicted as self-interested rational maximizers of their own utility and assumed to respond to both monetary and non-monetary incentives. More precisely, they are seen as pursuing pecuniary and career rewards such as an increase in salary or a future promotion (Higgins and Rubin, 1980; Kimenyi et al., 1985; Cohen, 1991, 1992). Political ideology is also assumed to guide them in their decision-making process (Posner, 1986; Macey, 1994; Ashenfelter et al., 1995) as well as prestige and power (Miceli and Cosgel, 1994; Rasmusen, 1994).

Contingent on judicial objectives, various outcomes may be expected in terms of independence. In this respect, no conclusive consensus prevails in the theory with regard to judicial objectives and their effect on independence. Several empirical tests, mainly focusing on the US system, lead to contrasting evidence. Testing the correlation between salaries and the content of judicial decisions, Anderson et al. (1989) find a positive correlation between the independence of judges, measured by the number of overturned regulations and statutes, and the level of their salaries. In contrast, Toma (1991) establishes a positive correlation between the level of the budget received by the US Supreme Court and the proximity of its decisions to the preferences of the Congress. Testing whether the prospect of promotion may limit judicial independence, Higgins and Rubin (1980) claim that although high reversal levels lower the judge's chance of promotion, the effects of promotion are very limited. In particular, they find that the decisions of older judges are not overturned more frequently than young ones, whereas the former could be expected to decide more according to their own preferences than according to their promotion prospects. Ashenfelter et al. (1995) confirm this result in establishing no significant relationship between case outcomes and an ap-

pointing president's ideology. However, Cohen (1991) establishes a correlation between a judge's attitude towards federal guidelines and his/her promotion prospects. Cohen (1989, 1992) also finds that the imposition of harsh penalties on parties enables young judges to signal themselves to political decision makers and, thus, increases their promotion prospects. By contrast, older judges do not need to send such a signal since decision makers have already observed it through past decisions. Ramseyer and Rasmusen (1997) find that political factors determine to some extent the turnover of lower court judges in Japan, their appointment being dependent on a politically appointed Supreme Court. Arguing that federal judges could easily move into the private sector and get substantially higher pay there, Greenberg and Haley (1986) find that judicial utility functions are not dominated by pecuniary wealth but that judges make a tradeoff between money and other compensations such as status and power. Revesz (1999, 2001) also underpins the importance of ideology in judicial decision making.

Whatever their objectives, the maximization of their utility by rational judges makes them responsive to political control through the manipulation of traditional incentive and compensation systems by other branches of government. In that view, the judiciary is a specific form of bureaucracy for which traditional incentives are effective, under the form of either *ex ante* selection or *ex post* monitoring. Political control may then take either a political or an administrative form, each creating different incentives for judges and, therefore, inducing them to different behaviours (Elder, 1987). Within this framework, the question of the contribution of political decision makers to judicial independence, or why they organize and respect independent courts while they have structural means available to influence judicial decision-making is of the utmost importance.

The law and economics literature suggests several explanations for the possible discrepancy between structural and substantive independence. First, an interest-group explanation suggests that an independent judiciary ensures that both politicians and interest groups can maximize their gains from legislative deals. Initially developed by Landes and Posner (1975), it considers that independent courts enforcing the original legislative understanding increase the value of political rulings in the eyes of interest groups within the political arena. In other words, since they prevent future parties from reneging on their predecessors' legislative contracts, independent courts help legislators to raise money successfully and, thereby, provide interest groups with a stronger incentive to invest in durable legislation.

Building upon the Landes–Posner analysis, Ramseyer (1994) questions the 'puzzle' of the various levels of independence among national judiciaries. Towards that end, he models the electoral game as a sequential one affected by uncertainty about the electoral outcome. Using a repeated prisoner's

dilemma game, Ramseyer shows that a politician in office and expecting to stay in power during several consecutive legislative terms has an incentive to cheat on judicial independence but has an interest in an independent judiciary when he/she is not in office. Therefore, rational politicians will find it advantageous to contribute to independence when they fear losing elections but will rather defect when they expect to win them indefinitely. In other words, the preservation of independence will result from an intertemporal calculus through which political decision makers agree to increase their control over the judiciary in the future by decreasing their control in the present. Finally, the expected degree of political stability (will elections continue indefinitely or not?) and the likelihood of winning the elections (political turnover and competitiveness of the political market) will determine politicians' contribution to independence. Within this analysis, independent courts represent a cooperative equilibrium to the game, but not a unique one, and cooperation will be very sensitive to the threats of defection and non-reciprocity. Ramseyer (1994) and Ramseyer and Rasmusen (1997) provide comparative evidence from imperial and post-1945 Japan and the twentieth-century United States for this theory. Hanssen (2001) corroborates the analysis in the case of American states where he finds the most independent judicial institutions to be associated with higher levels of political competition and greater difference between political platforms. Stephenson (2003) also confirms the negative correlation between political instability and political contribution to judicial independence as well as the negative correlation between an independent judiciary and political competition in the electoral market.

However, the interest-group theory of judicial independence faces several limitations and criticisms. On the one hand, the empirical tests of the theory appear inconclusive. While Landes and Posner (1975) do not find any evidence corroborating their own theory, North and Weingast (1989) provide evidence for seventeenth-century England. Crain and Tollison (1979a) find that the interest-group theory of an independent judiciary explains the production of constitutional amendments as a function of legislative and judicial organizations relatively well. Arguing that a constitutional amendment provides a particularly durable type of long-term interest-group legislation due to its high cost of repeal, and using the tenure of the chief justice of a state's highest court as a proxy for judicial independence, they test the process of constitutional change across state governments in the United States. They find that the more independent the judiciary, the fewer the number of constitutional amendments. Drawing an analogy between the executive veto and the independent judiciary, Crain and Tollison (1979b) show that, like independent courts, the veto power enhances the durability of legislation and, therefore, increases the returns from legislative contracts with special interests by making these contracts harder to renege on. On the other hand, in

addition to this lack of clear evidence, the Landes–Posner theory is also criticized for some conceptual flaws. For instance, it does not address the question of the origin of judicial independence (Boudreaux and Pritchard, 1994). Assuming it as a constitutional rule – as Buchanan (1987) does – is not entirely satisfying, since this amounts to simply considering it as a gift to interest groups and Congress from the Constitution's framers, without giving the reasons for such a gift. Furthermore, whereas it puts the judiciary right at the core of the rent-seeking system designed by politicians, the interest-group theory of judicial independence does not assume self-interested judges but presumes that independent courts will always attempt to faithfully enforce the original tenor of the enacting legislature.

Yet several authors doubt that such a consensus favouring the original legislative intent exists among judges (Macey, 1988, 1994; Epstein, 1990; Boudreaux and Pritchard, 1994). Pointing out the absence of credible mechanisms originated by politicians and intended to make judges enforce past deals, they contend that courts may be more responsive to the expectations of the current legislature rather than willing to issue decisions that please past legislatures. As a consequence, the judiciary may lack the collective incentive to produce independent decisions and may individually have an incentive to free ride, that is, to make decisions consistent with the preferences of the current politicians. Similarly, fearing that the political decision makers in office at the last period may defect, current legislators may not be willing to forgo lucrative political deals and reward courts for their independence. In other words, an iterated prisoner's dilemma game without the ability to sanction future legislators will lead to intertemporal free-riding and underinvestment in the contribution to judicial independence. The fact that the rents available from manipulating past legislative contracts may be far greater than those from new deals, due to the respective number of contracts of both types, coupled with the public choice assumption of short-term preoccupied politicians, underpin the argument. Finally, arguing that the most powerful obstacle to dependent decisions resides mainly in its cost in terms of reputation within the judicial audience and scholarship, Boudreaux and Pritchard (1994) conclude that a more traditional vision of judicial independence as serving the long-term public good and based on the genuine convictions of framers might be as justified as the Landes–Posner interest-group theory.

A second explanation for the independence of the judiciary relies on the interest of politicians to maintain independent courts to which they can delegate legislative and other public decision-making powers (Voigt and Salzberger, 2002). Extending the contribution of the literature on the delegation of decision-making powers to administrative agencies, Salzberger (1993) argues that – besides traditional motives invoked to justify delegation, such as the lack of time or expertise and the need for decisiveness and anti-cycling

procedures – delegation to independent courts may help politicians to maximize their political support and chance of re-election. Indeed, delegation may enable the politicians to shift responsibility for unpopular decisions onto the courts. Since at the same time, they may continue to claim the credit for those decisions from those among whom they are popular, delegation may thus help to find a compromise between legislators whose constituencies respectively win and lose from the decision. Furthermore, compared to other possible delegatee bodies, courts allow the politicians to transfer responsibility all the more easily since judges are perceived as independent. Building further along similar lines, Josselin and Marciano (2001) consider the constitution as an incomplete contract between citizens (principals) and politicians (agents) requiring delegation to the courts in order to fill gaps in constitutional arrangements and produce rules in the presence of an 'open area' of the law (Posner, 1999). To that extent, delegation partly accounts for the 'judicialization' tendency and judicial empowerment in sensitive political debates (Hirschl, 2000).

In addition to this delegation motive, a third reason for political decision makers to contribute to judicial independence resides in the informational benefits provided by independent judicial review. Namely, for institutional as well as chronological reasons, courts have an informational advantage concerning the actual consequences of enacted legislation relative to legislators themselves when they initially passed it. An independent judiciary therefore conveys information to legislators that they cannot acquire through the normal legislative proceedings, which in turn affects the quantity and informational quality of legislation (Rogers, 2001). Similarly, independent courts play an informational key role in the relation between political decision makers and public bureaucrats. By providing politicians with information about the actual bureaucratic performance, they assist them in controlling their bureaucracies and, thereby, in gaining electoral support (McCubbins and Schwartz, 1984; Ferejohn and Shipan, 1990; Gely and Spiller, 1990).

A fourth line of explanation resorting to law and economics analyses highlights public support for judicial independence as a key factor for explaining the political contribution to independence. The reasons behind this are twofold. On the one hand, when independent courts enjoy a high level of public support, political attempts to infringe upon judicial independence may turn out costly in terms of political support. Therefore, the fear of electoral sanctions may induce politicians to enforce judicial decisions. Thus, they may be all the more willing to respect independence since the probability and cost of a sanction are high; this happens when voters are able to effectively monitor political responses to judicial decisions, and may explain why other branches of government comply with judicial rulings made by constitutional courts (Vanberg, 2000, 2001). On the other hand, beyond the obvious elec-

toral gain associated with a situation that is valued and praised by voters, judicial independence brings several advantages for political decision makers and facilitates their activity. Since independence leads to greater acceptance and respect of judicial rulings among the public, it decreases the cost of their enforcement and makes the administration of justice less costly. More broadly, independent courts are also expected to enhance the general legitimacy of all political decisions and to make their implementation easier (Boudreaux and Pritchard, 1994).

The Political Economy literature

Also building upon the rational choice paradigm, the political economy literature develops a general equilibrium model of checks and balances to analyse strategic interactions between the branches of government. However, by contrast with law and economics analyses, it initially focuses on the executive–legislative relationship and on the role of voters and interest groups within this framework, at the expense of the role of an independent judiciary. Pointing out this shortcut in the political economy approach as well as the restrictive partial equilibrium dimension of law and economics models, Padovano et al. (2003) propose to reconcile the two approaches and explicitly introduce self-interested courts into Persson et al.'s (1997) model of separation of powers. Comparing the impact of a respectively dependent and independent judicial branch on the accountability of a democratic government, they conclude that independent courts improve political accountability whereas a dependent judiciary alters the distribution of political rents without improving the accountability of the system.

Independence versus judicial discretion

Also assuming rationality and self-interested judges, the spatial models of judicial discretion focus on the institutional constraints that weigh on judicial action. Borrowing from the positive political economy analysis of administrative agencies (Weingast, 1984; McCubbins et al., 1987, 1989; Hammond and Knott, 1996), they consider that the different preferences of political decision makers over the policy space shape a discretion interval for courts. As long as political actors holding veto powers disagree about a simultaneously preferred solution, they are not induced to veto judicial decisions. Whenever courts make their decisions strategically and retreat from politically untenable positions, they are able to avoid reversal. The set of veto-proof solutions will thus measure the range of judicial discretion. Therefore, the ability of a court to make stable decisions derives not from the grant of independence guarantees to the judiciary by framers or legislatures, but rather from the possibility of making decisions that are also structure-induced equilibria.

Within this framework, the judiciary can be analysed as any public agency subject to institutional constraints and the notion of judicial discretion substitutes for the notion of independence to explain most observed judicial behaviours (Ferejohn and Weingast, 1992). For instance, judicial activism is seen as the outcome of a wide set of possible decisions that are made possible by the dissenting preferences of the other institutional players. By contrast, restrained behaviours of courts are motivated by a narrow set of feasible decisions shaped by the other actors' preferences. According to that view, a 'shy' decision by a judge simply results from his/her expectation of a reversal when making another decision or, in other words, is simply the hallmark of rational self-restraint. Empirical studies corroborate this expected influence of the institutional structure on judicial discretion whenever courts take into account politicians' preferences and the threat of a veto. For instance, Gely and Spiller (1990, 1992) and Spiller and Gely (1992) find that the choices of the US Supreme Court are actually constrained by Congress. Cooter and Drexl (1994) study the impact of the evolution of the decisional procedures in the European Community in relation to the discretionary power of courts and verify the main results of the spatial theory. Using a panel of many countries, Cooter and Ginsburg (1996) also find a positive correlation between judicial independence and the number of institutional vetoes framing the political process.

In spite of a few notable exceptions, most of the aforementioned analyses focus on the American setting, in particular on the US Supreme Court. Albeit developing, the analysis of the organization of an independent judiciary in civil law systems is less advanced than in common law regimes, although it obviously provides an important leeway for political intervention. According to the conception of adjudication prevailing in the country, the organization of the judiciary may also convey different implications in terms of judicial independence from the litigants.

Independence from the litigants

Another reason for seeking judicial independence is to avoid judicial decisions being dependent on private interests and driven by specific economic or social forces. Judicial independence then directly refers to the notion of impartiality. From this perspective, the economic analysis appears less forthcoming about the means for achieving it than about the issue of the dependence of the judicial decision on the influence of litigants and the various forms it can take. Indeed, by analogy with rent-seeking contests, it usually considers litigants as interest groups in competition for the assignment of a 'rent'. In this logic, contending parties try to induce the court to make a favourable decision by spending resources on legal research to produce arguments that will help the judge's decision in their case. The probability of being found

guilty or not guilty is supposed to be a function not only of the objective merits of the case but also of the litigation expenditure (Katz, 1988). From this perspective, since litigation induces a fee shifting that does not occur in alternative rent-seeking situations, it constitutes a specific type of rent-seeking distinct from pressure activities oriented towards legislators or bureaucrats (Farmer and Pecorino, 1999). Following up this line, an analogy can be drawn with the analysis of corruption and bribing that assumes passive judges responding to the pressures exerted by parties (Rose-Ackerman, 1979). The rent-seeking approach to litigation also builds on common premises with signalling models of judicial behaviour that consider judges imperfectly informed about the reality and issuing decisions according to the probability that they affect the realization of various states of nature (that is, the probability of a party being guilty or not). This probability represents a signal that the defendant can modify by engaging in a certain level of litigation expenditures, so as to induce the judge to revise his/her prior belief. The posterior belief of the judge is a function of the information acquired during the trial process. Rubinfeld and Sappington (1987) show that the standard of proof and the penalty imposed on defendants who fail to meet the standard are very sensitive to the relationship between the litigation efforts of defendants and the judge's ultimate assessment of a defendant's guilt. Thus, within that framework judicial decisions are determined mainly by the expenditures of litigants.

At a more general level, not only judicial decisions but also law can be seen as the outcome of the rent-seeking activities of parties. The extension of rent-seeking models of litigation from one-shot to repeated games assumes that parties demand not only judicial solutions to conflicts but also durable precedents as a joint product of the conflict resolution (Landes and Posner, 1976). Therefore, according to whether they expect to be involved in future conflicts or not, parties value a favourable decision differently and are more or less prone to engage in high amounts of legal expenditure. From this perspective, the level of litigation expenditures accounts for the judicial development of the law towards rules that are more favourable to powerful interests. In particular, parties able to organize so as to internalize the gains of litigation may succeed in changing the law in a favourable direction (Bailey and Rubin, 1994). In that sense, rent-seeking models satisfy the results of the critical law studies according to which the law evolves to benefit the most powerful interest groups. They also provide contrasting predictions with the efficiency theory of the common law. According to the latter, litigants have a stronger incentive to challenge inefficient rules before the courts since in the aggregate they derive higher gains from the transition from an inefficient rule to an efficient one than from the reverse evolution (Priest, 1977; Rubin, 1977; Goodman, 1978; Landes and Posner, 1979). By contrast, the rent-seeking

approach to legal evolution does not predict that efficient rules have a tendency to prevail in the long run, but rather focuses on the redistributional nature of many judicial decisions and suggests reasons why a number of legal 'anomalies' such as inefficient doctrines or inefficient changes may exist and survive (Osborne, 2002). In addition, it shows that litigation and lobbying are not disparate activities but that the forces shaping common law are similar to those shaping statute law. In other words, an interest group may obtain legal changes through either litigation or lobbying and have an incentive to combine the two methods for achieving its goals in complex legal battles (Rubin et al., 2001).

Whereas the rent-seeking models of litigation emphasize the momentous role of pressure groups in shaping the law, they do not so far clearly provide a satisfying analysis of the issue of judicial independence. On the contrary, they focus on the motives of litigants and confine the judiciary to the role of a supine provider of conflict resolution and rules in response to the relative effort of the parties. This restrictive vision of the judge is consistent neither with the law and economics assumption of self-interested agents nor with the noble conception of courts advanced by the legal scholarship. Further developments explicitly assume a judge to be a rational self-interested rent setter. Von Wangenheim (1993) finds that if selfish judges base their decisions on the opinions of their peers, the evolution of judge-made law may follow irregular movements rather than tend towards efficiency. Bailey and Rubin (1994) also mention that judges may be ideologically motivated and able to thwart legal evolution such as it would occur if driven by the mere pressure of litigants. Even then, however, judges remain reliant on the litigation behaviours of parties, since cases must be litigated to enable courts to modify the law. As a consequence, parties aware of the willingness or unwillingness of judges to effect legal change may adapt their litigation and seizure behaviours so as to let or not let the preferences of judges shape the law. Whether judges pursue efficiency, as assumed by Posner, or redistributional goals is not conclusive in the literature.

Judicial independence versus judicial accountability

A direct effect of judicial independence is that judicial decisions entail consequences for the other agents that judges do not have to bear themselves (Kirchgässner, 1992). With very few exceptions, this immunity from damage liability for actions performed within their official function holds even in the case of judicial error or misbehaviour. Building upon this fact, critics of judicial activism argue that judicial independence enables courts to act like super-legislatures, issuing arbitrary decisions while escaping any democratic constraint that elected officials are usually subject to. In other words, judicial decisions may be driven by the personal and political views of judges, or by

their preferences on policy issues at stake, without judges being accountable to the representatives or their constituents in some direct or indirect way. In that view, since the powers of the judiciary ought to be controlled and balanced in order to avoid or at least to limit arbitrary discretion of judges, the democratic requirement for accountable decision makers justifies resorting to specific provisions aimed at increasing the popular control of judges. Judicial election provides such a direct form of popular control.

Legislation restricting judicial review, guidelines limiting judicial discretion, or the development of alternative resolution dispute programmes, can be seen as indirect mechanisms aimed at improving judicial accountability to the political decision makers and, thereby, to the citizens. In the same manner, screening potential judges *ex ante* should not always be considered as an attempt to 'pack' courts with friendly judges, but rather as a way of improving the quality of the judicial output *ex post*. Thus, in a quite provocative way, political 'incursions' into the judicial field and the empowerment of politicians to regulate the exercise of jurisdiction should no longer be misinterpreted as political infringements on judicial independence and manifestations of a struggle between two branches of government. On the contrary, they become justified by the need for judicial accountability and plausibly serve judicial legitimacy (Kahn Zemans, 1999). This conception of the relationship between judicial and political actors strongly relies on the assumption of benevolent politicians or at least politicians unwilling to abuse their power. This may be inconsistent with the rational self-interested agent paradigm, unless voters display strong preferences for judicial independence and are able to control politicians in the electoral market.

A second way to improve judicial accountability without promoting political intervention in the judicial field is to put a legal constraint upon judges. Beyond impeachment or resignation procedures, which remain very rare, a legal constraint helps to reduce judicial leeway in the decision-making process. It guarantees that the individual decision is consistent with the law rather than based purely on the judge's own preferences. For instance, the control of lower courts by higher courts through appeal mechanisms together with the corresponding possibility of reversal of erroneous decisions provide a legal constraint on the lower court's decision that can be modelled through spatial models of judicial discretion (Gely and Spiller, 1990 for an application to the US Supreme Court; Golub, 1996; Tridimas, 2001, for an application to the European Court of Justice). Reference to the precedent in common law regimes also ensures that a judge cannot break from the rule that prevails in the legal community without a convincing justification. As a consequence, precedent introduces dependence on the peer group. At a more general level, dependence on the social group may also arise from the influence of legal scholars, litigants, lawyers and other social actors. Reputational concern there-

fore provides a judge individually with an incentive to conform with the prevailing opinion as long as adherence is not too costly in private terms (Miceli and Cosgel, 1994).

Finally, judicial independence cannot be restricted merely to its political dimension but should also be considered in its judicial, professional and social dimensions. An extensive definition of independence taking into account the influence of the peer group and of other social groups makes it clear that the judicial situation is better depicted as a continuum from total lack of regard for external pressures or minimal influence to absolute subservience to external factors. This raises the issue of the relationship between individual independence of judges and institutional independence of the judiciary and also questions the possible influence of the prevailing conception of adjudication under various legal regimes on the practical organization of judicial independence.

References

Anderson, G., W. Shughart and R. Tollison (1989), 'On the incentives of judges to enforce legislative wealth transfers', *Journal of Law and Economics*, **32**, 215–28.

Ashenfelter, O., T. Eisenberg and S. Schwab (1995), 'Politics and the judiciary: the influence of judicial background on case outcome', *Journal of Legal Studies*, **24**, 257–81.

Bailey, M. and P. Rubin (1994), 'A positive theory of legal change', *International Review of Law and Economics*, **14**, 467–77.

Boudreaux, D. and A. Pritchard (1994), 'Reassessing the role of the independent judiciary in enforcing interest-group bargains', *Constitutional Political Economy*, **5** (1), 1–21.

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

Buchanan, J.M. (1987), 'Constitutional economics', in J. Eatwell, M. Milgate and P. Newman (eds), *The New Palgrave: A Dictionary of Economics*, vol 1. A–D, London: Macmillan, pp. 585–8.

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

Cohen, M. (1989), 'The role of criminal sanctions in antitrust enforcement', *Contemporary Policy Issues*, **7**, 36–46.

Cohen, M. (1991), 'Explaining judicial behavior or what's "unconstitutional" about the sentencing commission?', *Journal of Law, Economics, and Organization*, **7**, 183–99.

Cohen, M. (1992), 'The motives of judges: empirical evidence from antitrust sentencing', *International Review of Law and Economics*, **12**, 13–30.

Cooter R. and J. Drexl (1994), 'The logic of power in the emerging European constitution: game theory and the division of powers', *International Review of Law and Economics*, **12**, 13–30.

Cooter, R. and T. Ginsburg (1996), 'Comparative judicial discretion: an empirical test of economic models', *International Review of Law and Economics*, **16**, 295–313.

Crain, W. and R. Tollison (1979a), 'Constitutional change in an interest-group perspective', *Journal of Legal Studies*, **8**, 165–75.

Crain, W. and R. Tollison (1979b), 'The executive branch in the interest theory of government', *Journal of Legal Studies*, **8**, 555–67.

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

Elder H. (1987), 'Property rights structures and criminal courts: an analysis of state criminal courts', *International Review of Law and Economics*, **7**, 21–32.

Epstein, R.A. (1987), 'Beyond the rule of law: civic virtue and constitutional structure', *George Washington Law Review*, **56**, 149–63.

Epstein, R.A. (1990), 'The independence of judges: the uses and limitations of public choice theory', *Brigham Young Law Review*, 827–55.
Farmer A. and P. Pecorino (1999), 'Legal expenditure as a rent-seeking game', *Public Choice*, **100**, 271–88.
Ferejohn, J., C. Shipan (1990), 'Congressional influence on bureaucracy', *Journal of Law, Economics, and Organization*, **6** (S), 1–20.
Ferejohn, J. and B. Weingast (1992), 'A positive theory of statutory interpretation', *International Review of Law and Economics*, **6** (12), 263–79.
Gely, R. and P. Spiller (1990), 'A rational choice theory of supreme court statutory decisions with applications to the *State Farm* and *Grove City* Cases', *Journal of Law, Economics, and Organization*, **6** (2), 263–300.
Gely, Rafael and P. Spiller (1992), 'The political economy of supreme court constitutional decisions: the case of Roosevelt's court packing plan', *International Review of Law and Economics*, **12**, 45–67.
Golub, J. (1996), 'The politics of judicial discretion: rethinking the interaction between national courts and the European Court of Justice', *West European Politics*, **19**, 360–85.
Goodman, J.C. (1978), 'An economic theory of the evolution of the common law', *Journal of Legal Studies*, **7**, 393–406.
Greenberg, P.E. and J.A. Haley (1986), 'The role of compensation structure in enhancing judicial quality', *Journal of Legal Studies*, **15**, 417–26.
Hammond, T. and J. Knott (1996), 'Who controls the bureaucracy? presidential power, congressional dominance, legal constraints and bureaucratic autonomy in a model of multi-institutional policy-making', *Journal of Law, Economics, and Organization*, **12**, 119–66.
Hanssen, F.A. (2001), 'Is there a politically optimal level of judicial independence?', Department of Economics, Montana State University, mimeo (April).
Higgins, R. and P. Rubin (1980), 'Judicial discretion', *Journal of Legal Studies*, **9**, 129–38.
Hirschl, R. (2000), 'The political origins of judicial empowerment through constitutionalization: lessons from four constitutional revolutions', *Law and Social Inquiry*, **25** (1), 91–149.
Josselin, J.-M. and A. Marciano (2001), 'Displacing your principal: two historical case studies of interest for the future of Europe', *European Journal of Law and Economics*, **10** (3), 217–33.
Kahn Zemans, F. (1999), 'The accountable judge: guardian of judicial independence', *Southern California Law Review*, **72**, 624–55.
Katz, A. (1988), 'Judicial decision-making and litigation expenditure', *International Review of Law and Economics*, **8**, 127–43.
Kimenyi, M.S., W.F. Shughart and R.D. Tollison (1985), 'What do judges maximize?', *Economia Delle Scelte Pubbliche*, **3**, 181–8.
Kirchgässner, G. (1992), 'Towards a theory of low-cost decisions', *European Journal of Political Economy*, **8**, 305–20.
Landes, W. and R. Posner (1975), 'The independent judiciary in an interest-group perspective', *Journal of Law and Economics*, **18**, 875–901.
Landes, W. and R. Posner (1976), 'Legal precedent: a theoretical and empirical analysis', *Journal of Law and Economics*, **19**, 249–313.
Landes, W. and R. Posner (1979), 'Adjudication as a private good', *Journal of Legal Studies*, **8**, 235–321.
Macey, J. (1988), 'Transaction costs and the normative elements of the public choice model: an application to constitutional theory', *Virginia Law Review*, **74**, 471–578.
Macey, J. (1994), 'Judicial preferences, public choice, and the rules of procedure', *Journal of Legal Studies*, **23**, 626–46.
Mashaw, J. (1990), 'Explaining administrative process: normative, positive and critical stories of legal development', *Journal of Law, Economics, and Organization*, **6**, 267–98.
McCubbins, M., R. Noll and B. Weingast (1987), 'Administrative procedures as instruments of political control', *Journal of Law, Economics, and Organization*, **3**, 243–77.
McCubbins, M., R. Noll and B. Weingast (1989), 'Structure and process, politics and policy: administrative arrangements and the political control of agencies', *Virginia Law Review*, **75**, 431–82.

McCubbins, M. and T. Schwartz (1984), 'Congressional oversight overlooked: police patrols versus fire alarms', *American Journal of Political Science*, **28**, 165–79.
Miceli, T. and M. Cosgel (1994), 'Reputational and judicial decision-making', *Journal of Economic Behavior and Organization*, **23**, 31–51.
North, D.C. and B.R. Weingast (1989), 'Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England', *Journal of Economic History*, **49**, 803–32.
Osborne, E. (2002), 'What's yours is mine: rent-seeking and the common law', *Public Choice*, **111**, 399–415.
Padovano, F., G. Sgarra and N. Fiorino (2003), 'Judicial branch, checks and political accountability', *Constitutional Political Economy*, **14**, 47–70.
Persson, T., G. Roland and G. Tabellini (1997), 'Separation of powers and political accountability', *Quarterly Journal of Economics*, **112**, 1163–202.
Posner, R.A. (1986), *Economic Analysis of Law*, Boston, MA: Little, Brown.
Posner, R.A. (1994), 'What do judges and justices maximise? (The same thing everybody else does)', *Supreme Court Economic Review*, **3**, 1–41.
Posner, R. (1999), *The Problematics of Legal and Moral Theory*, Cambridge, MA: Belknap Press.
Priest, G.L. (1977), 'The common law process and the selection of efficient rules', *Journal of Legal Studies*, **6**, 65–82.
Ramseyer, M. (1994), 'The puzzling (in)dependence of courts: a comparative approach', *Journal of Legal Studies*, **23** (2), 721–46.
Ramseyer, J.M. and E.B. Rasmusen (1997), 'Judicial independence in civil law regimes: the evidence from Japan', *Journal of Law, Economics, and Organization*, **13**, 259–86.
Rasmussen, E.B. (1994), 'Judicial legitimacy: an interpretation as a repeated game', *Journal of Law, Economics, and Organization*, **10**, 63–83.
Revesz, R.L. (1999), 'Ideology, collegiality, and the D.C. Circuit: a reply to Chief Judge Harry T. Edwards', *Virginia Law Review*, **85**, 805–51.
Revesz, R.L. (2001), 'Congressional influence on judicial behavior? an empirical examination of challenges to agency action in the D.C. Circuit', *New York University Law Review*, **76**, 1100–141.
Rogers, J.R. (2001), 'Information and judicial review: a signaling game of legislative–judicial interaction', *American Journal of Political Science*, **45** (1), 84–99.
Rose-Ackerman, S. (1979), *Political Corruption*, Academic Press: New York.
Rubin, P. (1977), 'Why is common law efficient?', *Journal of Legal Studies*, **6**, 51–63.
Rubin, P., C. Curran and J. Curran (2001), 'Litigation *versus* litigation: forum shopping by rent seekers', *Public Choice*, **107** (3–4), 295–310.
Rubinfeld, L. and D. Sappington (1987), 'Efficient awards and standards of proof in judicial proceedings', *Rand Journal of Economics*, **18** (2), 308–15.
Salzberger, E. (1993), 'A positive analysis of the separation of powers, or: why do we have an independent judiciary?', *International Review of Law and Economics*, **13**, 349–79.
Spiller, P. and R. Gely (1992), 'Congressional control or judicial independence: the determinants of U.S. Supreme Court labor-relations decisions, 1949–1988', *Rand Journal of Economics*, **23** (4), 463–92.
Stephenson, M.C. (2003), 'When the devil turns…: the political foundations of independent judicial review', *Journal of Legal Studies*, **32**, 59–89.
Toma, E. (1991), 'Congressional influence and the Supreme Court: the budget as a signaling device', *Journal of Legal Studies*, **20**, 131–46.
Tridimas, T. (2001), 'The European Court of Justice under the Treaty of Amsterdam', in P. Lynch, N. Neuwahl and W. Rees (eds), *Reforming the European Union: From Maastricht to Amsterdam*, London: Longman, pp. 74–84.
Vanberg, G. (2000), 'Establishing judicial independence in Germany: the impact of opinion leadership and the separation of powers', *Comparative Politics*, **32**, 333–53.
Vanberg, G. (2001), 'Legislative–judicial relations: a game-theoretic approach to constitutional review', *American Journal of Political Science*, **45** (2), 346–61.

Voigt, S. and E. Salzberger (2002), 'Choosing not to choose: when politicians choose to delegate powers', *Kyklos*, **55** (2), 289–310.
Von Wangenheim, G. (1993), 'The evolution of judge-made law', *International Review of Law and Economics*, **13**, 381–411.
Weingast, B. (1984), 'The congressional–bureaucratic system: a principal–agent perspective (with applications to the SEC)', *Public Choice*, **44**, 147–91.

28 General norms and customs

Jean-Michel Josselin and Alain Marciano

Introduction

Economic interactions between individuals take place within institutions which can be defined either as informal constraints (sanctions, customs, taboos, traditions and codes of conduct) or as formal rules (constitutions, laws, property rights) (North, 1991). Social norms or general norms can be defined as the broader set of rules including informal constraints and formal rules. Therefore, the set of general norms can be delineated by the nature of agreements among individuals. A basic distinction is thus possible between conventions, which are tacit rules, and institutions, which result solely from an explicit and formal agreement among individuals. Customs belong to the first subset of general social norms, defined as informal and tacitly agreed rules. Using economic tools, we shall define customs and show how their properties are those of conventional norms. Then, we shall study the boundaries within which they are supposed to be obeyed. Finally, we shall examine the conditions under which customs ground legal rules and compare them to other sources of the law.

Customs as conventional rules

The formation process of customary law has frequently been analysed from the positive perspective of the description of some particular types of societies or practices. International trade is frequently quoted as a significant example (Benson, 1998a, 1998b). Primitive (Landa, 1983; Benson, 1991) or medieval (Greif, 1989; Milgrom et al., 1990, Greif et al., 1994) societies also provide evidence that help us to understand how customary rules emerge. These various customary orders interestingly reveal how customs are conventional rules: they emerge as the result of a spontaneous and decentralized process which functions like a market process. Customs facilitate coordination in allowing the harmonization and stabilization of mutual expectations about each other's behaviour. Emerging rules have a crucial role when a player wants to assess other players' preferences. They reduce the number of required interactions to a relatively low and manageable number of instances. However, such rules are not supposed to emerge in every kind of situation. Not only does the adoption of rules of conduct require a cooperative environment because individuals have to agree to interact or to initiate interactions, but also recognition and compliance with

rules are usually unlikely if individuals engage in discrete interactions such as one-shot games.

The first condition relates to the motivations of individuals who initiate interactions in a decentralized context. The players must be assumed to display a certain willingness to cooperate and, simultaneously, they must be able to identify those individuals who are characterized by the same feature. To get acquainted with others' characteristics, sympathy is of decisive importance. Sympathy can be defined as a principle of communication which allows individuals to feel – rather than understand – the way others behave. It implies not only that individuals have access to a knowledge they have not directly experienced but also that, as long as individuals sympathize with each other, they are capable of imagining themselves as being another person: when I sympathize, 'I consider what I should suffer if I was really you, and I not only change circumstances with you, but I change persons and characters' (Smith, 1976 , p. 317). In this context, inclination towards others replaces self-interested calculation. Therefore, in a society where individuals feel sympathy with others, social order is the product of a consensus which is grounded on secured expectations. It clearly indicates that interactions do not take place in an environment of conflict that could be represented by a prisoner's dilemma, but rather take the form of a coordination game. The situation of reference for understanding the emergence of customary rules is not that of the 'war against all' that Thomas Hobbes viewed as the manifestation of the state of nature, but that of a Humean, peaceful and already structured society.

It is thus a major assumption put forward by the defenders of the spontaneous order tradition, that the *ex ante* existence of some willingness to cooperate – for instance, under the form of sympathy – is a prerequisite for the emergence of customary rules. Once this first condition is satisfied, individuals can anticipate a secured repetition of interactions. Therefore, conventional customary rules may come into effect under the second condition that interactions be repeated. The process can be summarized as follows: customs emerge from unorganized interactions when a regularity of behaviour among a fixed number of individuals occurs in a recurrent situation. When individuals face these recurrent situations, trial and elimination of errors allow the emergence of patterns of actions which are progressively incorporated into customary rules. Thus, customs endogenously but spontaneously emerge from successful practices and attempts to solve recurrent problems. From this perspective, Vanberg (1989) rightly insists on the unintended nature of customs: while no one can be identified as their explicit creator, everyone uses them and it is publicly known that everybody does so. In other words, following the definition given by Lewis (1969), customs are common knowledge rules. An organizational benefit is identified here, on the one hand, because of the limitation in the

range of circumstances everyone has to pay attention to and, on the other, because of more secured interactions. Indeed, customs will not be challenged by players because, and as long as, everyone follows such rules of action and believes that the others do the same. Thus, successful coordination does not require rules to be explicitly known and discussed but solely to exist. No formal or explicit communication between the individuals about the positive meaning of the rule or about mutual intentions is required in so far as human actions spontaneously produce their own coordination.

Such coordination tells us about the impossibility of human beings explicitly knowing the sense customs convey. Customs acknowledge the irremediable ignorance of men, which is emphasized by Friedrich von Hayek but also dates back to the theories of the Scottish founding fathers of political economy, David Hume and Adam Smith. In conflict with Cartesian rationality, the Scottish Enlightenment sensualism proposes a theory of human nature in which men perceive the world through their senses rather than constructing it by means of reason. Knowledge is then viewed as the result of accumulation and association of impressions. Men are thus shaped by what they experience. Of course, since the domain of individual experience is limited, induction plays a major role in the accumulation of knowledge by requiring participation in various repeated interactions. Therefore, the process through which customs emerge rests upon an inductive accumulation of knowledge: repeated interactions and frequent non-formal communication between individuals lead to the emergence of a common tacit knowledge.

Customs as local general norms

There are two ways in which customs possess the characteristics of general norms. First, customs can be defined as general norms in that they are not built to solve specific problems but rather are abstract rules. In a second sense, as they incorporate successful practices, customs are assumed to be generalized by becoming applicable beyond the initial boundaries within which they emerged, or by becoming relevant to instances that were not concerned in the first place. To assume the possibility of such a generalization implies considering customs as pure public goods while, in fact, they are club goods or local public goods. Indeed, the space in which customs are used depends on the distance – physical as well as psychological – covered by sympathy. Hume and Smith, among others, have insisted on the fact that sympathy is a 'scarce' feeling restricted to individuals close to each other. Therefore, customs define homogeneous clubs whose members have participated in similar repeated interactions and thus display the same willingness to cooperate with one another. This 'local conformity effect' can be compared to the 'global diversity effect', which tells us that 'there is a positive probability that [several communities that do not interact with one another] will be

using different conventions' (Peyton Young, 1996, p. 112). From this perspective, the likely tensions between the different local (be they regional or national) traditions could be an obstacle to the generalization of customs.

Two types of problem are likely to occur. First, as we have shown previously, customs develop and spread in a cooperative environment because knowledge is not acquired through formal communication but through participation in interaction. Thus, a new player entering the game will have to participate in interactions with members of the group in order to become familiar with the different local customs. Alternatively, a new player entering the game may face induction problems because he/she is unable to positively know the meaning some other person gives to the rule or to infer this meaning from the observation of behaviours. There are asymmetries of information between individuals who belong to different groups, and individuals cannot acquire knowledge of local rules except through direct participation. The costs of acquisition of information through participation, when players of different groups are involved, are thus very high. Here, the problem faced by customary codification is that of the publicity of law beyond the limits of the group.

A second type of problem concerns variations in players' willingness to cooperate. Free-riding and opportunistic behaviours may be an increasing function of the size of the group. Large groups are assumed to promote efficiency in allowing more specialized production activities and increasing returns. However, as group size expands, transaction and control costs increase in a non-linear manner, perhaps offsetting the advantages of a larger group. The greater the number of individuals involved, the more difficult it is to know the rules they follow, in so far as it requires an ever greater number of experiences, and the more difficult it is to check that they respect the rules. Therefore, within larger groups, cooperative behaviours tend to disappear, and are replaced by free-riding. A 'constitutional' use of the model proposed by Hotelling (1929) explains how distance to the local public good influences the legal strength of the rule (Blum and Dudley, 1991; Josselin and Marciano, 1999). Thus, 'within a broader social setting decentralised law making encounters the problem that some individuals will tend to free-ride on the enforcement of others' (Ogus, 1999, p. 589). In this environment, non-cooperative behaviours can emerge and persist, whether or not the game is repeated (Witt, 1989). Brennan and Buchanan (1985, p. 60) called this 'Gresham's law of social interactions'.

Customs as legal general rules

The generalization of customs faces obstacles associated with the publicity of law and with non-cooperative behaviours. To overcome these, Hobbes proposes a contractual solution: the sovereign power makes law a national public

good and at the same time is the judge. Contractual codification solves both the problem of induction in that law is publicly and explicitly declared to the people (Hobbes, 1966) and the problems of non-cooperation because sovereigns establish themselves in the position of final arbitrator. However, the very nature of customs precludes any contractarian device. A social contract demands explicit agreements, which obviously involve individual knowledge of the content of the covenant. Hume's rejection of the social contract is justified by the lack of obligation it would convey, since it is plainly founded on interest. By contrast, the enforcement mechanism associated with customary rules has to respect the spontaneous nature of the latter.

Customary systems of law meaningfully exemplify the transformation of customs into customary legal rules. The function of the corresponding institutional arrangements is to ensure that law can effectively be allowed to spread in the geographical space as well as in the space of preferences. First, rules of law are considered as an extension of customs. From this perspective, judges do not have the responsibility even less the capacity to create rules. They merely act as arbitrators whose role consists in discovering rules through individuals' actions and clarifying already existing norms. In other words, it is impossible to distinguish between *de jure* and *de facto* rules (Hayek, 1980, p. 76): customs emerge *de facto* and may later become codified (*de jure*) legal norms. Obviously, judges cannot rely upon an effectively infinite knowledge of human behaviours and cannot pride themselves on being a central planner. The required knowledge to select the best rules and, after that, to evaluate the consequences of their choices, exceed their capacities. What is identified as the major ability of judges is their experience and their ability to refer to what was previously done in the same kind of circumstances. Judges should rely on prevailing customs and progressively amended practices. This position has been defended since the very origins of the spontaneous order tradition by jurists like Sir Matthew Hale or Sir Edmund Coke. For instance, Hale prefers 'a Lawe by which a Kingdome hath been happily governed four or five hundred years than to adventure the happiness and Peace of a Kingdome upon some new Theory on my owne' (quoted in Barry, 1982, pp. 15–16). At the same time, since no one can behave as a central planner, the rule of precedent will carry out this task. The importance of the *stare decisis* principle provides a clear illustration of the role of precedent in the decisions taken by judges. Every codification decision belongs to the tradition, confirms it and reinforces it.

Second, the role of oral transmission of tacit rules is particularly well developed. Iceland in the Middle Ages provides an example of such a periodical recitation of articles of law. Another part of the judicial device which is required for customs to be legal rules lies in the fact that justice is dispensed by groups of men, usually known as *Leiroir* (autumn assemblies). Whatever

their label, these groups perform the same function, namely to gather together individual members of different communities to impart to them the rules of law. Finally, until the twelfth and thirteenth centuries, the English system of customary law exhibited many specialized courts of justice (county courts: local law; church courts: canon law; borough courts: merchant law; courts of a baronial overlord: feudal customs). This was clearly a means of dealing with a problem of induction in the space of preferences. The aim was to provide an optimal allocation of judges' abilities between the different kinds of litigation.

Customs and other sources of the law

A traditional argument against the social contract tradition relates to the fact that the monopolist provider of law is expected to adopt strategic behaviours. Influenced by interest groups, he/she is likely to use power to promote private interests. The presence of a Leviathan (Brennan and Buchanan, 1985; Josselin and Marciano, 1997) threatens the rightness and the efficiency of public decisions and questions the legitimacy of law. On the contrary, customary codification is assumed to avoid unjustified, inefficient and unjust wealth transfers. Indeed, as an arbitrator, the judge does not stand in the position of a monopolist, a central authority benefiting from coercive powers in the provision of law. In this framework, one cannot clearly identify a genuine delegation process, and the related agency relationship, through which judges would become the agents in charge of law provision. The judge is the agent of the impartial spectator of the common law, in the sense given by Smith (1976) and Hume (1992). In the process of establishing customs as legal rules, no one is explicitly in a monopoly position, that is to say capable of controlling the provision of law. It is standard argumentation in economic analysis to affirm that decentralized market processes perform better than centralized allocative mechanisms. What is undoubtedly true for many activities may not necessarily cover all the dimensions of the market for law.

First, as far as rightness is concerned, it is questionable whether spontaneous order selects better rules than a contractual process. The emergence of norms is assumed to result from cultural selection, which, according to Hayek, is of the very same nature as natural selection. However, Hayek himself suggests that 'evolution cannot be just' (1988, p. 20) in the sense that rightness is extraneous to what is an evolutionary process rather than a constructed allocation.

Second, the natural selection mechanism is also assumed to promote economic efficiency (Priest, 1987; Rubin, 1987). However, one cannot have any certainty as to either the efficiency or the rightness of the selected rules. Among others, Brennan and Buchanan remark that

> [S]ocial conventions that emerge historically and take on the status of 'unwritten rule' do not necessarily produce the best conceivable pattern of outcomes. Some modern social analysts (notably Hayek and his followers) display an apparent faith in the forces of social and cultural 'evolution' to generate efficient rules. There seems to be no reason to predict that these forces will always ensure the selection of the best rules. (1985, pp. 9–10)

Even Nozick points out that chance can be a means of selecting a rule (1974, ch. 6). Then, once selected, inferior rules become stable and widespread in spite of their flaws. Problems come from lock-in effects, inertia and path dependence (see Harnay, 2002).

In this context, as noted by Backhaus, 'the local customs can no longer be regarded as the only or main source of the law' (1999, p. 7). In other words, created rules are necessary to complement customs. Nevertheless, rule creation does not refer to a pure or strict constructivist process in which customs would be ignored, but rather sees them as at least partially relevant. Constructivism must get along with pragmatism, thereby implying a reference to existing customs. Paying attention to social norms may avoid externally imposed rules that could crowd out cooperative behaviours endogenous to the customary setting (Frey, 1994, 1997).

In this respect, the *Code Napoléon* is exemplary in that it mixes customary and contractual considerations (Josselin and Marciano, 2002). Indeed, although it may be viewed as a constructivist, Cartesian and rationalist attempt to create a new legal system, the elaboration of the *Code Napoléon* really builds upon the two sources of the law, customary and contractual, that characterize the French legal system before the 1789 Revolution. At that time, the south of France was mainly under the influence of Roman law and its personality principle; contract law was well developed. The north was dominated by Germanic tribal customs. Drafted in four months under the close scrutiny of Napoleon himself, by François Tronchet, Felix Bigot (northern jurists from the *Tribunal de Cassation*) and by Jacques Maleville and Jean Portalis (southern jurists' respectively, from the *Tribunal de Cassation* and the maritime courts), the French civil code draws heavily on previous and pre-revolutionary works by Jean Domat (civil law in natural order) and Robert Pothier (general principles of *l'ancien droit*). Domat and Pothier had remarkably surveyed and formalized the pre-revolutionary system of law, hence the continuity from the old regime to the new one. Codified customs in the field of property law have been largely maintained while contract law still rests heavily on Roman law. What is outstanding in this process is both this continuity and the new role of the state in the control of law provision. Monopolization is achieved but without breaking down the old pillars of Germanic customs and Roman contracts.

Customary rules characterize a spontaneous order process in which no state intervention is required to back up the social order. In theory, customary

codification thus fundamentally differs from contractual codification. Conversely, the social contract should rule out customs and build the legal system from a *tabula rasa*. However, things are less clear-cut when history mixes practice and reasoned arguments.

References

Backhaus, Jürgen (1999), 'The German Civil Code of 1896: an economic interpretation', *European Journal of Law and Economics*, **7**, 5–14.

Barry, Norman (1982), 'The tradition of spontaneous order', *Literature of Liberty*, **5**, 7–58.

Benson, Bruce (1991), 'An evolutionary contractarian view of primitive law: the institutions and incentives arising under Indian law', *Review of Austrian Economics*, **5**, 41–65.

Benson, Bruce (1998a), 'Evolution of commercial law', in Peter Newman (ed.), *The New Palgrave Dictionary of Economics and the Law*, London: Macmillan, pp. 89–92.

Benson, Bruce (1998b), 'Law merchant', in Peter Newman (ed.), *The New Palgrave Dictionary of Economics and the Law*, London: Macmillan, pp. 500–508.

Blum, Ulrich and Leonard Dudley (1991), 'A spatial model of the state', *Journal of Institutional and Theoretical Economics*, **147**, 312–36.

Brennan, Geoffrey and James M. Buchanan (1985), *The Reason of Rules*, Cambridge: Cambridge University Press.

Frey, Bruno (1994), 'How intrinsic motivation is crowded out and in', *Rationality and Society*, **6**, 334–52.

Frey, Bruno (1997), 'A constitution for knaves crowds out civic virtues', *Economic Journal*, **107**, 1043–53.

Greif, Avner (1989), 'Reputation and coalitions in medieval trade: evidence on the Maghribi traders', *Journal of Economic History*, **49**, 857–82.

Greif, Avner, Paul Milgrom and Barry R. Weingast (1994), 'Coordination, commitment, and enforcement: the case of the merchant guild', *Journal of Political Economy*, **102**, 745–76.

Harnay, Sophie (2002), 'Was Napoleon a benevolent dictator? An economic justification of codification', *European Journal of Law and Economics*, **14**, 237–51.

Hayek, Friedrich August von (1980), *Droit, Législation et Liberté*, vol. I, Paris: Presses Universitaire de France (*Law, Legislation and Liberty*: first published 1973).

Hayek, Friedrich August von (1988), *The Fatal Conceit: The Errors of Socialism*, Chicago: Chicago University Press.

Hobbes, Thomas (1966), *A Dialogue Between a Philosopher and A Student of Common Law in England*, Paris: Dalloz.

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

Hume, David (1992), *Treatise of Human Nature*, Buffalo, NY: Prometheus (first published 1739–40).

Josselin, Jean-Michel and Alain Marciano (1997), 'The paradox of leviathan. How to develop and contain the future European state?', *European Journal of Law and Economics*, **4**, 5–21.

Josselin, Jean-Michel and Alain Marciano (1999), 'Unitary states and peripheral regions: a model of heterogeneous spatial clubs', *International Review of Law and Economics*, **19**, 501–11.

Josselin, Jean-Michel and Alain Marciano (2002), 'The making of the French civil code: an economic interpretation', *European Journal of Law and Economics*, **14**, 193–203.

Landa, Janet (1983), 'A theory of the ethnically homogenous Middleman group: an institutional alternative to contract law', *Journal of Legal Studies*, **10**, 349–61.

Lewis, David (1969), *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.

Milgrom, Paul R., Douglass C. North and Barry R. Weingast (1990), 'The role of institutions in the revival of trade; the medieval merchant law, private judges and the champagne fairs', *Economics and Politics*, **2**, 1–23.

North, Douglass (1991), 'Institutions', *Journal of Economic Perspectives*, **5**, 97–112.

Nozick, Robert (1974), *Anarchy, State and Utopia*, New York: Basic Books.

Ogus, Anthony (1999), 'Self regulation', in Baudouin Bouckaert and Gerrit de Geest (eds), *Encyclopedia of Law and Economics*, Vol. V, *The Economics of Crime and Litigation*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.
Peyton Young, Harvey (1996), 'The economics of conventions', *Journal of Economic Perspectives*, **10**, 105–22.
Priest, George (1987), 'The common law process and the selection of efficient rules', *Journal of Legal Studies*, **6**, 65–75.
Rubin, Paul (1987), 'Why is the common law efficient?', *Journal of Legal Studies*, **6**, 51–64.
Smith, Adam (1976), *The Theory of Moral Sentiments*, Indianapolis, IN: Liberty Press (first published 1759).
Vanberg, Victor (1989), 'Carl Menger's evolutionary and John R. Commons' collective action approach to institutions', *Review of Political Economy*, **1**, 334–60.
Witt, Ulrich (1989), 'The evolution of economic institutions as a propagation process', *Public Choice*, **62**, 155–72.

29 Science as a source of law

Peter R. Senn

So ubiquitous is science as a source of the law that few general statements are possible. Time, place and circumstance determine the role of science in the development of law. For example, the ancient Greeks did much to give meaning to modern ideas of both science and law. A good case can be made that they used their conceptions of science in the development of their laws. Despite this, it would be hard to support the claim that what they meant by both terms, law and science, then are broadly applicable to many of the issues of today.

Understanding the role that science plays as a source of law must depend on the meanings given to both 'law' and 'science'. Both terms are used in so many different senses that their denotation must always be specified. As used here, the term 'source' refers to the knowledge which science provides as the basis for changing or developing law. There is little agreement among scholars about precise definitions of 'law'.[1] Extreme caution in the use of the term 'law' is also required because in every language it has many connotations. Scholars from different disciplines use it to mean very different things. Legal scholars, for example, mean something quite different from philosophers or scientists when they use the same word. There are also many kinds of law. Among these are public, civil, natural, canon, divine, criminal, international and commercial law. Science has different influences on the development of each of these.

In the most elementary terms, the law consists of writings on a piece of paper. No serious discussion can begin with such a simple approach because law is a social institution. The law as expressed by words on paper is like the written words and music of a song. The words and notes only exist to provide a record of what the composer wrote. The hearing of a song bears about the same relationship to its written record as the words in the law have to its practice.

Some relationships of law and science

As used here, the term 'law' refers to rules that most of the people it is intended to encompass regard as covering all its members. These rules are accompanied by institutions and procedures for their interpretation and application and, in the case of disobedience, procedures and sanctions for enforcement.

Law is part of a social system that is unique in many respects for each country. Its relationships to science can only be understood in the context of the legal system for a given country or the developing international systems of law. By a legal system is meant a set of social organizations: legislatures, courts, the bar (lawyers and legal scholars), regulatory agencies and police that work together. Legislators originate the formal statement of the law. Judges interpret it and determine sanctions. Lawyers and legal scholars practise (study, interpret, manipulate) it. Regulators, bureaucrats and police administer it. Inevitably, they also interpret and manipulate it, but on a level different from that of the courts.

For each of the social organizations that make up a legal system, science plays a different role in the development of the law. As used here, the term 'science' denotes what passes for verifiable knowledge. It plays an important role in each of the social institutions that make up the legal system. The term is not used to stand for any kind of knowledge. In that usage, all law is based on it.

Science has several dimensions, each of which has a part in a system for producing verifiable knowledge. The goal of those who practise it is verifiable knowledge. Science requires argument and evidence that anyone can accept, one of the characteristics that makes it universal. In the long run, it is self-correcting. This is a result of its openness, accessibility and universality. Other investigators will, sooner or later, root out any errors.

Law is a social enterprise but of a type different from science. The goal of law is justice. The idea of justice differs according to culture, time and place. There has never been widespread agreement on what justice is. Many famous writers on law and economics have written about economic justice. For example, Gustav Schmoller (1838–1917) treated justice as an empirical concept. He felt that ideas of justice develop from the culture while 'the law can only uphold justice within its own range and can only execute it in a certain sense' (Schmoller, 1894, p. 725). Friedrich August von Hayek (1899–1992) treats the idea of social justice as a mirage (see Hayek, 1973, vol. 2, which contains a disparaging reference to Schmoller and his ideas of justice).

One of the most important functions of a legal system is to settle differences about what justice is in individual cases. In sharp contrast to science, argument and evidence in the legal system only need to satisfy the person or body before which they are given. No legal systems claim universality. All the sciences are bound by the same canons, which include a search for a closer approach to truth; honesty; verification procedures such as replication and falsification; and logical consistency. Different legal systems have very different standards for each of these. Torture or the use of electronic equipment or drugs as truth-finding devices are acceptable in some legal systems but not in the scientific world.

Scientists are also united in their agreement on many of the same assumptions, such as the existence of a physical and human world apart from that known to the individual through his/her senses and intellect; the idea of progress in the sense of a cumulative building of knowledge; causation, in the sense that things happen for reasons; and the reality of space and time.

Actors in the legal system make some very different assumptions. Most notably, they have to assume the ultimate reality of the world as it is known to the individual through the senses and intellect. In the law, motivation, which has little place in science, is often an important causal consideration.

Many elements of the methodology of law and science are the same, as with, for example, the use of different kinds of explanation, reliance on observation, tests of logical consistency and the use of mathematics. Scientific knowledge is, however, separated from other ways of knowing common in law. Unlike actors on the legal scene, scientists refuse to accept things as 'known' on the basis of authority, intuition, insight, hope, logic alone, personal behaviour or philosophy.

Another way to understand some of the differences between science and law is by way of the fact that different languages can make statements about the law which cannot be translated. But any scientific knowledge can be both stated and scrutinized in any modern language, using the same methods with the same results. To fully understand the differences between science and law, one has only to recognize that both the products and processes of each are quite different. The product of science is verifiable knowledge. The most obvious product of law is constitutionality. At its best, law ought to produce justice or at least equity and fairness. Science and law are so different and so important that investigations of science as a source of law are appropriate. Equally legitimate might be investigations of the law as a source of science, a subject not touched on here.

Direct and indirect influences

The processes through which science influences law are many. They depend on the kind of law being considered and the part of the legal system which is affected. The processes are both direct and indirect. Direct influences occur when science can be shown to have some use by, or effect on, actors in the legal system. Indirect influences occur when the use of science results in technology or social change which, in turn, can be shown to have some effect on actors in the legal system. Often direct and indirect influences work together.

An example of a direct process in which science is a source of law is the growth of knowledge which displaces what had been regarded as correct. A well-known historical example is that of witchcraft. As scientific knowledge grew, the laws about it changed. Another example is the changes in the canon

law of the Roman Catholic Church as the result of the victory of Galileo's assertion that the earth moves.

Many examples of the indirect processes through which science is a source of law can be found in the growth of technology. The most certain knowledge of science is embodied in technology. The reason is that successful technology must pass the ultimate empirical test – it must work. For at least the last hundred years, the growth of technology has been the main driving force by means of which verifiable knowledge has become a source of law. The process is almost always indirect, as a few examples will illustrate. The development of railroads required changes in numerous laws and the development of many new laws. Among these were laws of eminent domain, liability and finance. The changes brought about by industrialization gave rise not only to new laws about the tools that made it but also to the relationships between the government, employer and workers. Child labour laws are but one example.

The development of computers provides a dramatic example of the translation of scientific theory into a workable tool used by almost all the actors on the legal scene. In general terms, the process is this. Using the embodiment of verifiable knowledge in the form of computers, new forms and methods of communication develop, data bases become larger and access to them faster. They, in turn, facilitate the development of entirely new procedures, products, organizations and institutions. The proliferation of financial products called 'derivatives' is an example of this process. Again many new laws and changes in the old ones are required. The development of technology forces changes in the law by its creation of new procedures, products, institutions and relationships. As this happens, the legal system must accommodate the changes if it is to remain viable and the economic system is to grow.

There are no studies to tell us which process, direct or indirect, is more important in making science a source of law. Arbitrarily, examples of science as a direct source of law are given first. The main components of a legal system are legislatures, courts, the bar, regulatory agencies, police and other bureaucracies. Members of each often change their activities in the light of the knowledge science provides. Every actor on the legal scene is affected by such things as the development of new ways to gain access to and utilize knowledge such as the World Wide Web, the communications network based on computers.

Legislators must deal with the world of science and the often conflicting demands of politics. They commonly call upon scientists for their views about both the methods and the findings of science when drawing up laws. This is especially the case for technical issues about which the average person knows little. They must make laws about such things as cyberspace. Science, used in this way, is a source of law. The same is true for courts

where the use of experts is common. Judges in their reasoning sometimes use some of the theories and methods of science; occasionally, social science is used to provide context. They must also cope with new situations brought about by scientific advances, as for example computer-generated evidence.

Lawyers and legal scholars use the knowledge of science for many of the same purposes and in the same ways as legislators and judges. Lawyers, however, have a very different goal driving their uses, the victory of their client. In the United States, social science is used to plan both systems of litigation and the litigation of a case.[2] Legal scholars, who play a more important role in code law systems than they do in common law systems, often study law with scientific methods.

One of the most striking and consequential characteristics of modern legal systems is the inability of legislators to keep up with all the needed changes in the laws. This has resulted in the delegation of legislative duties to a large number of administrative organs of many varieties. These agencies make law in the form of regulations. The bureaucrats in these agencies also have the power to interpret the regulations and to determine procedures and sanctions for their enforcement, usually subject to an appeals procedure. A frequent complaint of administrators is that they are not provided with adequate funds for enforcement.

Many administrative agencies often directly use the findings of science in their work. Some governmental departments are absolutely dependent on the work of scientists, for example those in charge of regulating energy, telecommunications, the environment and drugs. Regulations without a solid scientific basis would be useless or harmful to those fields. It should go without saying that their uses of science are not always complete or impartial. Many other regulatory agencies, for example those in charge of health and welfare, often use the findings of science as inputs to the regulations and procedures they promulgate. It is hard to imagine any part of government that does not use the findings or results of science in some way.

For the common man, the police are the most visible sign of the law. They are the most immediately accessible interpreters of disobedience and appliers of sanctions for enforcement. They use both the methods and findings of science, every day and in many ways. Examples range from radar detection of speed through data collection and analysis to the most modern methods of communication.

In many cases, the indirect ways in which science becomes a source of law is more significant than the direct paths. One of the most important reasons forcing legislators to delegate their law-making powers to administrative agencies is that new technology makes old law obsolete and often unworkable. An example is the growth of the World Wide Web. Judges find that they must reason about problems never before encountered, as in space law. Court-

room conduct and procedures must be modified when modern technology such as television or other recording devices is used or introduced as evidence. At a step removed in terms of indirect influence is the spread of government, which has required the development of new systems of courts and laws, for example those for taxes, social security and commerce.

Lawyers and legal scholars often cannot rely on precedent for answers to many of their problems because of the growth of technology. Even their day-to-day practice is influenced by such things as the facsimile transmission (fax) machines and the use of computers. Both allow rapid access to, and manipulation of, amounts of data inconceivable before their use. Many administrative agencies constantly find themselves faced with new problems because of the growth of technology. For example, those in charge of health and welfare have had to cope with the growth of 'carpal tunnel syndrome' as the result of the increasing use of computers. Another is the problem of worker displacement caused by the development of tools like the fax machine or telephone answering systems which make workers' tasks more manageable. These often allow fewer people to do more work.

Many examples could be given of the way the police are affected by the growth of science. Sometimes their burdens are increased, as when cheap methods for making new addictive drugs become widespread. They need new ways to test for them. The new drugs often require new laws for their control. When their illegal sale is profitable, entirely new laws are required to cope with money laundering. Often new technology helps the police to do their job, as is the case with many of the improvements in communication and data handling.

Misuses of science as a source of law

Science is not an entirely benevolent source of the law. The knowledge science provides often forces changes in the law against the resistance of tradition. Sometimes science is neglected, sometimes misunderstood, often misused. There are many reasons for this. A legal system is a social institution. The actors on the legal scene have a common humanity. The inertia of tradition always weighs heavily. Characteristics such as those of selfishness, narcissism, sectionalism and parochialism are often found. Fear of the untried, devotion to special theories, lack of knowledge, misunderstandings, religion, dogma and ideologies are occasional obstacles to the use of science in the legal system.

Sometimes ethical and moral values are also felt by some to be antithetical to science. Victor R. Fuchs (1996, p. 1) has stressed the role of value differences 'as a major barrier to effective policy making' in the field of health economics. Much modern science requires some mathematical skills to understand it. These skills do not appear to be common.

Not to be ignored are motives. Knowledge is like a knife. It can be used for good or evil. The list of reasons for science not being used or misused in legal systems could be lengthened. It need not be if it is recognized that human fallibilities must always be taken into account. Some misuse and neglect of science is inevitable in a democracy. The legislature is the place where the formal statement of the law is developed. The politicians who make up the legislature must, if they are to remain in politics, give most of their attention to what they think will get them re-elected. The knowledge science provides has to take second place to putative political realities.

This is especially the case when social issues are involved. So insecure is the scientific knowledge about them, for example the causes of unemployment, that differing views can almost always be found. Even in cases where there is little or no dispute among scientists, the uses of knowledge, as in the case of atomic energy, are almost always subject to debate. Sometimes political pressures, usually from situations that are perceived as emergencies, require the passage of laws without the benefit of any scientific input.

Science influences law differently in different legal systems

It is certain that the influence of science as a source of the law is different in legal systems of different kinds. The *Shari'a* is the basic Islamic law founded on the Quran and *Sunnah*. The *Sunnah* comprises the traditions of the Prophet Mohammed. The other kind of law is the derived or substantive law, *Fiqh*, which comprises the temporal legislation. Most Islamic scholars would reject the idea that science played any substantial role as a source of *Shari'a*. Most Jewish scholars would also reject the contention that the foundations of Jewish religious law owe much to any common conception of science.

The laws of many African countries come from at least four very different sources of law. Each source, the indigenous legal institutions sometimes called 'native' or 'tribal', the laws of religious origin, for example those of Islam and the Hindus, the legal institutions that were introduced by the colonial powers and the legal systems developed after the various African states became independent, has a very different relationship to science.

It is very possible that the influence of science as a source of law is somewhat different in countries with code law systems and countries with common law systems. These two systems are the most important in the world today. The comments here are intended to apply to both systems on the assumption that the similarities are more important than the relatively minor differences for this topic.

The social context influences science as a source of the law

All modern legal systems have survived any number of dynamic political situations. The same law has taken on quite different meanings, in terms of its

interpretation and administration, in different political contexts. In general, changes in the law can be seen as responses to persistent, fundamental problems of social and political organization. The specific changes are selections made from a range of possible alternatives. In its best uses, the verifiable knowledge provided by science narrows the range of possible alternatives and conditions the final choice.

Although there is debate among historians of science about the importance of the social and political context in its development, the knowledge that science produces does not change in response to social and political contexts. The uses, the interpretation, the awareness and the internalization of knowledge by actors on the legal scene may change, but what is verifiable does not change, except by scientific consensus. This does not mean that knowledge does not change. It does. What is known is only a small part of what is potentially knowable. Thus there is no contradiction when it is said that verifiable knowledge does not change, while knowledge itself changes constantly.

In recent years economics, one of the social sciences, has been the source of many significant legal changes. Although the knowledge provided by economists has long been a source of law, both legal scholars and economists have increasingly contributed to the development of the separate discipline of what is now called 'law and economics'. The development of both the law and science guarantees that each will influence the other. In the future, just as in the past, science, because of its universality, is bound to influence the law more than the other way around.

The subject of science as a source of the law has not been systematically studied. The literature is vast and scattered. Every part of the modern legal system has a literature covering some of its relationships to science. Most of this literature is concerned with the application of scientific findings to parts of the legal system. For an example, see John A. Tarantino (1988). The classic bibliography, although dated, is that by Morris L. Cohen et al., still the best of its kind. The 'Introductory essay', by Cohen, remains the foremost overview of the general field. The Internet has many resources, for example see 'The seamless website – law and legal resources' at http//seamless.com/commons.html (May 1997). See also 'Social science as a source of the law' Chapter 30 in the present volume.

Notes

1. Lisa J. McIntyre, to give just one example, provides a list of 11 different definitions chosen from a sociological perspective.
2. For examples of these uses, see the traditional law school casebook edited by John Monahan and Laurens Walker (1994).

References

Cohen, Morris L., Naomi Ronen and Jan Stepan (1980), *Law & Science: A Selected Bibliography*, ed. Vivien B. Shelanski and Marcel C. La Follette, Cambridge, MA: MIT Press.

Fuchs, Victor R. (1996), 'Economics, values and health care reform', *American Economic Review*, **86** (1), March, 1–24.

Hayek, F.A. von (1973), *Law Legislation and Liberty: A New Statement of the Liberal Principles of Justice and Political Economy*, 3 vols, Chicago: University of Chicago Press.

McIntyre, Lisa J. (1994), *Law in the Sociological Enterprise: A Reconstruction*, Boulder, CO, San Francisco, CA and Oxford: Westview Press.

Monahan, John and Laurens Walker (eds) (1994), *Social Science in Law: Cases and Materials*, 3rd edn, Westbury, NY: Foundation Press.

Schmoller, Gustav (1894), 'The idea of justice in political economy', *Annals of the American Academy of Political and Social Science*, Vol. 4, March, 697–737; trans. Ernst L. von Halle and Carl L. Schutz.

Tarantino, John A. (1988), *Strategic Use of Scientific Evidence*, New York: Kluwer Law Books.

30 Social science as a source of the law

Peter R. Senn

The social sciences are, and long have been, a source of the law. The processes involved are unlike those of the sciences that study things or animals. Many of the same qualities that separate the different sciences lie behind the differences in processes. The most important of these scientific differences stem from their subjects and, secondarily, their methods. The roles of technology, indeed its definition, are special for each kind of science.

Social science studies human interactions. The other sciences study things or animals. For these sciences, the ultimate test of truth comes from the technology which is developed from the knowledge the science provides. If the knowledge can be shown to work, it is judged to be correct. If the knowledge does not survive empirical verification, something about it is thought to be lacking. The case is very different for the study of mankind. There is no ultimate test of truth. The time, the place, the circumstances and, above all, man's free will determine the outcome of events. Knowledge that stands up to empirical verification in one situation does not do so in another.

Social scientific knowledge is contingent in ways that the knowledge of things is not. Even cultural universals, such as music and exchange, change with time and are socially conditioned. Few social science generalizations are verifiable for all times and all cultures. Those that are universally true, such as 'all men die', are attributes more of the physical body rather than of behaviour. This does not deny the possibility, indeed, the necessity for the scientific study of men. The necessity stems from the fact that men, having free will, seek order and change. They have ideals and try to live by them. They try to improve and believe improvement is possible. What is more natural than to apply methods that are known to produce verifiable knowledge in the pursuit of these aims?

The kind of technology that is developed in the social sciences is a set of instructions for doing things. Although these instructions can be written down, they are not embodied in material things in the same way as is the knowledge developed by the other sciences. Often, the results of this kind of technology are embodied in techniques, social organizations and other linguistic and analytical tools. The development of computers provides a dramatic example of both kinds of technology. The physical parts reflect the translation of scientific theory into a workable tool. But they would not work without programs – the detailed list of instructions which manage the electri-

cal changes. These are the written reflection of ideas or techniques. They are the embodiment of ideas.

Popular expressions such as 'the knowledge revolution' refer to this kind of technology. Knowledge is embodied in techniques ranging from bowing for stringed instruments and tonguing for wind instruments to organization manuals for business and government, data bases, drug formulas, computer programs and much else often referred to as 'intellectual capital'. This kind of technology provides for the combination of limited resources in ways that produce ever more value.

The social sciences have a long history as a source of law

Today the disciplines of the social sciences are generally taken to be anthropology, economics, geography, history, political science, psychology, sociology and their applications: education, planning, public administration and social work. But it took a long time to arrive at the conclusion that people in their social roles could be studied scientifically.

By the late 1700s, the following disciplines now considered social sciences were reasonably well established: political science, economics, history and geography. Men had studied these subjects from time immemorial. They used what they took for knowledge about them in the development of their law. In the late 1700s, a great change occurred in Europe and America. The term 'social science' appeared. It was an accompaniment of new and changing goals that went along with an expansion of intellectual horizons. Much of mankind changed its authorities. Those that did felt emancipated. The American and French revolutions were only two of the many results of the new ideas about how people should live. It is especially significant that the term 'social science' almost always appears before 1800 in connection with education.

Most people and their leaders believed in progress. They thought that progress could be attained primarily through more and better education. They thought it possible that the mighty advances of physics, chemistry and the other physical and biological sciences made by the early nineteenth century could be utilized for the study of mankind. The cleavages between science and philosophy and those between philosophy and theology grew ever wider, until a thoroughly secularized knowledge system separated them.

The history of modern social science and the development of law are inexorably related. Better understanding of these relationships will almost certainly improve the theoretical basis of the connections between theory and practice in the social sciences. Study of them requires inquiry into the roots of our present-day concepts and hypotheses. It cannot be emphasized too often that most of the terms and ideas in use today in both law and social science evolved long before the modern development of either.

Modern social science and modern law

There are abundant examples of social science influencing law from the time when the idea of social science appeared. Many early economists were active and successful in their attempts to influence public policy. In England, the landed interests succeeded in passing the Corn Law of 1815. This law prohibited the importation of wheat until its domestic price was very high. Many important economists actively opposed it. For example, John Bright (1811–99) and Richard Cobden (1804–65) were active forces in the Anti-Corn Law League. It was primarily economic arguments for free trade that resulted in the repeal of these restrictive laws.

In Germany, important economists like Gustav von Schmoller (1838–1917) were major figures in the organization and operation of the Association for Social Policy (*Verein für Sozialpolitik*) organized for the express purpose of influencing public opinion. It sponsored and published many studies that were explicitly aimed at changing the law on what were then regarded as social problems. German economic thinking was at the root of many of the legal changes – ranging from state ownership of the railroads to social insurance – that were enacted in the latter half of the nineteenth century.

After the revolution in France, history and sociology were but two of the social sciences which also directly influenced the law. To cite just one example, François Pierre Gillaume Guizot (1787–1874) was a professor, historian and statesman. First, minister of home affairs, then minister of public instruction from 1832 to 1837, he became head of the government from 1841 to 1848. His interpretations of social science directly influenced much French legislation.

In the United States, all of the early leading statesmen based many of their ideas about the institutions they founded on their understanding of history and economics. Benjamin Franklin (1706–90), Alexander Hamilton (1757–1804) and Thomas Jefferson (1743–1826) are three who also have a place in the history of economic thought.

The Brandeis brief: a forerunner

In 1908, Louis Dembitz Brandeis (1856–1941) submitted a brief to the Supreme Court of the United States in support of the constitutionality of an Oregon law which limited the working day to ten hours for women. It was 112 pages long and had only two pages of discussion of legal issues. The rest was economic and social data. The sources were a wide variety of material from factory and other public commissioners, economists and statisticians, medical reports and psychological treatises.

In the words of John W. Johnson, 'The Brandeis brief was unprecedented' (Johnson, 1992, p. 85). It set out to 'prove' that it was 'reasonable' that the Oregon law met its stated objective of improving public health. Up to this

time, the majority of American judges had limited their views of what was reasonable to narrow legalities. The brief was successful in persuading judges to take into account the kinds of materials from the social sciences that legislators used when they drafted laws. This is generally considered the first time the Supreme Court recognized the need for facts to establish the reasonableness or unreasonableness of social legislation. In this, the United States was behind some other countries. German courts, for example, had recognized the importance of social facts earlier. As is the case in American legal proceedings, the brief was highly selective and one-sided. It was entirely consistent with the development of what came to be called 'sociological jurisprudence'. Increasingly, legal systems throughout the world now take the findings of social science into consideration.

In recent years one of the social sciences, economics, has been the source of many significant legal changes. Although the knowledge provided by economists has long been a source of law, both legal scholars and economists have contributed to the development of the separate discipline that is now called 'law and economics'. Jürgen Backhaus (1990, 1992) has examined some of its contributions. His findings support those of many other studies: economic analysis is regularly taken into account by legislators, regulators and the courts.

More significantly, Backhaus has shown that in many situations the methods and results of economic theory have been developed into legal principles. In an important study ('The German Waterpenny case', 1997) he describes the workings of the German constitutional court. Just as the Brandeis brief set a precedent in the United States, so the theory of functions used by the German court allows it to accept almost any kind of scientific argument. In brief, the theory of functions requires that the basic function of the institution in question be not impaired.

At least for the United States and Germany, some of the indisputable results of law and economics research have become legally binding. Among the many fields in which this has happened are those where laws require risk–benefit analysis and environmental impact statements. It is not uncommon for legal principles based on economic theory to be invoked, as for example in cases where it can be shown that the economic effects of an action result in a violation of the intended effects of a law.

The processes by which social science becomes a source of law

The process by which the social sciences influence the law begins with the declaration of a social problem.

A social problem is:

(1) a social situation
(2) involving a substantial number of persons

(3) which is declared
(4) to be inconsistent with the values
(5) of some influential group
(6) and labeled as a situation
(7) which must be dealt with
(8) by some kind of collective action. (Senn and Senn, 1993, p. 236)

Social problems have many different origins. They always begin with ethical and moral values. The process by which the sciences that study man influence the law starts with the declaration of a social problem based on a judgment that a social situation is inconsistent with the values of some influential group and must be dealt with by some kind of collective action. Arguments for the remedies sought are buttressed by what passes for social science knowledge at the time.

The process by which the sciences that study things influence the law begins with the verifiable knowledge all good science produces. The fact that the knowledge works is translated into new or improved technology. The development of technology forces changes in the law by creating new products, institutions and relationships. If it is to remain viable and the economic system is to grow, the legal system must accommodate the changes.

The process by which the sciences that study man influence the law are not the same as those of the sciences that studies things. A social problem can be the result of either of the two main kinds of technological change – those stemming from knowledge embodied in material things or those stemming from the knowledge embodied in techniques. The contrasts between the two processes can be illustrated by the different kinds of issues which arose in the nineteenth century from the development of railroads and opposition to the use of child labour in factories.

When the use of the labour of children became recognized as a social problem, laws in the industrialized countries were passed to regulate it. Typically, they set limits on the hours that could be worked for children of different age groups. Arguments from many of the social sciences were influential in the debates which preceded their passage. The initial impetus for their passage was based primarily on moral and ethical grounds.

The evolution of most laws to regulate railroads stemmed more directly from the technological advances they represented. As railroads developed, they required not only changes in numerous laws but also many new laws. Among these were laws about eminent domain, liability, ownership and finance. The initial impetus for their passage was based on legal voids that needed filling if the benefits from railroads were to be obtained. It often happens that developments in material technology cause social problems. The process of industrialization not only gave rise to new laws about the tools that made it, but also brought about other social changes. Among those were

new relationships in the family and between the government, employers and workers. All of these social developments were reflected in laws that were based on social science.

The recognition of social problems is a consequence of ethical and value judgements. Practical political considerations far outweigh any kind of logic in their treatment. This is in sharp contrast to social science, which is essentially logical and empirical. The nexus between its concepts is a logical nexus. The grounds for its detailed assertions are logical and empirical.

The legal system is a social institution with logics of a very different nature. Neither the politics of social problems nor the workings of a legal system can be rationally justified in the same way as the theories internal to the disciplines. In a democracy, the formal concerns that motivate political choices must always be more important than the knowledge social science provides.

No legislature will pass new laws simply on the basis that a social situation is 'bad'. Other arguments are required, and social science provides them. The complexities of changing societies, the way political institutions and legal systems function and the contingent nature of much social science knowledge are the main reasons for the less than successful resolution of many social problems.

The influence of social science on the law is context dependent

Both law and social science are the result of social processes. Complex interactions between them are inevitable. These interactions are not well understood. Generalizations about them must be more of an observational nature than soundly based in metaphysics, philosophy or robust empirical findings. There are many kinds of law. Commonly distinguished types of law are public, civil, criminal, commercial, international, natural, canon and divine. Social science has had and continues to have different influences on the development of each.

A legal system is a set of social organizations – legislatures, courts, the bar (lawyers and legal scholars), regulatory agencies and police – that are all supposed to work together. For each of them social science plays a different role in the development of the law. Legal systems are unique in many respects for each country. The way the components of a legal system are influenced by social science can only be fully understood in the context of a given country at a given time.

The processes through which social science influences law are many. They depend on the kind of law being considered and the part of the legal system which is affected. The processes are both direct and indirect. Direct influences occur when social science can be shown to have some use by, or effect on, actors in the legal system. An example of this is the famous case of *Brown* v.

Board of Education of Topeka (347 U.S. 483, 1954). The Supreme Court changed the law when it decided that racial discrimination in the public schools was illegal. Not only did the judges make new law, but they supported their decision with the social science research of Gunnar Myrdal (1898–1987). Like the Brandeis brief, it is an example of an instance in which social science is a direct source of law. It also confirms Backhaus's theory that social science knowledge or methodology has been developed into legal principles.

Indirect influences occur when the knowledge of social science results in techniques which, in turn, can be shown to have some effect on actors in the legal system. Often direct and indirect influences are at work at the same time. Many examples of the indirect processes through which social science is a source of law can be found in the growth of the kind of technology which results in new techniques. One example is the increasing attention given to laws governing intellectual property rights and reputations. Social science is also an important indirect source of the law when the growth of the knowledge it produces causes changes, as for example in the treatment of mental problems.

When knowledge causes change, judges, lawyers and legal scholars cannot rely on precedent for answers. How are property rights to be assigned to ideas embodied in a quasi-public good which has value because of the ideas behind it? An example is a new form of business organization. There are no studies that tell us which process, direct or indirect, is more important in using social science as a source of law.

It is very possible that the role of social science as a source of law is somewhat different in countries with code law systems and countries with common law systems. These two systems are the most important in the world today. The comments here are intended to apply to both systems on the assumption that the similarities are more important than the relatively minor differences for this topic.

As previously mentioned, most Islamic scholars would reject the idea that social science played any substantial role as a source of *Shari'a*, and most Jewish scholars would also reject the contention that the foundations of Jewish religious law owe much to social science. However, both religions are increasingly utilizing it, although in very different ways, as they adjust to changing conditions. The laws of China, India and Japan are all influenced by social science in very different ways.

The uses of social science in a legal system

Each of the main components of a legal system – legislatures, courts, the bar, the police and regulatory and other government bureaucracies – uses knowledge from the social sciences. Every actor on the legal scene sometimes

changes in the light which social science provides. Legislators use both the methods and the findings of social science when they consider legal changes. They often call upon social scientists for their views. The same is true for courts where the use of experts is also frequent. Judges in their reasoning often use the theories and methods of social science.

Legal scholars and lawyers use the methods and knowledge of social science for many of the same purposes and, often, in the same ways as legislators and judges. Legal scholars, who have a larger role in code law systems than they do in common law systems, often use the methods and knowledge of social science. Lawyers frequently use the methods and knowledge of social science, sometimes in controversial ways. Because their goal is the victory of their client, the knowledge from the social sciences they use in arguing their cases is often less than impartial. They regularly employ social scientists who provide testimony favourable to their cause. It is not uncommon to find two equally qualified social science experts on opposite sides. This is one result of the contingent nature of social science knowledge. In the United States, lawyers also use social science to plan both systems of litigation and the litigation of a case. For examples of these uses, see the traditional law school casebook edited by John Monahan and Laurens Walker (1994).

One controversial aspect of the use of social science is the employment of social scientists who specialize in the process of jury selection. These specialists have the job of choosing jurors who are most likely to render the verdict desired. Another controversy concerns the use of mock juries chosen to have the same characteristics as the jury trying the case. Arguments are then tested on the mock jury before they are used in court.

The police are the most visible sign of the law for most people. They often must interpret behaviour and apply sanctions for enforcement. They use both the methods and findings of social science in many ways. Examples range from constructing psychological profiles of criminals to the use of the findings of anthropology and sociology in their understanding of the behaviour and treatment of different cultural groups. Increasing traffic in drugs has required police to learn the economics of money laundering. They must also try to satisfy those who support them with constrained budgets.

The same is true of regulatory and other government bureaucracies. All face the fundamental economic issue of attempting to do their jobs with limited resources. The inability of legislators to keep up with all the needed changes in the laws and the administration of those they have passed has resulted in the growth of bureaucracies. The delegation of legislative duties to a large number of administrative organs of many varieties has had the consequence that many bureaucracies not only administer laws but make them in

the form of regulations. Bureaucrats in these agencies have the power not only to interpret the laws but also to make regulations and to determine procedures and sanctions.

Many administrative agencies are absolutely dependent on the work of social scientists. Perhaps the most important of these are the agencies in charge of financial matters. Regulations without a solid basis in economic theory would be worthless. Much the same can be said for the government departments in charge of regulating energy, telecommunications and the environment. Many other agencies, for example those in charge of education and health and welfare, often use the findings of social science as the basis for the regulations and procedures they decree. Both legislatures and administrative agencies find themselves faced with new problems, because the growth of technology has social consequences. For example, the problems of worker displacement and unemployment cannot be understood without the help of economics. It is hard to imagine any part of the legal system that does not use social science in some way.

Social science, the law and public policy

How effective have laws based on social science been in remedying social problems? Most of the influence of social science on the law is to be found in laws and regulations about public policy. The problems of the social uses of social science begin with the definition of what a social problem is. Anyone can declare a social problem. But the philosophical and theoretical criteria for assessing the importance of such declarations are not in place. Most declarations of social problems are therefore insulated from reasonably profound theoretical and critical analysis.

Well-thought-out answers to elementary policy questions are lacking. Why is one problem more important than any other problem? Social science provides no simple, consistent or reliable theory or method for establishing the importance, either relatively or absolutely, of any given problem statement. In practice this means that modern democracies leave the determination of social problems to those they elect. The politicians who make up the legislature where the formal statement of the law is developed decide what social problems will be treated or left alone. They must, if they are to remain in politics, give most of their attention to what they think will get them re-elected. The knowledge social science provides has to take second place to perceived political realities.

There are many other reasons for social problems never ending. They change over time. Change over time also means that no single approach to them will be appropriate. Modern democracies are tolerant of a very wide range of ethical and moral values, with the result that perceptions of them differ. Social problems are typically very complex. Every element of their

definition requires clusters or groups of features, each of which poses its own philosophical and scientific questions.

The last century has seen some progress in defining social problems. But this progress has been relatively little and very slow. The theoretical problems of defining social problems and the political nature of their resolution have many consequences for the influence of social science on the laws that make public policy. One is the virtual certainty of unanticipated consequences of their adoption and the generation of other social problems caused by putatively ameliorative social actions.

There is much room for more use of social science in the legal system. As Ludwig von Mises (1881–1973) explained long ago, 'Nobody will deny that the social sciences and especially economics are far from being perfect. Every economist knows how much remains to be done'. He emphasized, however, that 'the present unsatisfactory state of social and economic conditions has nothing to do with an alleged inadequacy in economic theory. If people do not use the teachings of economics as a guide for their policies they cannot blame the discipline for their own failure' (Mises, 1942, p. 253).

The development of both the law and social science guarantees that each will influence the other. In the future, just as in the past, social science is bound to influence the law more than the other way around. This is because it is a superior, although far from perfect, way to knowledge about social problems. Despite all the difficulties with social science as a source of the law, it is still the best hope we have for the amelioration of social problems.

There is an enormous literature on every topic discussed above. Every part of the modern legal system has a literature covering some of its relationships to social science. There is also a substantial literature that discusses law and some of the disciplines of the social sciences, notably anthropology, economics, history, political science and sociology. Almost all of the social sciences have journals specializing on the relationship of their discipline to law. Examples are *Law and Anthropology*, *European Journal of Law and Economics*, *Law and History Review*, *Law and Psychology Review*, *Behavioral Sciences and the Law* and *Journal of Law and Economics*.

For an early and perceptive view of how some aspects of the subject were perceived, see Michael and Adler (1933). For another sample, see Nagel with Bievenue (1992). For a study of how social science was used to develop new legal theory in the school desegregation cases, see Chesler et al. (1988).

Despite these, and other, efforts, the processes by which knowledge from the social sciences develops into law have not been systematically studied. The scanty material that exists is scattered and valuable mainly for the hints it provides about the need for more work. The best general sources are still the two social science encyclopedias (Seligman, 1934; Sills, 1968).

References

Backhaus, Jürgen G. (1990), 'The requirement to correct dysfunctional legislation' (De eis tot correctie van ondoelmatige wetgeving), *Ars Aequi*, **39** (10), 660–65.

Backhaus, Jürgen G. (1992), 'Integration, harmonization and differentiation of law within the European context from an economic point of view', in Bruno D. Witte and C. Forda (eds), *The Common Law of Europe and the Future of Legal Education/Le Droit commun de l'Europe et l'Avenir de l'Enseignement juridique*, Deventer: Kluwer Law and Taxation.

Backhaus, Jürgen G. (1997), 'The German Waterpenny case: a paradigm for the emerging common law of Europe', in Sabine Urban (ed.), *Europe's Challenges*, Wiesbaden: Gabler.

Chesler, Mark A., Joseph A. Sanders and Debra S. Kalmuss (1988), *Social Science in Court: Mobilizing Experts in the School Desegregation Cases*, Madison, WI: University of Wisconsin Press.

Johnson, John W. (1992), 'Brandeis brief', *The Oxford Companion to The Supreme Court of the United States*, ed. Kermit L. Hall, New York and Oxford: Oxford University Press.

Michael, Jerome and Mortimer J. Adler (1933), *Crime, Law and Social Science*, New York: Harcourt, Brace/London: Kegan Paul, Trench, Trubner & Co. Ltd.

Mises, Ludwig von (1942), 'Social science and natural science', *Journal of Social Philosophy and Jurisprudence*, **7** (3), April, 240–53.

Monahan, John and Laurens Walker (eds) (1994), *Social Science in Law: Cases and Materials*, 3rd edn, Westbury, NY: Foundation Press.

Myrdal, Gunnar (1944), *An American Dilemma: The Negro Problem and Modern Democracy*, New York: Harper.

Nagel, Stuart S. with Lisa A. Bievenue (1992), *Social Science, Law and Public Policy*, Lanham, MD: University Press of America.

Seligman, Edwin R.A. (ed.) (1934), *International Encyclopaedia of the Social Sciences*, 17 vols, New York: Macmillan.

Senn, Mary S. and Peter R. Senn (1993), 'What is a social problem? A history of its definition', *Jahrbuch für Soziologiegeschichte 1993*, Leske Verlag: Leverkusen, pp. 211–46.

Sills, David L. (ed.) (1968), *International Encyclopaedia of the Social Sciences*, 17 vols, New York: Macmillan and Free Press.

31 Cognitive science

John N. Drobak

Economists have looked to cognitive science and psychology to help understand the rational actor assumption in neoclassical theory. Much of legal theory, like economics, assumes that people act rationally or at least can be induced to act rationally by the correct rules. The 'reasonable person' standard is a hallmark of tort law. A natural extension of the economic rationality scholarship has been the re-examination of legal rules using cognitive science (see Langevoort, 1998). Sometimes the use of cognitive science has complemented the neoclassical law and economics literature and reinforced the conclusions reached under neoclassical theory. In connection with contract law, however, the teachings of cognitive science raise the question of whether some parties can truly assent to some contracts. This undercuts a basic premise of both contract law and law and economics theory.

Cognitive scientists have reached a consensus on many aspects of human decision making that are useful for analysing the law. The following brief summary highlights a few of these findings. Humans have bounded rationality, sometimes inherently as a result of limitations in information processing or lack of adequate information, and sometimes intendedly as a way to simplify complexity (Simon, 1982; Clark, 1997). Occasionally, people are rationally ignorant because it helps to simplify a complex world. Likewise people often use heuristics or simple rules of thumb (Kahneman et al., 1982). Invariably, we are overoptimistic, believing that we will do better than statistics show. Cognitive scientists have shown that defective decision-making capabilities stem from a number of attributes. People tend to give more weight to recent experience, to overemphasize short-term consequences and to believe that small samples are representative. People are also not very good at risk estimation, especially for low probability events (Eisenberg, 1995, pp. 213–25).

Although these limitations on human decision making are relevant to legal rules in virtually all the various disciplines of the law, focusing on tort and contract law will illustrate the consequences of using cognitive science. Laws mandating the use of seat belts in cars have been a controversial issue throughout the United States. Opponents of the laws viewed them as an unnecessary government intrusion on private decision making. They believed that government should not impose its will on those who were willing to take the risk that came from driving without seat belts. The implication was that these

drivers were making rational decisions. A number of scholars justified these mandatory seat belt laws by relying on principles from cognitive science. Overoptimism and faulty risk estimation make it very difficult for people to understand the risk of being involved in an accident. Since a car accident is a low probability event, and a fatal or otherwise serious accident is of even lower probability, drivers cannot make rational decisions about whether to wear seat belts (Ulen, 1989, 403–6). Thus the seat belt laws were viewed as a modern equivalent to Ulysses protecting himself and his sailors from the enticing singing of the sirens (Elster, 1984).

The choice between a tort regime of negligence law versus a regime of strict liability has been examined using cognitive science principles. This involves a determination of whether people can be deterred from negligent conduct and whether they can be induced to take precautions. Cognitive science has also been used to analyse the rules governing warnings on potentially harmful products, such as lawnmowers. This, in turn, requires an understanding of when and why consumers fail to read product warnings, fail to understand them and fail to follow them (Latin, 1994). The use of cognitive science has generally confirmed the premise of tort law that people can be induced to act in a safer manner, although it is shown that safety is sometimes much harder to achieve than the law seems to assume (for example, Schwartz, 1994, pp. 378–9, 434–6, 443–4).

The biggest impact of cognitive science has been on contract law because it often brings into question the basic contract requirement of assent. Contracts come into effect when the parties truly agree to their terms. Assent has seldom been a problem, since a party's signature or verbal agreement are enough to show the existence of assent. Cognitive science has begun to question whether there is true assent in many contracts. Under traditional contract law, the doctrine of unconscionability allows courts to undo a contract in those rare instances when the bargain 'shocks the conscience'. The notion is that something about the contract itself or in the process of reaching agreement shows that there could not have been assent to the contract. That doctrine has been soundly criticized because it leaves so much discretion to judges and juries to overturn an agreement. The vast majority of contract law scholars, as well as law and economics scholars, emphasize the great importance of enforcing agreements as they are written. Cognitive science has helped our understanding why some agreements seem so outrageous that they should be unenforceable. By identifying the precise attributes of those transactions, cognitive science is helping to limit the court's discretion and to establish parameters to guide the use of the unconscionability doctrine. In this manner, cognitive science has advanced contract law.

Cognitive science has, on the other hand, raised a question about the validity of a large body of contracts for their potential lack of true assent. The

case of *Carnival Cruise Lines, Inc*. v. *Schute* (1991) illustrates the problem vividly. Eulala Schute purchased a ticket for a seven-day cruise through her travel agent in the state of Washington. Later she flew to Los Angeles, where she boarded the vessel for Puerto Vallarta, Mexico. Off the coast of Mexico, Schute slipped and was injured. After she returned to her home in Washington State, she sued the cruise line. The back of her cruise ticket had contained a contractual forum selection clause, which required all disputes of any kind to be litigated in the courts in the state of Florida. The cruise line argued that Schute's suit was improper in Washington; rather, it had to be brought in Florida. The Supreme Court of the United States relied on the contractual provision in agreeing with the cruise line as to the proper forum. This result is what ordinarily would be expected under principles of contract law. If the case is viewed from a cognitive perspective, however, it is clear that Schute never agreed to the forum selection clause, probably never knew about the existence or terms of the forum selection clause and was rational in those acts.

An ocean cruise is a complex product, with many dimensions. It is likely that Schute was most concerned about the destination, the dates, the price and the amenities offered by the class of service. There were probably many other dimensions of the cruise that were of greater interest to her than where she would have to bring a lawsuit against the cruise line. Even if she had been interested in the forum, it would have been time consuming and expensive for her to attempt to bargain with the cruise line over the forum selection clause. It is also highly likely that the company would not have altered its standard terms. Bounded rationality meant that she should exclude consideration of the forum in making her decision. Even if she had thought about the forum selection clause, overoptimism and defective decision-making capabilities make it likely that she would have assumed she would never have to sue the cruise line. The clause would have been irrelevant to her. For all these reasons, one would have to conclude that Schute never really assented to the forum selection clause and that it should not be used against her as a matter of contract.

This kind of approach would give similar reasons for disregarding many other kinds of contractual provisions. Consumer form contracts, such as those used to establish credit purchases, often contain terms dealing with non-performance that relate to future uncertain low-probability risks. What we know from cognitive science suggests that the vast majority of consumers who sign these kinds of form never truly understood or agreed to what they were signing, nor can the law change that result. In recognition of the potential unfairness to consumers, there have been proposals in the United States to treat consumer contracts differently from business ones in the part of the Uniform Commercial Code that governs contracts (Greenfield, 1997). Simi-

larly, many US states contain uniform landlord–tenant laws that establish rules for residential leases, which cannot be modified by the written lease. These follow from the view that tenants are unable to protect themselves, not only from a lack of bargaining power but also from cognitive factors. Agreements by people dealing with future unknown changes also raise problems. Prenuptial agreements are probably the best example of this. Although courts routinely treat prenuptial agreements as binding (with limited exceptions), there are many cognitive limitations on understanding the actual future implications of the agreement. Undoubtedly, people enter a marriage bursting with undue optimism about their fate. People in love do not expect to end up in divorce, notwithstanding the abstract statistics that show the probability of that outcome (divorce). Bounded rationality is also a problem, since no-one can really predict future changes that would bring the agreement into effect (Eisenberg, 1995, pp. 254–8).

These same cognitive issues arise with business contracts as well. Many law and economics scholars have criticized the courts' special scrutiny of liquidated damage provisions, based on the view that parties can bargain over those provisions just as well as they bargain over other terms. The lesson of cognitive science is that liquidated damage clauses are more likely to be affected by faulty decision making, given the uncertainty of future events. Bounded rationality, the difficulties of obtaining information and representative samples of similar problems, unrealistic optimism and the tendency to give undue weight to recent events all lead to questions about the validity of these provisions. A similar problem arises with the 'battle of the forms'. Purchases often take place through an exchange of forms, with the purchaser first sending a purchase order, followed by receipt of an invoice with the goods, and then sometimes followed by an acknowledgement of the invoice. Generally, the forms tendered by the two parties show agreement on the key terms of the contract (such as the price and the description of the goods), but differ significantly on many peripheral terms. For decades, contract scholars tried to devise methods to encourage people to draft contracts that embodied the complete terms of their purchase and sale (White, 1977). The Uniform Commercial Code in the United States was eventually changed in recognition of the impossibility of the parties ever agreeing on all the terms. Now the code sets up a way to fill in the gaps and to resolve inconsistency between the various forms (Greenfield, 1997, p. 292). What we have learned from cognitive science has confirmed the code's current approach to resolving this problem of inconsistent forms.

Both contract law and economic theory rely on the courts to enforce the bargains reached by people. Trade and a market system depend upon the enforceability of contracts. In addition, transactions costs are minimized by bargains being enforceable according to the terms embodied in the contract.

The lessons from cognitive science, that people often never truly assent to contract terms, nor understand or pay attention to them, undermine the enforceability of contracts. They also increase the likelihood of opportunistic behaviour by parties who *ex post* decide that they do not like the consequences of a contract and so claim that they never assented, even though they understood the transaction and willingly took the risks *ex ante*. It is, of course, unlikely that much of contract law will be undone by cognitive science. Most likely, there will be adjustments at the margin, as with special provisions for consumer contracts. There may also be a greater recognition that contract terms should be enforced for reasons other than reflecting the true bargains between the parties. In *Carnival Cruise Lines, Inc.* v. *Schute*, the United States Supreme Court was careful never to say that Schute had truly understood and agreed to the forum selection clause. Rather, the court indicated that enforcing the clause made good business and economic sense. It was a way for the cruise line to limit its litigation expenses by directing all of its litigation to Florida, where its corporate headquarters was located.

The law and economics movement will continue to benefit from what we learn from cognitive scientists. Recently, cognitive and behavioural science has been used to shed light on the incoherence of administrative regulations (Sunstein et al., 2002), death penalty decision making (Rachlinski and Jourden, 2003) and tax policy (Fennel, 2003), among other topics (see Sunstein, 2002; Ross and Shestowsky, 2003). The actual law will, however, only change slowly from that learning.

References

Carnival Cruise Lines, Inc. v. *Shute*, 499 U.S. 586 (1991).

Clark, Andy (1997), 'Economic reason: the interplay of individual learning and external structure', in John N. Drobak and John V.C. Nye (eds), *The Frontiers of the New Institutional Economics*, San Diego, CA and London: Academic Press, pp. 269–90.

Eisenberg, Melvin A. (1995), 'The limits of cognition and the limits of contract', *Stanford Law Review*, **47** (2), January, 211–59.

Elster, Jan (1984), *Ulysses and the Sirens*, Cambridge: Cambridge University Press.

Fennel, Lee Anne (2003), 'Death, taxes, and cognition', *North Carolina Law Review*, **81** (2), January, 567-652.

Greenfield, Michael M. (1997), 'The role of assent in Article 2 and Article 9', *Washington University Law Quarterly*, **75** (1), spring, 289–314.

Kahneman, Daniel, Paul Slovic and Amos Tversky (eds) (1982), *Judgment Under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Langevoort, Donald (1998), 'Behavioral theories of judgment and decision-making in legal scholarship: a literature review', *Vanderbilt Law Review*, **51** (6), November, 1499–515.

Latin, Howard (1994), '"Good" warnings, bad products and cognitive limitations', *UCLA Law Review*, **41** (5), June, 1193–295.

Rachlinski, Jeffrey J. and Forrest Jourden (2003), 'The cognitive components of punishment', *Cornell Law Review*, **88** (2), January, 457–85.

Ross, Lee and Donna Shestowsky (2003), 'Contemporary psychology's challenges to traditional legal theory and practice', *Northwestern University Law Review*, **97** (3), Spring, 1081–113.

Schwartz, Gary T. (1994), 'Reality in the economic analysis of tort law: does tort law really deter?', *UCLA Law Review*, **42** (2), December, 377–444.
Simon, Herbert (1982), *Models of Bounded Rationality*, Cambridge, MA: MIT Press.
Sunstein, Cass (2002), *Risk and Reason: Safety, Law, and the Environment*, Cambridge: Cambridge University Press.
Sunstein, Cass, Daniel Kahneman, David Schkade and Illana Ritov (2002), 'Predictably incoherent judgements', *Stanford Law Review*, **54** (6), 1153–214.
Ulen, Thomas S. (1989), 'Cognitive imperfections and the economic analysis of law', *Hamline Law Review*, **12** (2), spring, 385–410.
White, James J. (1977), 'Form contracts under revised Article 2', *Washington University Law Quarterly*, **75** (1), spring, 315–56.

32 Connections with law and society research

Jürgen G. Backhaus

This entry will point to some connections between law and society research, on the one hand, and law and economics work, on the other. In emphasizing general similarities, we are trying to connect different bodies of literature that stem from different disciplinary backgrounds but which, in complementing each other, might be fruitfully combined in interdisciplinary law and economics–law and society projects.

'The study of legal change can be considered to be at the heart of sociology of law as an enterprise' (Cotterrell, 1995, p. 351). As a matter of fact, the same holds for a substantial body of literature in law and economics. Although most law and economics work discusses the end point of legal change as the outcome of purposeful interaction (of legislatures, courts, plaintiffs and defendants, and so on), prominent scholars in particular in the field of economic history have interpreted the course of economic history in terms of a change of structures (of property rights assignments) in order to capture externalities and thereby more fully facilitate economic growth. A leading proponent of this approach is Douglass North (see North, 1981). Specifically, since economics is a social science (Frey, 1992) devoted to any circumstance in which alternatives have to be weighed against each other from the point of view of an identifiable agent with sufficiently clear objectives (Buchanan, 1969), law and economics research is not confined to the sphere of the purely economic. In this respect as well, the difference between law and economics and law and society research does not lie in the demarcation of the subjects to be studied. What can be said about legal sociology is equally applicable to law and economics research:

> [M]any of these social relationships (and the institutions, practices, doctrines and understanding that is around them) are centred on or grounded in economic relationships (for example trade, employment, property, fiduciary relationships involving economic risk and corporate organisation). But many are not, or at least not obviously: for example, many social relationships protected by civil rights of free speech, freedom of assembly or freedom of religion. Political, civil and human rights that are the direct concern of law are generally assumed to go beyond the sphere of economic relationships and to have a wider significance than the economic. (Cotterrell, 1995, pp. 347–60)

Despite this wider significance, all the topics mentioned are the subject of current law and economics research (for a survey, see Posner, 1987). The

difference lies, not in the demarcation of subject areas, but in the specific approach offered by the emphasis on choices and opportunity cost.

As for legal development, it is correct to state that 'studies of legal evolution by Henry Maine, Friedrich Karl von Savigny and other historical jurists helped to set in train a development of classical sociological theory brought to fruition by Ferdinand Tönnies, Émile Durkheim and Max Weber, and in which legal development remained an important focus for wider studies of social change' (Cotterrell, 1995, pp. 347–60). Yet, when we discuss Weber, we also have to look at Werner Sombart who, in his *modern capitalism*, (Figure 32.1), gave a detailed analysis of the institutional prerequisites of modern capitalist organization. (See Sombart, 1916 [1927]; Backhaus, 1989.) Even the connection between the protestant ethic and the spirit of capitalism (Weber, 1930) is a subject first discussed by Sombart as one of the prerequisites of capitalist development. If we look at the organization of Book 2 of his first volume on proto-capitalism, the evolution of entrepreneurship is shown to have substantially benefited on the one hand from agents with good access to the means of production such as the princes, the landed gentry and the townsmen, but in addition from agents who had only in common that they shared substantially different ideas and

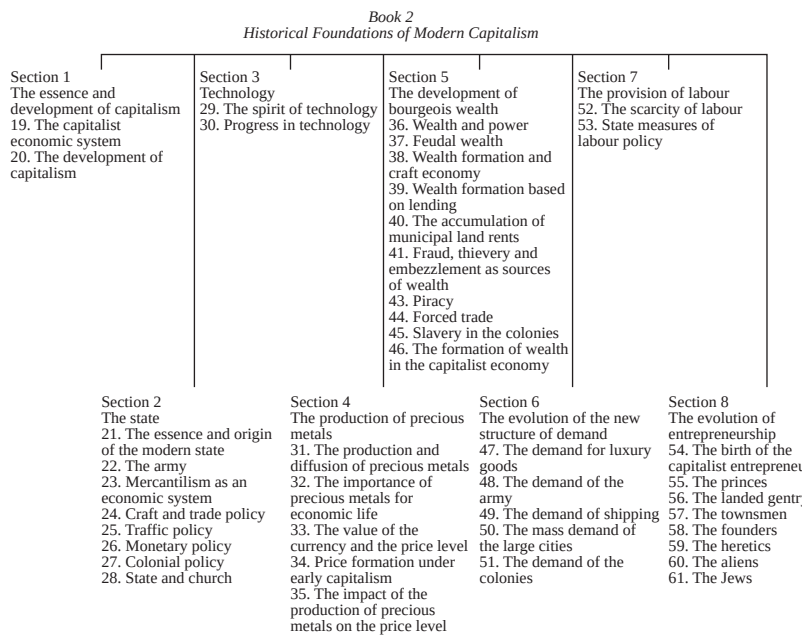


Figure 32.1 Sombart's modern capitalism

beliefs from those of the general population, that is the founders, the heretics, the aliens and the Jews, where the protestants obviously form part of the heretics.

Finally, modern sociological theory emphasizes the function of law as a communication system. Niklas Luhmann and Gunter Teubner have made important contributions in this area. Again, a tradition of economic thought should not be overlooked. Friedrich von Hayek, in his articles during the 1930s, culminating in his 'The use of knowledge in society' (1945), had already emphasized the market, as facilitated by legal institutions, as being the means through which economic agents effectively communicate in a modern economy.

References

Backhaus, Jürgen (1989), 'Sombart's modern capitalism', *Kyklos*, **42** (4), 599–611.
Buchanan, James M. (1969), *Cost and Choice*, Chicago: Markham.
Cotterrell, Roger (1995), 'Sociological interpretation of legal development', *European Journal of Law and Economics*, **2**, Norwell, MA: Kluwer.
Frey, Bruno S. (1992), *Economics is a Science of Human Behavior: Towards a New Social Science Paradigm*, Boston, MA: Kluwer.
Hayek, F. von (1945), 'The use of knowledge in society', *American Economic Review*, **35**, 519–30.
North, Douglass (1981), *Structure and Change in Economics History*, New York: W.W. Norton.
Posner, Richard A. (1987), 'The law and economics movement', *American Economic Review*, **77** (2), May, 1–13.
Sombart, Werner (1916, [1927]), *Der moderne Kapitalismus: historisch systematische Dastellung des gesamteuropäischen Wirtschaftlebens von seinen Anfängen bis zur Gegenwart* [The Modern Capitalism: A Systematic Study of European Economic History], Vols I–III, Leipzig and Munich: Duncker & Humblot.
Weber, Max (1930), *The Protestant Ethic and the Spirit of Capitalism*, trans. Talcott Parsons; reprinted, London: Unwin Reprint, 1985.

PART VIII

TOWARDS AN IDEAL ECONOMIC ANALYSIS OF A LEGAL PROBLEM

33 Towards an ideal economic analysis of a legal problem

Jürgen G. Backhaus

Introduction

Economic analyses are needed for a variety of different applications. Their format will therefore differ according to the application for which they are needed. In this 'how-to' guide, a survey of the most important applications is offered, followed by a discussion of the steps necessary for an economic analysis of a legal problem at a professionally acceptable level.

Three types of analysis

In principle there are three different kinds of analysis: an evaluative analysis, a positive analysis of legal structures (economic reconstruction of legal argument), and a normative (welfare) economic analysis. These three types differ in the economic method and approach to the problem. An evaluative analysis tries to analyse, on the basis of an economic model, the consequences of a particular legal decision or set of decisions, or else an act. A positive analysis aiming at reconstructing the structure of a legal argument or doctrine aims at illuminating complex legal reasoning that cannot be reduced to one or a few organizing principles of legal doctrine. Hence a legal theory is replaced by an economic theory. While these first two types of analysis differ in their level of abstraction, they both belong to the realm of positive (and therefore not normative) analysis. This implies that the analytical conclusions are in principle testable and therefore need to be presented in a testable form. Although it will not always be possible to run empirical tests for all the relevant conclusions of an economic analysis, an appropriate test procedure and the relevant data sources should be indicated in conjunction with a statement indicating which test results would refute the analytical conclusion.

A normative (welfare) economic analysis is different. Its purpose is to distil normative conclusions from a limited set of value judgements. Since legal reasoning is based on value judgements, the task of the economist is to explore the relationship between the various value judgements underlying legal discourse and to indicate where and how they may conflict. If, for example, a particular value judgement is to be given precedence over others, the normative economic analysis can show to what extent this priority will compromise attainment of other traditional goals. In this case, the

economist tries to give expression to lawyers' values to their fullest possible extent.

Since this exercise is actually a sophisticated tool of good lawyering and thus by necessity decision oriented and normative, it does not lead to testable applications. Whether the analysis is correct or not can still be ascertained by testing its logic.

Different applications

The first distinction concerns the economic method or approach. A second distinction refers to the different uses of legal economic analysis. Economic analyses of legal problems are used in vastly different contexts. In court cases, economic experts have traditionally played a role when specifically economic expertise (such as the likely effects of mergers) was required. With the expansion of law and economics as a subdiscipline to virtually every aspect of legal applications, the economist's role has likewise become a broader one. The scope of a legal economic analysis of a court expert will depend on whether the expertise is required for elucidating the particulars of a single case or whether an assessment of the likely consequences of setting precedent is called for. In the first case, the application of received economic knowledge and exercise of standard techniques is normally sufficient, whereas in the second case a model may have to be developed (and tested).

A legal economic analysis plays an important part in administrative applications as well. Again there is a substantial distinction between the analysis of the application of a specific rule to a particular case or whether the impact of the rule on all the cases to which it is applied has to be analysed. Again, in the second case, an explicit model will have to be constructed. The most genuine applications of an economic analysis to legal problems occur at the legislative level. Since at this level we will normally observe a competition between different proposals, the task of the economist is to sort out which aspects of the differences between them are economically relevant. It is hoped that a single model of relevance can be developed that encompasses the competing proposals. The broader the impact of the legislation under consideration, the more thorough hypothesis testing will be required.

General guidelines

Apart from the differences between these various types of analytical approaches, to be discussed shortly, there are six recommendations which should be heeded in any event.

1. The purposes of a particular decision or rule need be strictly differentiated from the outcomes of that decision or rule. A frequent

misunderstanding grows out of the assumption that the two coincide, and it is often the task of the economist to show how and why they do not.

2. The economic analysis must be kept pure, in the sense that no additional assumptions not contained in the model must enter. In particular, the theoretical model must be kept separate from the empirical base.
3. The economic analysis of a particular legal rule or decision must not be kept to the confines of that rule; very often, there are consequences going beyond the area to which the rule or decision was meant to apply. These wider consequences may actually be more important than the narrower and intended ones. It is the task of the economic analyst to render as complete a picture as possible.
4. At every step the analysis should be kept empirically open in the sense that empirical knowledge that is or might be available can be systematically introduced.
5. Legal rules and decisions have to be analysed in terms of whose decision making they are able to affect. Economics is a science about decisions taken by agents, and the consequences of legal decision making are the composites of those decisions with respect to rule making. The legal economic analysis can only be institutionally relevant if the deciding actors have been correctly identified.
6. Finally, the economic analyst, in presenting his/her results, needs to keep in mind how much and what information his/her lawyer counterparts are able to digest and work with. A careful process of translating and simplifying without falsifying may be necessary.

Preliminary steps

Step 1 Define the problem to be analysed and state what types of answers can be expected.

Step 2 Disassemble the problem and reduce it to a sequence of legal questions with as small a residue as possible (see step 4).

Step 3 Make every legal problem correspond to an appropriate set of legal norms and indicate the norms.

Step 4 Now structure the residue:

- (a) list all the questions of fact and indicate the most appropriate method for establishing these facts;
- (b) list all the questions of theory, state the questions precisely in terms of and in conjunction with the appropriate theory or theories to be invoked.

Analytical steps

Step 1 State the basic problem in economic terms, develop one unifying model of reference and indicate the iterative steps needed to develop

the model in several variations until it fits the problem at hand. State the possible outcomes of your model-based analysis.

Step 2 Select and list the norms and legal terms the interpretation of which will be critical to the outcome of the analysis and give an economic analysis of each in turn. You must precisely state the theory which you use for this analysis and ensure that the premises remain the same as in the general model of reference.

Step 3 List the solutions (of step 2) and translate them into the language of the legal norms. Draw the conclusions.

Step 4 Now carry through the analysis as a whole by using the model developed in step 1 and the interpretative results from step 3. State the outcome.

Model variations and sensivity analysis

The four analytical steps may have to be repeated for the main alternatives under consideration.

Step 1 Try to incorporate the main variations of the conclusions (analytical step) into the same model and discuss the outcomes.

Step 2 Highlight the critical assumptions on which the solution depends.

Step 3 Ascertain whether the model is robust with respect to those assumptions and, if it is not, indicate precisely where the model applies.

Step 4 Give a survey of the most important competing models not used in this analysis and indicate the reasons for your preference. If that preference is only slight, you may have to repeat the analysis in terms of the second-best model.

Step 5 Translate the results of your analysis and the main steps by which it came about into legal terms.

Specific observations

Some observations apply specifically to the different approaches and uses, as displayed in Table 33.1.

Table 33.1 Different approaches to law and economics

Evaluative	Positive	Normative	Analysis applications
1	[2]	[3]	[Courts]
4	5	6	Administrative
7	8	9	Legislative

1. Most of the analysis needed in court situations is straightforward applications of economic techniques. Here, reliance on empirical data is particularly called for. Since the expert is brought in to help determine the case one way or the other, he/she should expect to be confronted by a counter-expert. Critical assumptions of the economic analysis are eagerly picked up by lawyers, particularly in adversarial procedures. This tends to confuse all parties involved, including the judges. Therefore one has to be careful to avoid jargon and to state assumptions and conclusions in clear terms. The model itself should be kept basic.
2. When trying to reconstruct involved court decisions in terms of positive economic theory – this is the art Judge Posner has developed to the point of mastery – it is important to trace the roots of the conflicting aspects of court decisions. The conflicting outcomes may result from any one of the following causes:
   (a) faulty reasoning (unlikely in high court decisions, but see Tullock, 1993),
   (b) implicit factual assumptions that have to be brought to the surface and proved correct or incorrect,
   (c) incorrect use of non-legal terms,
   (d) the incorrect generalizations from non-legal theories,
   (e) uncertainty of the courts about the likely outcomes of their decisions,
   (f) conflicting intentions of traditional decision making.

For the economic analyst it is important to pinpoint one or several of these problems and focus on their resolution. The results of the economic analysis will be welcome to the court decision makers as long as they clearly improve upon their lawyering. The judicial intentions have to be clearly formulated and need to enter the analysis. Much of the opposition to the economic analysis of legal problems such as Posner's stems not so much from an aversion to economic analytical techniques as from a sense that the economic analysis is a Trojan horse carrying ideological value judgements not shared by the court.

3. Welfare economic analyses often revolve around highly technical questions that can have important consequences. An example is the proper choice of a discount rate in compensation cases. The economic analyst should be aware of these arguments and state them explicitly in his/her brief, while not relying excessively on dubious assumptions (see previous section).
4. In administrative decisions, typically, more time is allowed to prepare an argument. Therefore the level of sophistication can be substantially

higher. This will refer in particular to the empirical corroboration of the analysis. The administrative agency may actually be helpful and assist in data gathering and processing.

5. When attempting to reconstruct in economic terms a whole set of legal norms such as an administrative decree or an act, the analyst should determine what shaped the act and its current interpretation and implementation. It is unrealistic to assume that an economic analysis can bring about a change in either the act itself or its interpretation or its implementation if that change runs counter to the intentions of powerful pressure groups. It is not the task of the economist to change the political balance. The economist has to take the political status quo as given and try to improve upon it in the Pareto sense. Hence a political economic analysis of the legal norms under consideration is likely to be necessary. This does not mean that the economic analyst should skirt the difficult 'political' or distributive issues. It is quite appropriate and sometimes welcome to elucidate distributive consequences of alternative decisions in conjunction with the emphasis on allocation. Only where one alternative is clearly superior to the other in the Pareto sense should the economic analyst make this judgement, and explicitly so.
6. In some administrative procedures, explicit normative economic analyses are called for. When, for example, environmental impact statements are required, the analytical techniques used should be clearly described. Confusion about those techniques can lead to a perversion of their results.
7. There is no doubt that the economic analyst can play a role of the greatest importance in the legislative arena. Economics is most useful when applied to a large number of similar cases, and this is the case when a piece of legislation is to be enacted. Since most legislation is controversial, the economic analyst should try to stay outside the political turbulences while being aware of the different pressures involved. Nothing hurts the standing of an economic analyst more than being involved in a heated discussion with another economic analyst *in public*. Questions of economic methodology cannot always be resolved. They can never be resolved in the political arena. In order to be successful, it is important to stress the areas of agreement among economists and state, but not controversially discuss, the points of disagreement. In an evaluative analysis of this kind, heavy reliance on empirical techniques with large data bases is particularly useful. The economic analyst who is also politically involved should state clearly where the analysis ends and the political persuasions begin.
8. When analysing legislative proposals or laws that have to be passed or may have to be amended, the economic analyst should keep in mind

that the impact of a law begins much earlier than the date at which it gets passed. The more important the legislation is, the more forceful will be the reactions to it before its enactment. The economic analyst should therefore try to construct a model which takes into account the likely reactions of those affected by the law, and spell out the consequences of alternative proposals in these terms. Far-reaching legislative proposals often have consequences even if they are not enacted. In order to avoid such unintended consequences, it is wise and makes economic sense to provide for loopholes and clauses making it possible to opt out of the provisions of the law. Often the distributive consequences of particular pieces of legislation foster so much opposition that a well-thought-out and balanced piece of legislation cannot be passed. In that case, a long transition period and an effective date pushed far into the future can prove to be an attractive solution. (See Frey, 1983, p. 29.)

9. In the legislative process, the economic analyst can play an important and helpful role by spelling out the implications of different political persuasions. In this way, deadlocked committees can sometimes be got moving again. In order to make welfare economic analyses useful for these applications, it is important to translate them into plain everyday language.
10. A short guide to the literature is in order here. A legal economic analysis which follows roughly the approach outlined above can be found in Backhaus (1987), a book which gives a legal–economic analysis of one court case. The classical works in law and economics are, of course, Calabresi (1970), which he followed up with Calabresi and Bobbitt (1978), and Posner's *Economic Analysis of Law* (1986). An alternative, institutionalist approach can be found in Samuels and Schmid (1981). Texts commonly used at American universities include Polinsky (1983) and Shavell (1987). The text by Ogus and Veljanovski (1984) offers a collection of relevant readings, and widely used, although somewhat opaque, is Veljanovski (1982). Most texts de-emphasize procedural law, but Tullock (1980) is a notable exception. His book is particularly accessible to the non-economist. An excellent case study is Reuter (1983), which has since appeared in a student edition. While most of the literature is in English, the German literature is also extensive, most of it being quoted in Adams (1985) and Backhaus (1987). The most important journals in the field are the *Journal of Law and Economics*, the *Journal of Legal Studies*, the *Journal of Law, Economics, and Organization*, the *International Review of Law and Economics*, and last, but not least, the *European Journal of Law and Economics*.

References

Adams, M. (1985), *Ökonomische Analyse der Gefährdungs- und Verschuldenshaftung* [Economic Analysis of Absolute and Encumbrance Liability], Heidelberg: R. v. Decker's Verlag, G. Schenck.

Backhaus, J.G. (1987), *Mitbestimmung im Unternehmen: Eine ökonomische Rechtsanalyse des Verfassungsgerichturteils vom 1. März 1979 als Beitrag zur Theorie der Wirtschaftlichen Rechtspolitik* [Co-determination : A Legal and Economic Analysis], Göttingen: Vandenhoeck & Ruprecht.

Calabresi, G. (1970), *The Costs of Accidents: A Legal and Economic Analysis*, New Haven, CT and London: Yale University Press.

Calabresi, G. and P. Bobbitt (1978), *Tragic Choices. The Conflicts Society Confronts in the Allocation of Tragically Scarce Resources*, New York: W.W. Norton.

Frey, B.S. (1983), *Democratic Economic Policy: A Theoretical Introduction*, Oxford: Basil Blackwell.

Ogus, A.I. and C.G. Veljanovski (1984), *Readings in the Economics of Law and Regulation*, Oxford: Clarendon Press.

Polinsky, A.M. (1983), *An Introduction to Law and Economics*, Boston, MA: Little, Brown.

Posner, R.A. (1986), *Economic Analysis of Law*, 3rd edn, Boston, MA: Little, Brown.

Reuter, P. (1983), *Disorganized Crime: The Economics of the Visible Hand*, London and Cambridge, MA: MIT Press.

Samuels, W.J. and A. Schmid (1981), *Law and Economics – An Institutional Perspective*, Boston, MA: Martinus Nijhoff.

Shavell, S. (1987), *Economic Analysis of Accident Law*, Cambridge, MA: Harvard University Press.

Tullock, G. (1980), *Trials on Trial: The Pure Theory of Legal Procedure*, New York and Guildford: Columbia University Press.

Tullock, G. (1993), 'Court errors', *European Journal of Law and Economics*, **I**, 9–22.

Veljanovski, C.G. (1982), 'The Coase theorem and the economic theory of markets and law', *Kyklos*, **35** (1), 53–74.

PART IX

CLASSICAL AUTHORS IN LAW AND ECONOMICS

34 Cesare Beccaria (1738–94)

Francesco Parisi and Giampaolo Frezza

It has now become a question of style for American authors to list Cesare Beccaria among the forerunners of the economic analysis of criminal law,[1] with specific reference to the work entitled *Dei delitti e delle pene* (On crimes and punishments) (1764).[2] Beccaria was born in Milan, Italy to a wealthy aristocratic family. He studied at a private college in Parma and graduated from the University of Pavia Law School at the age of 20. During his lifetime Beccaria held leading administrative posts in the Council of Lombardy and served as provincial magistrate, while remaining an active member of cultural circles and a proponent of legislative reform.

Beccaria's theories of social welfare

In the history of economic thought, Beccaria is occasionally included among the fathers of modern utilitarianism. Scholars have debated the appropriateness of such inclusion, given the frequent use of contractarian theories in his work. Beccaria's work shows that utilitarian theories can be built on contractarian premises, and his combined use of contrasting paradigms proves that the two perspectives can successfully coexist.

According to some scholars, Beccaria utilized a hypothetical contractarian framework to justify the purely utilitarian choices of positive law. In reality, he successfully combined the utilitarian and contractarian perspectives into a single coherent framework. In this respect, contractarianism helped him avoid the paradoxes of pure utilitarianism. Additionally, Beccaria provided a powerful inspiration to Jeremy Bentham's utilitarian approach to social welfare, yet anticipating the more modern contractarian justification of utilitarianism adopted by contemporary philosophers such as John Rawls.

The contractarian starting point of Beccaria's work served at the same time to legitimize[3] and to limit the domain of utilitarianism. From the social contract, it followed that the purpose of government was to promote social welfare. The social contract, however, required that the happiness of every individual be taken into account. Thus the maximization could not be undertaken on the basis of mere aggregate values. In this respect, Beccaria's notion of social welfare – although providing the inspiration for Bentham's definition[4] – differs substantially from the Benthamite approach, anticipating the contemporary paradigms of social welfare formulated by Rawls and John Nash. It is not enough to maximize aggregate utility in order to maximize the happiness of society.

It is not clear, however, that Beccaria was fully conscious of the dilemma of social welfare maximization.[5] In his work he alternatively appears to refer to a conception of social well-being given by the summation of individual well-being, and to different paradigms where the internal division of wealth enters as a competing criterion of social welfare. The concern for the latter, however, appears to be mostly semantic.[6]

Beccaria on crimes and punishments

Beccaria's emphasis on utilitarian and contractarian principles constitutes the entire treatment of his work on crime and punishment. The following subsections are designed as a summary analysis of his work with the aim of bringing to light his use of economic reasoning. Other minor writings by Beccaria reveal economic intuitions that anticipate the application of modern economic methodology to legal problems.[7]

On crimes and punishments[8]

In the introduction to his principal work, *On Crimes and Punishments*, Beccaria clarifies the objective, anticipating the method of his analysis. At the outset, he poses the fundamental question of whether the punishment of crimes is worthwhile and, moreover, whether the death penalty is truly useful and necessary for the general level of security and social order. In his subsequent analysis, Beccaria raises additional questions about whether torture and other torments are just or unjust. He brings his analysis to a level of generality addressing the best crime prevention methods.

Beccaria uses generic terms such as 'worthwhile, useful and necessary'. Behind these common terms, he addresses crucial methodological problems which are common to the modern analyses of the efficiency of the criminal law system. Furthermore, the resolution of these questions, on the basis of the specifications he proposes, deserves to be analysed 'with geometric precision': already in line with the initial approach, the pertinence of the economic mathematical method is evidently fundamental to the text, as a detailed analysis of the various chapters will also reveal. This exposition is formulated in the same way as his original work, namely, in brief and concise paragraphs. For the sake of brevity, we have omitted the analysis of paragraphs that do not specifically relate to the economic approach.

The origin of punishments: the right to punish

According to Beccaria, analysis of the origin of criminal law represents a *utility* for society: laws, including criminal laws, are the conditions uniting independent, isolated men and women in society, weary of their unbridled state and continuous wars with each other. It is precisely in this regard that the word 'law' must not be seen as contradictory to the word 'strength': on

closer analysis, the former is, rather, a modification of the latter, that is, a modification most *useful* to the greatest number.

The social contract among men, then, represents a utility for men who renounce freedom in the total sense to create an aggregate which is also penal, designed to give form to the organized social group that, in turn, gives everyone a greater sense of security and freedom.

Consequences (omissis)

The interpretation of the laws (omissis)

The obscurity of the laws

According to Beccaria, laws must be formulated clearly, since the more people who understand the 'sacred code of the laws', the fewer the crimes. Thus, ignorance and uncertainty about punishments are inversely proportional to the frequency of crimes committed.

On detention awaiting trial (omissis)

Evidence, and form of judgments

According to Beccaria, there is a very useful general theorem for calculating the certainty of a fact – for example, the force of evidence of a crime. He formulates the theorem in the following terms:

> When the pieces of evidence for some matter are interdependent, that is, when the pieces of evidence cannot be tested except against each other, then, the more evidence is adduced, the less credible is the matter in question, because anything which would make the earlier parts fail will make the later parts fail too. When all the pieces of evidence for some matter depend equally on a single piece, the number of pieces neither increases nor decreases the probability of the matter, because their joint value as evidence is included in the value of the piece on which they all depend. When the pieces of evidence are independent of each other, that is, when the evidence can be tested other than by each other, then, the more evidence is adduced, the more credible is the matter in question, because the falsity of one piece of evidence does not affect the validity of the others.

Of witness credibility

If legislation is to be useful it must have a clear notion of witness credibility. The true measure of such credibility is the witness's degree of motivation for telling the truth or not. Thus, the credibility must diminish in proportion to the degree of hostility or friendship or to the degree of closeness of the relationship between the witness and the accused. In the same way, this credibility is inversely proportional to the atrocity level of the crime or to the unlikelihood of the circumstances.

Thus, in short, man is as cruel as he is motivated by the hatred or fear that the case in question arouses in him.

Secret denunciations (omissis)

Leading interrogations, depositions (omissis)

Of oaths

Beccaria pinpoints the disutility of oaths which represent a contradiction between the laws and natural human sentiment: the accused has the best possible reason to lie because telling the truth could lead to his destruction.

Thus Beccaria affirms:

> Experience has shown how useless oaths are. Any judge will testify that no oath has ever made a guilty man tell the truth, and so does reason, which rules that every law which runs counter to men's natural feelings is useless and therefore pernicious. Such laws share the fate of dykes which are built straight in the line of a river's flow: they are either flattened and engulfed straight away, or they are eroded and gradually undermined by the eddies which they themselves set up.

Of torture

In Beccaria's view, torture is considered of no utility. Torture involves the punishment of an individual before he has been proven guilty. In this setting, Beccaria observes:

> [This] dilemma is not a novelty: either the crime is certain or it is not; if it is certain, then no other punishment is called for than what is established by law and other torments are superfluous because the criminal's confession is superfluous; if it is not certain, then an innocent man should not be made to suffer, because, in law, such a man's crimes have not been proven.

An additional consideration, therefore, is that if a greater number of men and women – either out of fear or virtue – respect the law than break it, it is equally true that 'the risk of torturing an innocent ought to be accounted all the greater, since it is more likely that any given man has observed the laws than that he has flouted them'.

More specifically, torture is a matter of temperament and calculation which varies in every person in proportion to his physical resilience and sensibilities: on this basis, Beccaria maintains:

> [The] result, therefore, of torture depends on a man's predisposition and on calculation, which vary from man to man according to their hardihood and sensibility, so that, with this method, a mathematician would settle problems better than a judge. Given the strength of an innocent man's muscles and the sensitivity

> of his sinews, one need only find the right level of pain to make him admit his guilt of a given crime.

The use of torture, then, has another consequence: an accused who is innocent is placed in a worse position than the self-confessed criminal. Since both are susceptible to torture, the innocent accused would certainly find himself at a disadvantage after torture:

> [The] innocent is put in a worse position than the guilty. For, if both are tortured, the former has everything against him. Either he confesses to the crime and is convicted, or he is acquitted and has suffered an unwarranted punishment. The criminal, in contrast, finds himself in a favorable position, because if he staunchly withstands the torture he must be acquitted and so has commuted a heavier sentence into a lighter one. Therefore, the innocent man cannot but lose and the guilty man may gain.

Beccaria's summary historical analysis demonstrates the soundness of his conceptions: torture is not believed necessary or useful by Roman law, by British law, by Swedish law or even by military regulations.

Trials and prescriptions

Beccaria maintains that there is a proportional relationship between the false condemnation of an innocent subject and the defects in legislation: the former increases with defects in the laws. For this precise reason the laws, and not judges, must establish a certain period of time for defence and proving crimes.

Given this, and having experimented with lenience in punishing (see below), the author affirms:

> [In] a nation which has discovered the usefulness of moderate punishments, laws which extend or shorten the period available for prosecution in proportion to the gravity of the crime, using remand and voluntary exile as part of the punishment, will be able to provide a simple and restricted class of lenient punishments for a wide range of crimes. But the periods in question shall not increase in direct proportion to the seriousness of the crime, since the likelihood of a crime is in inverse proportion to its seriousness. The period of investigation ought to diminish accordingly, therefore, and the time within which a prosecution must occur increase, which may seem to be in conflict with what I have said about equal punishments being given for unequal crimes if we count the period of remand or period of limitation before the verdict as part of the punishment.

In explanation of this affirmation, Beccaria distinguishes two categories of crime: crimes of atrocity (for example, homicide) and lesser crimes:

> This distinction has its foundation in human nature. The safety of one's own life is a natural right, the protection of property is a social right. The number of motives

> which impel men to overstep the natural feelings of pity is far fewer than the number of motives which impel them by the natural desire to be happy to violate a right which they do not find in their hearts but in social conventions.

That there is a maximum degree of difference between these two categories means that they must be treated differently.

In the case of crimes of atrocity, the preliminary examination schedules increasing the probability of the accused's innocence must be cut back and the prescription period increased: 'But in minor crimes, since the accused's innocence is less likely, the time set aside for investigation should increase, and, since the harm caused by impunity is the less, the period for preparing the trial should decrease'.

Attempted crimes, accomplices and immunity

Beccaria points out that some courts offer immunity to accomplices who betray their companions in crime. But this, on analysis, presents both advantages and disadvantages. One advantage is that of preventing major crimes since a person who has not kept faith with the public domain, that is, the laws, will in the future be held equally faithless in his private relationships. A disadvantage of this procedure, is that a nation is authorizing betrayal, which is detestable in any context. Beccaria sustains that a general law as opposed to a special law is preferable in this matter.

Lenience in punishing

The objective of criminal laws is to prevent the criminal from causing further harm. Thus the punishment must fit the crime and be such as to make an effective, lasting impression on men and women and, therefore, must be the less tormenting prospect for the guilty. That is, the punishment must be as lenient as possible: if this were not so, the above function would not be fulfilled.

On this basis, Beccario is bitterly critical of severe punishments:

> [T]he harsher the punishment and the worse the evil he faces, the more anxious the criminal is to avoid it, and it makes him commit other crimes to escape the punishment of the first. The times and places in which the penalties have been fiercest have been those of the bloodiest and most inhuman actions. Because the same brutal spirit which guided the hand of the lawgiver, also moved the parricide's and the assassin's.

This is to say that the cruelty of the 'suffering inflicted' renders the human soul even more ill-intentioned. Beccario continues:

> If a punishment is to serve its purpose, it is enough that the harm of punishment should outweigh the good which the criminal can derive from the crime, and into

> the calculation of this balance, we must add the unerringness of the punishment and the loss of the good produced by the crime. Anything more than this is superfluous and, therefore, tyrannous.

The death penalty

At this point in his inquiry, Beccaria moves on to analyse whether the death penalty is truly useful and just in well-organized governance. In his work, the death of any citizen is deemed necessary in only two cases:

> There are only two grounds on which the death of a citizen might be held to be necessary. First, when it is evident that even if deprived of his freedom, he retains such connections and such power as to endanger the security of the nation, when, that is, his existence may threaten a dangerous revolution in the established forms of government. The death of a citizen becomes necessary, therefore, when the nation stands to gain or lose its freedom, or in periods of anarchy, when disorder replaces the laws. But when the rule of law calmly prevails, under a form of government behind which the people are united, which is secured from without and from within, both by its strength and, perhaps more efficacious than force itself, by public opinion, in which the control of power is in the hands of the true sovereign, in which wealth buys pleasures and not influence, then I do not see any need to destroy a citizen, unless his death is the true and only brake to prevent others from committing crimes, which is the second ground for thinking the death penalty just and necessary.

In substance, according to Beccaria, more effective than the idea of death is a criminal law which would influence the human soul over an extended period, which is to say, human sensibility is more easily and permanently stirred by lesser though frequently repeated impressions than by a strong, though transient evil. The effects of punishments, then, must not be strong but frequent.

In other words,

> [I]f a punishment is to be just, it must be pitched at just that level of intensity which suffices to deter men from crime. Now there is no one who, after considering the matter, could choose the total and permanent loss of his own freedom, however profitable the crime might be. Therefore, permanent penal servitude in place of the death penalty would be enough to deter even the most resolute soul: indeed, I would say that it is more likely to. Very many people look on death with a calm and steadfast gaze, some from fanaticism, some from vanity, a sentiment that almost always accompanies a man to the grave and beyond, and some from a last desperate effort either to live no more or to escape from poverty. However, neither fanaticism nor vanity survives in manacles and chains, under the rod and the yoke or in an iron cage; and the ills of the desperate man are not over, but are just beginning. Our spirit withstands violence and extreme but fleeting pains better than time and endless fatigue. For it can, so to speak, condense itself to repel the former, but its tenacious elasticity is insufficient to resist the latter.

Still on capital punishment, Beccaria observes:

> With the death penalty, every lesson which is given to the nation requires a new crime; with permanent penal servitude, a single crime gives very many lasting lessons. And, if it is important that men often see the power of the law, executions ought not to be too infrequent: they therefore require there to be frequent crimes; so that, if this punishment is to be effective, it is necessary that it not make the impression that it should make. That is, it must be both useful and useless at the same time. If it be said that permanent penal servitude is as grievous as death, and therefore as cruel, I reply that, if we add up all the unhappy moments of slavery, perhaps it is even more so, but the latter are spread out over an entire life, whereas the former exerts its force only at a single moment. And this is an advantage of penal servitude, because it frightens those who see it more than those who undergo it. For the former thinks about the sum of unhappy moments, whereas the latter is distracted from present unhappiness by the prospect of future pain. All harms are magnified in the imagination, and the sufferer finds resources and consolations unknown and unsuspected by the spectators, who put their own sensibility in the place of the hardened soul of the wretch.

According to Beccaria, therefore, the death penalty 'is not useful because of the example of savagery it gives to men'.

On the basis of these arguments, he goes on to conclude by bringing the readers back to the fundamental question of his contractarian approach: 'what are the true and most useful laws? Those contracts and terms that everyone would want to obey and to propose so long as the voice of private interest, which is always listened to, is silent or in agreement with the public interest'.

In sum, Beccaria's aversion to the generalized use of the death penalty is based on utilitarian, rather than dogmatic, principles, linked to his belief that capital punishments are fundamentally wasteful and hardly deterrent. Executions do not have a lasting impact on the public memory and they occasion the loss of a human life that could have been instrumental to repaying the loss inflicted on society.

Banishment and confiscations

Beccario wonders whether anyone who is banished and excluded for ever from society must also be deprived of ownership of the goods and assets belonging to him. He answers in negative terms because the loss of assets represents a worse punishment than banishment: there must, then, be cases which, in proportion to the crime, admit the renunciation of ownership of assets on the part of the person who is banished; and there must be others where succession of ownership by his legitimate heirs is possible rather than by the state.

Of honour (omissis)

Of prompt punishments

Beccaria's conception of punishment as the price of crime is extremely germane to the modern conception of punishment in the law and economics tradition. He considers the need for a timely imposition of punishment after wrongdoing, and observes:

> The swifter and closer to the crime a punishment is, the juster and more useful it will be. I say juster, because it spares the criminal the useless and fierce torments of uncertainty which grow in proportion to the liveliness of one's imagination and one's sense of one's own impotence. More just because, loss of freedom being a punishment, a man should suffer it no longer than necessary before being sentenced. Remand in custody, therefore, is the simple safe-keeping of a citizen until he may be judged guilty, and since this custody is intrinsically of the nature of a punishment, it should last the minimum possible time and should be as lacking in severity as can be arranged. The minimum time should be calculated taking into account both the length of time needed for the trial and the right of those who have been held the longest to be tried first. The stringency of the detention ought not to be greater than what is necessary to prevent escape or to save evidence from being covered up. The trial itself ought to be brought to a conclusion in the shortest possible time.

More specifically, prompt punishment is more just because it spares the criminal the futile, ongoing torments of uncertainty and because it would be necessary to reduce the detention of the person awaiting sentence to the maximum.

In addition, prompt punishment is of greater utility because the shorter the time between the misdeed and the punishment, 'the stronger and more lasting the association in the human mind between the two ideas crime and punishment. The former will come to be seen as the cause and the latter as the necessary, inexorable effect'. In this respect, Beccaria follows the empiricist argument of John Locke attributing human knowledge to the operation of experience over human senses.[9] The need for clear laws and speedy punishment was instrumental to effective deterrence. Beccaria's utilitarian argument is based on his belief that a timely imposition of punishment ensures a mental association between pain and crime.

Pardons

Beccaria's view of pardons anticipates the modern conceptions of rational expectations, credibility of threats and subjective beliefs in criminal punishment. According to him, 'to show men that crimes can be pardoned, and that punishment is not their inevitable consequence, encourages the illusion of impunity and induces the belief that, since there are pardons, those sentences

which are not pardoned are violent acts of force rather than the products of justice'.

Asylum (omissis)

On setting a price on men's heads (omissis)

The proportion between crimes and punishments

According to Beccaria, there must be proportion between crimes and punishments: punishments must punish only those crimes for which they were created:

> It is in the common interest not only that crimes not be committed, but that they be rarer in proportion to the harm they do to society. Hence the obstacles which repel men from committing crimes ought to be made stronger the more those crimes are against the public good and the more inducements there are for committing them. Hence, there must be a proportion between crimes and punishments.

In Beccaria's view,

> It is impossible to foresee all the mischiefs which arise from the universal struggle of the human emotions. They multiply at a compound rate with the growth of population and with the criss-crossing of private interests, which cannot be geometrically directed towards the public utility. In political arithmetic, we must substitute the calculus of probabilities for mathematical exactitude. Even a cursory look at history shows that disorder grows as the boundaries of empires expand. As patriotic sentiment correspondingly wanes, there is a growth in the motives for crime insofar as each individual has an interest in that very disorder: therefore, the need to stiffen the punishments continually increases.

Beccaria strongly emphasizes the need for punishment to be proportionate to the gravity of the offence. In his utilitarian approach, the gravity should be evaluated on the basis of the harmfulness of the criminal conduct. Crimes, as he indicated in other sections of his work, must be categorized according to the harm they impose upon society. His utilitarian approach was in this respect focused on the internalization of the harm done by the criminal. In other portions of his work, however, he evidences a tension between optimal internalization of the harm and effective deterrence. Beccaria indeed believed that penalties should also create proper incentives for the marginal choices of criminals: the threatened punishment should guarantee that individuals choose lesser rather than greater crimes.

The measure of crimes

According to Beccaria the only true measure of crimes should be the harm that they have caused to society.

In this respect, Beccaria's position is clearly in line with the prevalent economic idea of efficient punishment, where the primary criterion of analysis is the social harm occasioned by the criminal act, rather than the benefit or utility drawn by the criminal.

The classification of crimes

At this point in the inquiry, Beccaria reviews the principal crimes committed in the society of his own time and according to specific categories and classification.

This may be omitted from the present analysis but, in the interests of inquiry, the following is a list of the crimes reviewed (paragraphs XXVI–XL): 1. crimes of treason; 2. violent crimes; 3. defamation; 4. duelling; 5. theft; 6. contraband; 7. indebtedness; 8. against the public order; 9. parasites; 10. suicide; 11. crimes difficult to prove; 12. crimes of a particular kind; 13. legislative errors; false ideas of utility; 14. family feuding; 15. fiscal crime.

Crime prevention

It is better to prevent crime than punish it. This, in Beccaria's view, 'is the principal goal of all good legislation, which is the art of guiding men to their greatest happiness, or the least unhappiness possible, taking into account all the blessings and evils of life'.

To prevent crime, however, the spectrum of potential crimes should be restricted, which is to say:

> [T]he likelihood of crimes is proportional to the number of motives a man might have for committing them, broadening the range of crimes only increases the likelihood of their being committed. The majority of the laws are mere privileges, that is to say, a tribute from everyone for the comfort of the few.

Beccaria continues his analysis by posing the following question:

> Do you want to prevent crimes? Then see to it that enlightenment and freedom go hand in hand. The evils which arise from knowledge are in inverse proportion to its diffusion, and the benefits are in direct proportion. A daring impostor, who is always an uncommon man, wins the adoration of an ignorant people and the jeers of an enlightened one. By facilitating the making of comparisons and multiplying the points of view, knowledge counterpoises different sentiments, which modify each other reciprocally, a process that becomes all the easier as we learn to anticipate the same views and the same objections in others. In the face of widespread enlightenment within a nation, foul-mouthed ignorance is silenced and the authority which has no defenses in reason trembles.

Conclusion

Beccaria's conclusions on the issue of crime and punishment are very modern in approach and noteworthy for their affinity with economic ideals of certain and efficient criminal sanctions: 'In order that punishment should not be an act of violence perpetrated by one or many upon a private citizen, it is essential that it should be public, speedy, necessary, the minimum possible in the given circumstances, proportionate to the crime, and determined by the law'. In his view, a modern legal system requires precise laws, with a minimum use of discretionary evaluation on the part of judges. Every citizen should be able to predict with confidence the consequences of his or her actions, and should be in a position to act accordingly.

Conclusions

The utilitarian and contractarian components of Beccaria's thought constitute the modernity of his approach to the formulation of his theory of punishment. Beccaria avoids the temptation to utilize a purely utilitarian approach in his treatment of crime and punishment. He intelligently gives legitimacy to his theories of economical deterrence, stressing the voluntary and contractarian basis of adjudication and legislation.

Beccaria's work is attentive to the sensitivity of the audience of his time and guides the reader through difficult steps in order to unveil the reasonableness of utilitarian principles in criminal law. In this respect, his work has been likened to the work of contemporary philosophers such as John Rawls and H.L.A. Hart, where the hypothetical or real contractarian origin of law and legislation legitimize utilitarian choices.

Beccaria perceptively unveils the shortcomings of retributive theories of punishment. Such theories are retrospective and are aimed primarily at imposing pain on the wrongdoer in proportion to their guilt and the gravity of their offence. Retributive theories of punishment are not acceptable by utilitarian standards and are not likely to be met with a contractarian justification. Retribution conceives punishment as a way to undo the harm, in a context where no harm can effectively be undone.

On the contrary, Beccaria argues for the appropriateness of utilitarian theories of punishment. Utilitarianism is prospective in its approach. Punishment is imposed on the wrongdoer to prevent future crimes, not to recompense society for its loss. Crimes can be prevented by rendering criminal behaviour less attractive to potential wrongdoers, and punishments should be chosen in order to achieve the optimal level of deterrence. Beccaria, however, was well aware of the limits of utilitarian theories in criminal law. If deterrence is the sole objective of criminal sanctions, occasional exemplary punishments should be a perfect alternative to frequent lenient penalties. Beccaria's contractarian dimension saved him from pushing utilitarianism to such a paradoxical conclusion.

The utilitarian and contractarian foundations of Beccaria's theory of punishment further contributed to the formulation of a theory of division of powers in the context of criminal law. In addressing the respective functions of courts and legislators, he appears to endorse the adoption of purely utilitarian criteria in the drafting of criminal rules, allowing for some retributive consideration by judges who have more direct information on the particulars of the cases.

Finally, Beccaria's utilitarianism resists any paternalistic temptation. If the purpose of criminal laws – and of laws in general – is to deter wrongdoing, then they should contain a negative, rather than positive, command. His warning against paternalistic legislation is most explicit:

> Everything that is publicly useful does not need to be directly commanded, although one should prohibit everything that is harmful. Therefore, all laws that restrict the personal liberty of men have their limit and rule in necessity; and the laws that aim solely at positive utility must restrict personal liberty.[10]

In sum, according to Beccaria legal prescription should only specify what individuals cannot do, rather than attempt to guide human choices. The law should mark the confines between legitimate and illegitimate choices, discouraging the latter through criminal punishment. The search for happiness will motivate human action in the remaining cases and no legal intervention should influence such human choices. In this respect, Beccaria's work is in perfect harmony with the thought of his British contemporary Adam Smith and the inspirations provided by the earlier work of Bernard Mandeville.

Notes

1. For an assessment of the impact of Cesare Beccaria on the American law and economics approach to criminal law, see, F. Cosentino, *Analisi economica del diritto ritorno al futuro?* (Economic analysis of law, return to the future?), in *Foro Italiano*, 1990, V, 153ff. Examples of American references to Beccaria's work in the law and economics tradition are S. Shavell, 'Criminal law and the optimal use of non-monetary sanctions as a deterrent', 85, *Columbia Law Review*, 1232 (1985), n. 2, 53, which states that 'the most important writers on the use of sanctions as a deterrent are probably Charles Montesquieu, Cesare Beccaria, Jeremy Benthan and Bery Becker. Montesquieu and Beccaria were among the the first to adopt a utilitarian style of analysis. While they were often vague and unanalytic, they suggested that, because non-monetary sanctions are costly to impose (they considered mostly the disutility to punished parties), sanctions should be used sparingly and only when likely to accomplish deterrence'; R. Posner (1985), 'An economic theory of the criminal law', *Columbia Law Review*, **85**, 1193–231.
2. The standard English translation of the work is Cesare Beccaria, *On Crimes and Punishments and Other Writings*, Richard Bellamy (ed.), translated by Richard Davies (with Virginia Cox and Richard Bellamy) (Cambridge: Cambridge University Press, 1995).
3. Bellamy (ibid., p. xviii), observes that 'Beccaria employed the ideas of a social contract more as a theoretical device for setting limits to the legitimacy of law than as an actual historical act to explain its origins'.

4. It should be noted that the first English translation of Beccaria's book wrongly translated his notion of social welfare with the use of words 'the greatest happiness of the greatest number'. The phrase is now generally used to identify the so-called Benthamite criterion of social welfare, which served as Bentham's primary purpose for legislation. Ironically, Bentham took inspiration from Beccaria relying on a mistaken translation. As Richard Bellamy, the editor of Beccaria's standard English translation, recently observed 'in spite of the fact that all subsequent English translators have continued to attribute it to Beccaria, the Italian never employed the phrase. His exact words were "the greatest (*massima*) happiness shared among the greater (*maggior*) number"' (see Bellamy, supra, note 2).
5. According to Bellamy, supra, note 2, Beccaria's definition of welfare offers a poor guide to the understanding of his philosophy on the point: 'The notion of division might suggest that he had in mind average as opposed to aggregate utility, whilst the use of the comparative *(maggior*) instead of the superlative (*massima*) potentially indicates that he might have meant "the greatest happiness of the majority" rather than of the "largest number"'.
6. Even when Beccaria appears concerned about the fair division of welfare among the members of society, he conceives no tradeoff between the two dimensions of the problem.
7. For a brief economic analysis by Cesare Beccaria of contraband and customs duties, see: *Scrittori classici italiani di economica politica* (Classical Italian writers of economic policy), Milan 1804, XII, 237, 241.
8. The quotations are taken from the integral work edited by the 'Biblioteca universale', Sonzogno, 1884.
9. See, Bellamy, supra, note 2, at 16.
10. Cesare Beccaria, 'Reflections on manners and customs', translation of Beccaria's *Scritti filosofici e letterari* in Cesare Beccaria, *On Crimes and Punishments and Other Writings* (1995), Richard Bellamy (ed.).

35 Franz Böhm (1895–1977)

Heinz Grossekettler

A survey of Böhm's life and work

Franz Böhm was born in Konstanz, in Baden, on 16 February 1895, the son of a jurist who was later to become the Badenese Minister of Education and Cultural Affairs. He passed the school-leaving examination in 1913, and after serving in the war he studied law in Freiburg. In 1922 and 1924 he passed the two examinations that German law students are required to take in order to complete their legal education. He then became a public prosecutor in Freiburg. As early as 1925, however, he moved to the Reich Ministry for Economics, where he was put in charge of the antitrust enforcement department. While he was employed there he worked on his doctoral thesis about the conflict between monopolists and outsiders.

The dissertation was supervised by Heinrich Hoeniger, a professor of law, and completed in 1931. Böhm then wrote a post-doctoral thesis which was appraised by Hans Grossmann-Doerth, a jurist, and by Walter Eucken, a well-known economist. The two theses were published together in a book entitled *Wettbewerb und Monopolkampf* (Competition and the struggle for monopoly) (Böhm, 1933).

In order to become a professor at a German university, it is usual to gain a doctorate (*Promotion*) and qualify as a lecturer (*Habilitation*), and generally the professorship must be offered by a different university from the one where the scholar qualified. In 1937 Böhm was offered a chair at the University of Jena, but he worked there for only a short time before being dismissed by the National Socialists after protesting against the treatment meted out to Jews.

During his Freiburg period, he had founded, together with Eucken and Grossmann-Doerth, a series entitled *Die Ordnung der Wirtschaft als geschichtliche Aufgabe und rechtsschöpferische Leistung* (Economic order as a historical task and a law-creating achievement). In this series he published a book with the same title (Böhm, 1937). The new series brought into being the Freiburg school of thought which was profoundly influenced by Böhm and Eucken, and which is now regarded as the most important branch of the German neoliberal movement known as '*Ordoliberalismus*'. Since Germany's neoliberal economists cooperated with the anti-Hitlerian resistance movement during the Second World War, the Freiburg school remained untainted by Nazism. For this and other reasons, the German neoliberals

became highly influential during the post-war period and laid the theoretical foundations for the institutional framework of the German economy – the economic system known as the 'social market economy' (Grossekettler, 1997). From the late 1960s onward, however, this institutional order was overlaid by welfarist influences and consequently lost much of its original neoliberal (*ordoliberal*) character.

Neoliberal ideas continued to hold sway over the social market economy until the mid-1960s, and this was due in large measure to the fact that Böhm himself worked resolutely to refine and elaborate the theories expounded by the Freiburg school. In 1945 he was appointed professor at the University of Freiburg, and shortly afterwards he became Minister of Education and Cultural Affairs in what is now the federal state of Hessen. He then took up a chair at the University of Frankfurt am Main, where he continued to teach until his retirement in 1962. In 1952, he conducted the reparation negotiations with Israel and did what he could to secure the most generous settlement possible. From 1953 until 1965 he was a CDU[1] member of the Bundestag, and as a member of parliament he was one of the fathers of the German antitrust law (*Gesetz gegen Wettbewerbsbeschränkungen*). Until his death on 26 September 1977, he was also a member of the Academic Advisory Board (*Wissenschaftlicher Beirat*) in the Federal Ministry for Economics and co-editor of the yearbook *ORDO*, the organ of the German neoliberals. Throughout his academic career, he sought to attain two objectives: (i) to translate into juridical parlance economic proposals for an institutional order as a framework for economic processes, and (ii) to combat every kind of economic power and the conversion of such power into political influence. Tribute was paid to Böhm's work in a special issue of the *European Journal of Law and Economics* (vol. 3, no. 4/1996). This publication contains, among other things, references to Böhm's life, a list of his books and his contributions to learned journals and collaborative volumes, and the titles of the *Festschriften* and symposia that were devoted to him.

How Böhm became a pioneer in law and economics

Böhm's association with the ORDO school of law and economics was determined by an institutional tradition which was still very much alive in Germany during his lifetime, but it is hardly an exaggeration to say that the genesis of his neoliberal ideas was also due to a stroke of luck.

The institutional determinant consisted in the fact that in Germany economics was normally taught at law schools up to the 1930s, so that jurists and economists were able to draw heavily on both law and economics (Backhaus, 1996a, p. 459). As a result,

> [a] generation before the American Bar Association and the American Economic Association decided to come together to devote themselves to the economic

> analysis of antitrust law, this kind of cooperation was perfectly natural at the University of Freiburg … where jurists and political economists taught and did research together in the thirties. (Möschel, 1985, p. 45)

It was a stroke of luck that while Böhm was working on the problems associated with the struggle for monopoly, Eucken was looking for ways and means of setting up institutions to ensure that the price system would remain workable, and Grossmann-Doerth was analysing the excesses of the 'self-created law of the economy' which left its mark in the 'general terms and conditions' of powerful business enterprises. The thread of common interest which bound together these initially separate research programmes was later described by Böhm (1957, p. 99) in the following terms:

> The issue with which we were all concerned boiled down to the question of private power in a free society. It inevitably raised further questions concerning the nature of order in a free society, the various types of economic order that exist, and the dysfunctions which may occur within these various types …

These questions were examined by the founder members of the Freiburg school. The problems were considered on an interdisciplinary basis, not only from a legal viewpoint, but also from the viewpoint of the professional economist. The representatives of the two academic disciplines in question exchanged their views on the problems associated with an institutional order for economic processes. This exchange was fruitful because it enabled all those concerned to remedy their weaknesses and combine their strengths.

When jurists deal with questions concerning private law, their main attention is given to lawsuits in which litigants' property rights are clarified and attempts are made to settle disputes between those concerned. By contrast, when jurists look into problems associated with public law, they are primarily concerned with lawsuits whose purpose is to dispel doubts about what the state is entitled to demand from individual citizens and to what extent the latter are entitled to defend themselves against the state. In both cases, we are talking about a specific dispute between two specific litigants at a specific point in time. Lawyers who concern themselves with such matters therefore tend to overlook the fact that by influencing transaction cost structures in a more or less appropriate manner, regulations may trigger off, accelerate, retard or even completely hinder production and exchange processes in economic systems based on the division of labour. Excessively high transaction costs, for instance, may jeopardize the provision of public goods, produce external effects, create or preserve informational asymmetries, or endanger the functional dynamism of market processes. Excessively low transaction costs may give rise to a situation where unwelcome processes (for example, cartelization) unfold much too smoothly. Inappropriate transaction cost struc-

tures may therefore result in various types of market failure (Grossekettler, 1997, pp. 103ff.). Such coordinative deficiencies are the upshot of human decisions, but in the final analysis no one really wants such things to happen. Thus, for instance, a decision taken by the *Reichsgericht*[2] gave rise to a situation where functional market processes were replaced by cartel agreements in many sectors of the German economy (Nörr, 1994).

Economists, on the other hand, are aware that market processes are self-regulatory processes that depend on a large number of decisions. They know that under favourable circumstances such self-regulatory processes can stabilize desired equilibria in economic systems, provided that we have an appropriate institutional disposition of decision-making bodies, information rights and duties, and economic incentives. However, economists tend to overlook the fact that in a *Rechtsstaat* (state based on law), the institutional prerequisites of appropriate economic processes are actionable and must therefore be formulated in such a way that judicial decisions can be taken in individual cases. There is also the further complication that when the law is applied to individual cases within the juridical framework of a *Rechtsstaat*, the objects of the economic system as a whole are entitled to consideration only in the second place unless there are meta rules[3] which ensure that judges will adhere to the *same* interpretational and organizational rules in *every* individual case. If the choice of meta rules is left exclusively to the experts who are responsible for specific matters, the general legal framework is bound to be riddled with inconsistencies. Good examples are provided by antitrust and regulation law (which is of fundamental importance for the allocation of rights and duties), accounting and bankruptcy law (which is vital for the flow of information), or national and international tax law (which has a major influence on incentive structure) and the specific, economically inconsistent approaches that may be evolved in these domains.

The foregoing analysis is conducive to a better understanding of the characteristic strengths and weaknesses of jurists and economists in respect of the construction of an institutional framework for economic processes. Once we are aware of these strengths and weaknesses, we can comprehend why Böhm opted for a particular law and economics approach. He believed it was his duty to act as an interpreter between jurists and economists. He wanted to make it clear to jurists that a state's *economic order*[4] comprises a nucleus of legislation passed by parliament. This legislation forms the *economic constitution*, a fundamental system of laws which exerts a decisive influence on a country's economic development. This being so, it is imperative that the system in question should be consciously organized in respect of economic functions and adapted to changing circumstances in order to meet new challenges. In nineteenth-century Germany the medieval guild system gave way to economic freedom.[5] This transition was the result of a *conscious* decision

in favour of a new free market economic constitution. By contrast, the way Germany slid into the corporatist economy of the Weimar Republic shows what can happen when people modify the economic constitution *without* making a conscious assessment of the possible consequences. If jurists want to make this clear to their colleagues, it is essential that they should familiarize themselves with the economic coordinating functions of the price mechanism, and they should also learn how to present a legal analysis of economic problems.

Böhm wanted to show economists that jurists have legal organizational instruments which can enable a state to resist dysfunctional influences exerted by lobbies. The institutional draft proposals put forward by economists may well be couched only in general or indefinite terms, but jurists can convert such documents into practicable regulations that can be applied by judges. However, jurists can only formulate such regulations correctly if economists provide them with meta rules for the consistent construction of the entire legal framework for economic processes. Böhm believed that some of the requisite rules could be found in the fundamental principles for economic policy elaborated by Eucken.[6] When these principles were enunciated it was not yet clear how such maxims could be put into practice in the sphere of politics, and even today vagueness and want of precision are par for the course in this domain. Nonetheless, Eucken and Böhm hoped that in the fullness of time observations on the construction of a practicable economic order would acquire an influence comparable to that of ideas about the establishment of *Rechtsstaaten* in the seventeenth and eighteenth centuries. At the outset, after all, these ideas had been confined to the realm of philosophical speculation, yet in the nineteenth century they became a force to be reckoned with.

Böhm was also anxious to show economists how important it is to recognize jurists' capacity for organization and to take advantage of this capacity in order to construct a suitable economic order. Jurists – especially judges in superior courts of law like the German Federal Constitutional Court or the EU's European Court of Justice – constantly modify the economic constitution by their regular decisions, and such decisions have an impact on the workability of the economic constitution. If the idea of an efficiency-oriented economic constitution can be firmly anchored in the literature of the legal profession, much may yet be achieved. Since juridical literature is continually expanded and updated by legal scholars, the odds are that the innovation in question will ultimately achieve more than spectacular political decisions, for politicians are forever chopping and changing, and their decisions tend to be oriented towards short-term electoral goals. However, if any progress is to be made in this domain, it is essential that jurists should work together with economists who appreciate specifically legal problems and who have learned how to present an economic analysis of such problems.

Böhm showed how these basic ideas may be applied in two fields of law, namely antitrust law and codetermination legislation.

Specific contributions: Böhm's influence on German and European antitrust law and codetermination legislation

Böhm not only elevated law and economics to an academic programme. As a politician, he also put forward practical proposals for framing laws which were consciously geared to the economic functions of law. His greatest successes were in the domain of *antitrust law*. The approach he adopted here was typical of the law and economics school. First, he explained the origin of the property rights structure which was commonly accepted in Germany until the 1930s, was reconstituted after the Second World War and remained in existence until the 1950s (see the historical sections in Böhm, 1933). He subsequently presented an analysis which showed to what extent this structure favours or jeopardizes the attainment of the goal of efficient allocation (see the analytical sections in Böhm, 1933 and 1950). On the basis of this work, he then formulated proposals for legislation that would be more efficiency oriented (Josten et al., 1949; Böhm, 1955).

Although Böhm's proposals were never implemented in their pristine form, they exerted a very strong influence on German antitrust legislation (Möschel, 1985, p. 37), and since German antitrust law had a strong influence on European law we can say that Böhm has had an impact on European antitrust law, thereby inhibiting economic processes that might lead to cartellization at the European level.

Antitrust authorities are regulatory bodies, and as such they risk being infiltrated by pressure groups or exposed to the influence of politicians whose activities are oriented towards personal profit rather than the common good. Whenever such vested interests gain the upper hand, they induce grave *infirmities in the body politic*. Böhm sought to avert such dangers by framing new institutions with great dexterity. In order to illustrate his *modus operandi*, I shall give a brief summary of his proposals concerning the organization of the Federal Cartel Office (Böhm, 1950, p. 91).

The function of the Federal Cartel Office (Böhm uses the term *Monopolamt*: Monopolies Office) was to be strictly defined and formulated in a law containing both an injunction and a prohibition: the injunction concerned the elimination of avoidable and the control of unavoidable economic power, while the prohibition was to ensure that the powers conferred upon the office would not be used for any other purposes such as the implementation of industrial policies.

Since the function of the office was thus clearly predetermined with regard to goals and means, Böhm believed that it could and ought to be separated from the rest of the administration with a status comparable to that of the

European Central Bank. Thus the office 'is accountable neither to Parliament nor to the Government, but – like a law court – it has no discretionary powers of a political nature. As far as its measures are concerned, it is subject [only] to the law'.

> Consequently, the Monopolies Office ought to be subject to the same kinds of control as the law courts. Firstly, it ought to be obliged to justify all its decisions in writing, and its justificatory legal analysis should furnish proof that it has abided strictly by the law. Secondly, it should be required to make its decisions in collegial divisions. This is primarily to ensure that the Monopolies Office fulfils its function properly, resisting lobbyists and influential politicians who might try to pressurize officials into showing leniency to powerful business groups.

In the same passage, Böhm goes on to say that 'this kind of body adapts itself to the conditions of political democracy and (in particular) to the tenet of the *Rechtsstaat*, which acts in conformity with issue-related principles established by law'. Speaking from a present-day vantage point, one might add that this type of organization is in accordance with the maxims of new public management, which aims at drawing a clear-cut distinction between two species of decision: (i) *political* decisions about objectives and (ii) *operative* decisions about the means by which these objectives are to be attained in particular cases.

Böhm was less successful in framing *labour legislation in general and codetermination legislation in particular*. This statement holds at any rate for the political upshot of his efforts and for the acceptability or completeness of the arguments he put forward. But not so with his methodology, which in this domain was fully in accordance with the law and economics approach.

Little in the restricted space here available can be said about the intricacies of German labour and codetermination legislation. This body of laws is based on the idea of *social partnership*, and most of the statutes in question were enacted between 1951 and 1976. The statutes contain a catalogue of rights: rights to obtain information and rights to share in policy decisions. These fall into two categories: (i) rights which depend on the decision-making level (political, managerial, and shop-floor levels), and (ii) rights which depend on company size and the economic sectors concerned. A certain number of general trends can be observed. As a rule, workers' representatives have more rights at the lower decision-making levels and fewer at the higher levels; in big firms, they have more to say than in small and medium-sized businesses; and in the coal, iron and steel industries they wield much more influence than in other sectors. It should also be noted that codetermination rights at the managerial level were and – to a certain extent – still are the subject of much controversy.

Böhm commented on the codetermination problem in a number of publications,[7] particularly in a more than 200-page article in the journal *ORDO*

(Böhm, 1951). In this essay, his *modus operandi* is more or less the same as in his discussion of questions relating to antitrust law: he begins by discussing the history of relevant property rights. However, the focus of attention is not on the history of the rights which were then enjoyed by German workers. Böhm's main concern is with the history of proposals concerning the granting of codetermination rights (proposals that date back to the Weimar Republic). He then goes on to analyse the influence that he thinks the granting of codetermination rights would have on the coordinative efficiency of the pricing mechanism. On the basis of this work, he endeavours to ascertain which proposals for the improvement of social relations might be deemed compatible with a free market order and which might not.

Viewed from our present-day vantage point, Böhm's contributions appear somewhat one-sided. To begin with, when Böhm studies the history of the origins of the proposals concerning codetermination, he underestimates the importance of the fact that employers and employees joined forces in the battle to stop the dismantling of industrial plants, particularly in the Ruhr coal basin. The dismantling of these plants was part of a plan devised by the Allies – a plan to cripple the German economy after the end of the war. Even after the Federal Republic of Germany had been established, the dismantling continued for a while. As a result, capital and labour made a united stand against their common foe. This greatly increased the German people's willingness to adopt proposals relating to codetermination, and it subsequently made it much easier for politicians to push through legislation on the subject (Backhaus, 1996b, pp. 369f.).

Second, Böhm's assessment of the impact of codetermination on the coordinative efficiency of the pricing mechanism now appears somewhat unsatisfactory. He resisted demands for the establishment of an 'economic democracy', which were first made in Germany in the 1920s, and reiterated during the discussion about codetermination. Böhm objected that employers do not reign over their employees as monarchs reign over states. He said that employers should be considered as parties to a contract under civil law. As suppliers of goods or working hours, employers and employees ought to heed those price signals whose economic function is to ensure that production is geared to consumers' wishes. This, he said, also necessitates modifications in capacity and job structures. This is certainly true. As modern labour market theory shows, several points have to be taken into consideration when one attempts to assess the efficiency of contractual structures in labour markets:

1. Human capital is a plastic resource whose productivity is determined not only by wage incentives, but also by the organization of human relations.
2. Employment contracts are normally incomplete long-term contracts which must be drafted in such a way that they can be adapted to changes in

collateral circumstances. Hence the need for adjustment mechanisms which have to be devised by institutions.

3. There are informational asymmetries between employers and employees.
4. It is also necessary to take account of the fact that highly specific human capital is often developed only when it is protected against the dangers of exploitation. The acquisition of such know-how is in the interests of employers and employees alike.

These are all research findings which Böhm did not yet have at his disposal in this form, and which now lead us to take a more optimistic view of the efficiency of rules for codetermination. However, what this reveals is not a problem peculiar to Böhm's way of thinking; it is a general problem posed by the relationship between law and economics. Judgements concerning efficiency can be formed only on the basis of available economic knowledge, and the knowledge that economists had in 1951 was different from what we have now.

Notes

1. The CDU is the German Conservative Party.
2. This was the German Supreme Court (established in 1879), which in 1897 declared cartels to be legal.
3. These are guiding principles for drafting and administering laws.
4. The term 'economic order' denotes all the institutions that regulate the allocation of rights and duties, open channels of information, and create incentives (Böhm, 1946, pp. 146f.; Neuberger and Duffy, 1976).
5. 'Economic freedom' here means liberty to establish and carry on any trade or industry.
6. On this point and related questions, see Grossekettler (1997, pp. 44–52 and pp. 111–15; 2002, pp. 33–67).
7. Meijer (1996) offers a broad survey of these publications.

References

Backhaus, J.G. (1996a), 'Good economics, bad economics, and European economics', *Kyklos*, **49**, 449–63.

Backhaus, J.G. (1996b), 'Co-determination in Germany 1949–1956', *European Journal of Law and Economics*, **3**, 369–72.

Böhm, F. (1933), *Wettbewerb und Monopolkampf. Eine Untersuchung zur Frage des wirtschaftlichen Kampfrechts und zur Frage der rechtlichen Struktur der geltenden Wirtschaftsordnung* (Competition and the struggle for monopoly), Berlin: C. Heymann.

Böhm, F. (1937), *Die Ordnung der Wirtschaft als geschichtliche Aufgabe und rechtsschöpferische Leistung. Nebst Einleitung der Herausgeber* (Economic order as a historic task and a law creating achievement. With an introduction by the editors), Stuttgart/ Berlin: Kohlhammer.

Böhm, F. (1946), 'Die Bedeutung der Wirtschaftsordnung für die politische Verfassung' (The importance of the economic order for the political constitution), *Süddeutsche Juristenzeitung*, **1**, 141–9.

Böhm, F. (1950), *Wirtschaftsordnung und Staatsverfassung* (Economic order and national constitution), Tübingen: Mohr.

Böhm, F. (1951), 'Das wirtschaftliche Mitbestimmungsrecht der Arbeiter im Betrieb' (The Co-determination Right of the Workers in an Enterprise), *ORDO*, **4**, 21–250.

Böhm, F. (1955), 'Entwurf eines Gesetzes gegen Wettbewerbsbeschränkungen der Abgeordneten Dr. Böhm, Dr. Dresbach, Ruf und Genossen' (Draft of the German Act against Restraints on Competition by the Members of Parliament Dr. Böhm, Dr. Dresbach, Ruf and others), *Bundestagsdrucksache 2/1269*, 16 March.

Böhm, F. (1957), 'Die Forschungs- und Lehrgemeinschaft zwischen Juristen und Volkswirten an der Universität Freiburg in den dreißiger und vierziger Jahren des 20. Jahrhunderts' (Joint research and teaching of jurists and economists at the University of Freiburg in the 1930s and 1940s), in H.J. Wolff (ed.), *Aus der Geschichte der Rechts- und Staatswissenschaften in Freiburg i.Br.* (On the history of the Faculty of Law and Economics in Freiburg), Freiburg: Albert, pp. 95–113.

Grossekettler, H. (1997), *Die Wirtschaftsordnung als Gestaltungsaufgabe. Entstehungsgeschichte und Entwicklungsperspektiven des Ordoliberalismus nach 50 Jahren Sozialer Marktwirtschaft* (Economic order as creative task. Origins and perspectives of Ordoliberalism after 50 years of social market economy), Hamburg: Lit-Verlag.

Grossekettler, H. (2002), 'Die österreichische Finanzverfassung vor dem Hintergrund des internationalen Standortwettbewerbs' [The Austrian economic order of public finance facing global competition], in E. Theurl, H. Winner and R. Sausgruber (eds), *Kompendium der österreichischer Finanzpolitik* [Handbook of Austrian Public Finance], Vienna and New York: Springer, pp. 32–89.

Josten, P., F. Böhm, W. Bauer, C. Fischer, W. Köppel, W. Kromphardt and B. Pfister (1949), *Entwurf zu einem Gesetz zur Sicherung des Leistungswettbewerbs und zu einem Gesetz über das Monopolamt* (Draft of laws concerning antitrust and the Monopolies Office), submitted to the administration of economic affairs, Herrn Prof. Dr. Erhard, 5 July.

Meijer, G. (1996), 'Co-determination and the market economy', *European Journal of Law and Economics*, **3**, 361–8.

Möschel, W. (1985), 'Das Kartellgesetz heute – aus ordoliberaler Sicht' (The antitrust law today – comments from an Ordoliberal point of view), in Ludwig-Erhard-Stiftung (ed.), *Recht und Gesittung in einer freien Gesellschaft. Zur Erinnerung an Franz Böhm aus Anlaß des 90. Geburtstages* (Right and civilized behaviour in a free society), Bonn: Ludwig-Erhard-Stiftung, pp. 35–48.

Neuberger, E. and E.J. Duffy (1976), *Comparative Economic Systems: A Decision-Making Approach*, Boston, MA: Allyn & Bacon.

Nörr, K.W. (1994), *Die Leiden des Privatrechts. Kartelle in Deutschland von der Holzstoffkartellentscheidung zum Gesetz gegen Wettbewerbsbeschränkungen* (Complaints of civil law. Cartels in Germany from the 'Holzstoffkartell' decision to the Act against Restraints on Competition), Tübingen: Mohr.

36 John R. Commons (1862–1945)

Warren J. Samuels

John R. Commons was a historian and theoretician of the economic system in general and of capitalism in particular. The core of his legal–economic theory was presented in the title of one of his major works, *Legal Foundations of Capitalism* (1924). Central to his analyses were understandings that markets are formed and structured by institutions and power structures which also operate through them; that chief among these institutions were those of law and government; that economy and polity coevolved through a legal–economic nexus; that states and politics were essentially processes of collective bargaining among powerful interested parties; that the economy was an object of legal control and the law was an object of economic advantage; that politics was a contest to control the state, often for economic objectives; and that legal–economic policy making was a process through which the organization and control of the economic system was worked out, constituting in part authoritative determinations of public purpose – and it was only through articulation of public purpose that choices between conflicting private interests could be established. Commons developed these ideas in a series of publications issued throughout his life. Together with his work on labour unions, these ideas were the foundation of his version of institutional economics.

Commons derived his theoretical insights from his practical, historical and empirical studies, particularly in the field of labour relations and in various areas of social reform. He drew insight not only from economics but also from the fields of political science, law, sociology and history. A principal advisor and architect of the Wisconsin progressive movement under Robert M. La Follette, Commons was active as an advisor to both state and federal governments. He was instrumental in drafting landmark legislation in the fields of industrial relations, civil service, public utility regulation, workmen's compensation and unemployment insurance. In all such fields, Commons learned about, participated in and theorized about the fundamentals of the economic sources and roles of law in all its forms. A fundamental element of his approach was his 'method', as he called it, of 'look and see'. What he looked for were the details of a subject and/or problem, and among the details which were important to him were the details of its legal history. For the present state of a topic – subject or problem – for him was but a stage or point in a continuing process of becoming, and law and government were critical elements in that process. Case-study work for Commons thus in-

cluded historical research to determine how we got where we are, as that would reveal both the legal–economic foundations and details of the present and the mode of change operative in both the past and, arguably, the present.

In the *Legal Foundations*, Commons provides a theory of the origins and foundations – the emergence and evolution – of capitalism. The analysis is in part historical in character and relies on empirical legal history and a system of legal terminology for its materials. The legal history is combined with theories of social control, social change and conflict resolution; theories of systemic and institutional organization; a theory of markets, especially of the institutions which form and operate through them; a particular legal–economic model of interpersonal relationships; and a theory of behaviour. For Commons, the law is both influenced by practice and channels economic organization and practice, accomplishing the latter by effectively choosing between alternative customs or modes of organization and practice. The keys for Commons are judicial determinations of public purpose in choosing between parties to litigation, as a mode of transforming property and liberty and forming the legal bases of transactions and going concerns through adopting and changing the working rules.

Commons makes two overriding points with regard to government, points which he considers historically empirically demonstrated, given that they must be comprehended within the larger model of the legal–economic nexus. One point is the fundamental and inexorable role of government in the creation of the economic system through its generation of institutional details, notably the working rules. This directly contradicts the dominant individualist and non-interventionist ideology of Western civilization, and appears as statism to some. Commons believes that the ideology misrepresents both the legal history of capitalism and the role of law in apportioning economic relationships. Government is both fundamental and important to the economic system. Law, like all institutions, both enhances and restricts individual freedom, typically for different persons in any particular situation. More specifically, Commons argues, first, that law is important in determining which individuals' interest will be protected by government as rights, and, second, that individuals operate only within larger social structures and processes in which government and other social control forces operate. Commons also argues that there is no escape from some conception of social or public purpose in choosing between conflicting interests and claims. If one accepts Commons's theory, most conventional arguments about the economic role of government are either naive or disingenuous, in either case functioning to channel the use of government even when government is explicitly denigrated and minimized. It is no conclusive objection to this that Commons himself had an agenda (discussed below); he attempts both to make his premises explicit and to ground his arguments historically.

The second point is that, given the fundamental systemic role of government, the critical question of policy is decidedly not government versus no government, or minimal government, but legal change of law, which is to say, legal change of the interests protected by law. Commons insists that there is (or if there is not yet, there will be) a law(s) pertaining to all economic relationships and phenomena, such that proposals for policy overwhelmingly represent legal change of law and not the introduction of law into a situation or system from which it hitherto has been absent.

Commons emphasizes the importance of the working rules which apportion opportunity and regulate interpersonal relationships: for example, access to and exercise of power. He argues that the most fundamental values and value decisions are not those with regard to the prices of commodities. They are the valuations ensconced within working rules which represent choices between alternative interests and between alternative organizational and power structures. The *Legal Foundations* contains an enormous amount of material illustrating how governments, the courts in particular, have made decisions about interests, relations and organizational structures, often through the reformulation of the concepts of property and liberty. He subsequently devoted a great deal of effort to producing a theory of 'reasonable value', to a significant extent using the same materials. He identifies the valuation process undertaken by government in general and by the judiciary in particular through which changing values are newly embodied in the working rules as part of the central process of the legal foundations of capitalism, a historical process of social (re)construction.

Commons adopts a social constructionist approach to the legal–economic nexus. He argues one principal and four correlative points. The principal point is that both the polity and the economy are neither given nor transcendental to mankind but are worked out through human action. Both the polity and the economy are artifacts and therefore the product of human social construction. The first correlative point is that the social construction has been a continuous process; it is a historical process of continual, and continuing, reconstruction. The second point comprises a rejection of given undifferentiated transcendent metaphysical absolutes, such as property, freedom, the economy, capitalism, and so on. The third point is his emphasis on the story residing in the details. Capitalism as it is at present constituted exists in the form generated by the details of human individual and collective action, especially the cumulative product of court decisions and legislation. Apropos of the second and third points, Commons rejects the existence, for example, of property as a generalized category with pre-existing content, in favour of an understanding that property is the name given to bundles of legal definitions and protections known as rights. In each respect, property is not protected by law because it is property but it is property because it – and not

some other interest – is given legal protection. The fourth point is that, contrary to the retrospective judicial emphasis on past cases, especially on precedent, the fundamental meaning of law is its role in creating the future, its futurity. Central to the exercise of legal discretion is not reliance on precedent, jurisprudential ideology notwithstanding, but judicial selection among alternative lines of precedent.

These views of Commons are evident in a famous statement in the *Legal Foundations* in which he refers to a certain neglect of the role of the common law in choosing between and standardizing the customs of people that he considered so central to the genesis and evolution of capitalism:

> This oversight of the Physiocrats, of Adam Smith and the classical economists, is explicable in the fact that what they mistook for the order of nature or divine providence was merely the common law silently growing up around them in the decisions of judges who were quietly selecting and standardizing the good customs of the neighborhood and rejecting the bad practices that did not conform to the accepted rules of reason. Legislatures and monarchies are dramatic, arbitrary and artificial; courts are commonplace and natural. (Commons, 1924, pp. 241–2)

Commons was interested in government and in the interrelations between nominally legal and economic processes and activities as a problem in its own right. But he also had a normative objective: to make an intellectual case for the recognition of labour interests alongside interests historically designated 'property'. He believed that capitalism, in succeeding feudalism, represented not just a transformation of the economic system but an extension of the range of interests given significant legal protection. Protection of the interests of workers was seen by Commons as a continuation of this process of extending the range of interests given legal status and protection. This orientation pervades his work as the leading labour economist of his time.

He was motivated by the desire to show that, just as land and commodity markets have undergone enormous transformation in their legal foundations, so too have labour markets and, moreover, that it is legitimate to consider modern developments as part of that process. The emergence of the labour market was due to the recognition by law of the right of the worker to his own labour. But labour is bought and sold in the market under conditions importantly established by law, that is, the legal foundations of the labour market. And here Commons points to the differential treatment by the courts in the late nineteenth century of corporations and unions. Each represents collective action, but one is abetted and the other hindered by the courts. This is all a matter of legal choice, says Commons, just as it was in the evolution of the corporation itself. The courts have simply chosen the customs of business over those of workers, at least to that date. Law is presented as the result of a choice of customs.

Commons's social constructivism has two elements in it. One comprises the non-deliberative elements of habit and custom, such that components of the legal foundations of capitalism are reinforced by use. The other is the exercise of deliberative judgment by individual economic actors, including, and analytically most critically, government in the form of legislatures, courts and administrative agencies. The social construction and reconstruction of the economy is the continuing result of both types of behaviour. For Commons the economic system has been transformed in a manner consistent with the principle of unintended and unforeseen consequences, though in a complex and non-ideological manner. Individual buyers and sellers and individual litigators were presumably seeking only their own interests (as they have come to know their interests) throughout the centuries during which capitalism was formed and evolved. But actions and decisions were taken which cumulatively resulted in the transformation of late feudalism into capitalism and eventually in the transformation of capitalism itself. Part of this complex process undoubtedly included the behaviour and decisions of various actors, especially legislators and judges, motivated by conceptions of the kind of economic system they wanted, at least in so far as the matter arose in connection with the legislation or case currently under consideration. Images of the economic system entered into their decisions both consciously and unconsciously, deliberatively and non-deliberatively. Expressed differently, the forces of social construction generated an 'order' through both 'spontaneous' or organic modes and deliberative assessment of existing arrangements by legislatures and courts.

Commons adopted a social constructivist and not a representational conception of language. 'Words, prices and numbers,' he wrote, 'are nominal and not real. They are signs and symbols needed for the operation of the working rules. Yet each is the only effective means by which human beings can deal with each other securely and accurately with regard to the things that are real. But each may be insecure and inaccurate' (ibid., p. 9). Certain words embody and give effect to theories – for example, of property, liberty and value – which are sometimes erected into metaphysical and ontological absolutes but, however held, serve as the basis for the formulation and reformulation of rights, duties and so on. It is out of these processes that both the economic system known as capitalism emerged and evolved and the words were given ascriptive meaning.

Commons analysed legal discourse: how words, as artifacts, encapsulate changing interpretations of experience and of values – and have done so as part of the transformation from feudal to capitalist society. Changes in legal semantics thus incorporate more or less subtle but typically important changes in law and therefore in relative rights, opportunities, exposures and immunities. Commons also indicates the reification and privileging of certain specifications

of concepts like private property as natural and therefore antecedent to and independent of government, in contrast to the actual process of the human social construction of property as an institution; the selective privilegings function as part of that process. He principally emphasizes the growth of the legal definition of property to include incorporeal and intangible property, the transformation of the meaning of property from use-value and exchange-value (and the correlative change from an emphasis on one's own use to others' potential use and the role of withholding). Changes in the legal definition of property were a function of changes in judicial theories of property and liberty, all involving words with socioeconomic consequences.

Commons is thus a social constructivist with regard to jurisprudence as well as language, that is, the language and meanings of law. Changes in definition are in part the means whereby both legislature and, especially, the courts can legislate; such changes in definition, Commons writes, 'are of course not arbitrary. They spring from new conditions. Yet they are discretionary' (ibid., p. 356). Legal change for Commons is more than a matter of changing conditions; it is also a matter of changed perceptions and evaluations of experience, for example, changes in which/whose customs the courts will embody in law and enforce on others, and so on. Changing definitions can arise from new conditions but conditions can be variably experienced and evaluated, depending upon purposes and values, so that even new conditions can be variably interpreted and lead to different changes in language or to no change. All this is central to the social reconstruction of reality.

Commons's general model has several elements, which may be identified and summarized as follows:

1. a legal–economic nexus in which nominally economic and nominally legal (political, governmental) activities are not only mutually determinative and interactive but, especially, coevolve from a common set of sources;
2. a model of economic relationships grounded in legal relationships, ultimately in terms of rights, powers, duties, opportunities, liabilities, exposures and immunities;
3. these relationships are the foundation, the changing foundation, of transactions and of going concerns, the former being for Commons the fundamental unit of analysis and the latter the embodiment of organizational, structural and change variables;
4. conflicts between claims of relative rights and of power and immunity, and so on, historically have involved an inexorable necessity of choice, typically by courts, of which/whose custom will prevail;
5. also inexorably involved are determinations of public purpose, on the basis of which effective choices between conflicting claims of right and so on are made;

6. determinations of public purpose are embodied in the working rules of law, which, among other things, govern the distribution, access to, and use of power among economic actors;
7. the working rules apportion power within the nominally private sphere, within the nominally public or governmental sphere, and between the two spheres;
8. other forms of working rules exist, including those formulated within organizations, such as businesses, albeit always within the zone of authorized discretion permitted by the external working rules, especially those of law;
9. combining recognition of both the variety of forms of working rules and the diffusion of power within the economy, the economy is seen as a system of power and therefore of combined private and public governance, each selectively perceived and each given its selective discursive and symbolic (mythic) expression and rationalization, but together constituting the total legal–economic nexus;
10. historical change within this system is brought about or driven by changing practices, changing beliefs, changing values and so on, especially as ensconced within changing theories of law, property and liberty;
11. the crux of historical change is the legal change of law, that is, the legal change of the interests given legal definition and protection, epitomized in the historical transformation of property.

A central thread of historical change for Commons was the transformations and inversions of legal/political organization and concepts. What had originally been monarchical prerogative became transformed and constrained during the rise of republican, or representative, government; from autocratic monarchy came the modern state. What had initially been legal privileges of the aristocracy under Magna Carta were later reinterpreted, selectively, into common rights. Correlative with the transformation of government from feudal to modern was the transformation of the economy from feudal to commercial/capitalist.

What had been the combination of 'property' and sovereignty that constituted feudalism – lords and their subjects – now became free citizens and their government. Commons recounts the initial predominance of more or less absolute monarchies with their royal prerogatives; the emergence of royal courts and courts at the levels of the lower aristocracy; conflicts between monarch and lesser nobility, between monarch and courts, and between monarchical and other courts; and the growth of the common law of free men arising from the use of their customs and beliefs in the resolution of local inter-party conflicts. This historical story can be understood as several stages of the Anglo-American economic system represented by different systems of

law and government. But economy and polity coevolve: the changing system and structure of government led to systemic and structural economic change, and changing economic change led to changing political change.

Commons portrays a broad, complex and deep series of transformations of English society and does so in terms of the transformation of property through the emergence of several bargains. One was the rent bargain, the origin of modern private property in land, with an enormous social and economic diminution of the rights of feudal landlords correlative with the growth of fee simple property ownership of land. Here the landlords kept their physical land but with greatly reduced social and economic power, or rights.

Another was the price bargain, with the monarchy and feudal lords retaining their physical landed property (diminished as just described) but now, along with the guilds, having relatively negligible control over private economic activity in an economy of free people and not subjects and serfs, and so on. In a correlative bargain, the landed aristocracy, including the monarch, would retain their physical land but lose much of their control over government and its policy. Governments were increasingly in the hands of both a parliament (representative government), in whose operation the business or middle class predominated, and courts whose judges were increasingly amenable to recognizing and protecting the interests of the middle class.

Eventually, the interests of the landed and non-landed (capitalist) property owners were challenged by the working class, and another bargain was worked out: the owners of property retained their physical property and many of the rights associated therewith, but now had, through the extension of the franchise and the resulting greater responsiveness of elected politicians to worker interests, increasingly to share the goals of government policy with a wider range of interests. One result was the formation of what has been called the welfare state, meaning thereby the passage of legislation and programmes promotive of the interests of workers and others in a manner comparable in substance, though not in name, to the promotion through property rights of the interests of those who came to own property. Another result was the growth of a system of industrial governance, centring on the rise and increasing legal recognition of labour unions. All of this took centuries, the negotiation, as it were, over the welfare state and labour unions continuing to the present day.

Commons believed that modern government was a ‘collective bargaining state’. Government was the arena in which these bargains (solutions) were worked out. Commons contrasts two theories of law, one maintaining that law is made by the command of a superior, the other holding that law is found in the customs of the people. His analysis effectively rejects the conventional juxtaposition. He argues that courts make law by choosing between the customs of different groups of people and in that way ‘reconstruct soci-

ety' (ibid., p. 299). Customs are for Commons the raw material out of which justice is constructed. But customs differ, customs change, customs are good and bad, and customs conflict. They are uncertain, complex, contradictory, and confusing. A choice must be made. Law is the result of a choice between customs, ergo a choice between different psychologies and between different interests.

Bibliography

Commons, John R. (1896), 'Political economy and law', *The Kingdom* (Minneapolis), 24 January, 225.
Commons, John R. (1924), *Legal Foundations of Capitalism*, New York: Macmillan; reissued, New Brunswick, NJ: Transaction Publishers, 1995.
Commons, John R. (1925), 'Law and economics', *Yale Law Journal*, **34**, February, 371–82.
Commons, John R. (1927), 'Legal and economic job analysis', *Yale Law Journal*, **37**, December, 139–78.
Commons, John R. (1931), 'Institutional economics', *American Economic Review*, **21**, December, 648–57.
Commons, John R. (1932), 'The problem of correlating law, economics and ethics', *Wisconsin Law Review*, **8**, December, 3–26.
Commons, John R. (1936), 'Institutional economics', *American Economic Review*, Supplement, **26**, March, 237–49.
Commons, John R. (1965), *A Sociological View of Sovereignty*, New York: Kelley.
Samuels, Warren J. (1996), 'Reader's guide to John R. Commons's *Legal Foundations of Capitalism*', *Research in the History of Economic Thought and Methodology*, Archival Supplement, **5**, 1–61.

37 Walter Eucken (1891–1950)

Leland B. Yeager

Was Walter Eucken a pioneer in law and economics? Perhaps not, not if the field is conceived narrowly as 'the use of economics to explain common – that is, judge-made – law' and 'to understand and explain the emergence of legal doctrines in areas of law that do not seem, on the surface, to be susceptible to economic reasoning' (Boudreaux 1994, p. 264). On a broader conception, however, according to which the field overlaps the work of public-choice scholars (Boudreaux 1994, p. 264), Eucken and his Freiburg colleagues are indeed notable contributors.[1]

With reference to the yearbook *Ordo* that Eucken and Franz Böhm launched in 1948 as the 'discussion platform' of the Freiburg school, its adherents are often called Ordoliberals (Grossekettler 1989, p. 42).[2] 'Order', a central concept in their thinking, means something like 'system', 'basic structure', 'institutional framework', or 'organizing principles'.

Ordoliberals emphasize that just as various economic magnitudes are interdependent, so are economics and politics and all other areas of life. Policymakers must take this interdependence among subsystems of society into account when shaping and adapting the corresponding institutions. Eucken and Böhm conceived of an economic constitution corresponding to the political constitution; both require checks and balances to prevent arbitrary use of power (Jöhr 1950, pp. 274–5; Streit 1994, p. 512). Eucken called for 'Unity of Constitutive Principles'. Political decisions on the economic constitution must apply, for example, not only to basic laws but also to special laws like those of bankruptcy. Lawyers can properly help draft the required legal instruments only if they understand and apply the findings of economic research (Streit 1994, p. 13; Eucken 1982a, pp. 130–31; Eucken 1952, pp. 13–16; Böhm et al. 1936, p. 24). Ordoliberals called for basic training in economics for lawyers and basic legal training for economists (Grossekettler 1989, p. 56).

The early membership of the Freiburg school reflects Eucken's vision of close relation between the two fields (Jöhr, p. 273). Key personalities included at least two lawyers, Böhm[3] and Hans Großmann-Doerth. A joint seminar of jurists and economists began meeting as early as 1933–34 (Streit 1994, pp. 508–509). A 1936 manifesto by Böhm, Eucken, and Großmann-Doerth explained why the problems they dealt with required both the methods of reasoning and research of law and economics. Yet they knew that blurring

the boundary between them would disfigure both (1936, p. 25). Economic and legal thought could work together on, for example, the treatment of cartels, general business law, the international monetary system, and other questions of economic and legal policy (Eucken 1950, p. 316).

Eucken's insights about law and economics especially overlap those of Rutledge Vining, 'the Founding Uncle of Public Choice' (so called by Forman 1987; to my knowledge, however, neither Eucken nor Vining ever cited the other's work). Vining's central concern is economists' role as policy advisors, whether official or informal. They try to evaluate the performance of an economic system and suggest ways of improving it. They realize, however – or should – that the government has no direct handle on the outcomes of the system's operation, such as patterns of production and resource allocation and levels and distributions of income, consumption, and investment. What policymakers *can* do is tinker with the rules and institutions that constitute the system. Vining distinguishes among the economic system itself, the outcome of its workings as a largely random process, and statistical measures of that outcome. He thinks of economic and social life as a game whose rules are subject to piecemeal modification in hopes of making the game more rewarding or fairer or otherwise more satisfying. Vining gives great attention to hearings in which legislators and witnesses discuss legal modifications that might improve economic performance (Vining 1984; Forman 1987). Similarly, Eucken conceived of an 'economic order' as the totality of institutional constraints that economic actors must take into consideration. The Ordoliberals did not want the game to be planned, only the rules (Grossekettler 1989, pp. 43, 63).

It is not enough, in the view of Eucken and his colleagues, for economists to explain how the price mechanism works. 'In addition they should win the public over to the cause of realizing conditions under which this functioning proves satisfactory' (Grossekettler 1989, p. 71). Academics should not shy away from taking stands on policy. If science neglects issues of economic order, power groups and literary types press in to fill the vacuum (Eucken 1952, p. 346). (In saying so, Eucken might almost have been anticipating and countering the scorn for 'policy' or 'journalistic' efforts voiced nowadays by academic economists who congratulate themselves on working at the supposed 'frontiers' of their discipline, as well as anticipating how socialist thinking would persist in academic departments of humanities.) Still, Ordoliberals do not expect much from today's politicians. They put more hope in influencing the ideas that tomorrow's politicians, civil servants, and journalists will be working with (Grossekettler 1989, p. 69).

Eucken believed in the power of ideas. He warned against three bad prejudices, as he saw them, of contemporary social science: positivism, historicism, and pointillism. These traits impede understanding of how the economic,

political, and social orders intertwine. Positivism is too well-worn a topic to need rehashing here. The 'relativistic hypothesis of historicism' denies any enduring truth and seems to disparage resisting the supposed tides of history. Pointillism (or punctualism [*Punktualismus*]) in policy means applying expertise separately in different narrow fields. These three prejudices must be overcome to make social science into a decisive ordering power (Eucken 1952, pp. 338–46). Although economics had long ago refuted the old complaint of Saint Simon and other socialists that decentralized private enterprise is chaotic, Eucken saw the anachronistic notion still persisting that central direction is 'social' (Eucken 1982b, pp. 272–3, 275). He also warned against some churchmen's economically ignorant preachments, as about interest on capital (1952, pp. 349–50).

Eucken distinguished between two basic types of economic order: the system of decentralized plans and decisions linked by market processes – 'the exchange economy' – and the centrally administered economy. He further divided the latter type into a small, primitive, and nearly self-sufficient economy such as a medieval manor or monastery or a pioneer farm and a 'planned' national economy of the Soviet or Nazi type (1950, esp. pp. 117–33).[4] He classified systems according to who writes the plans and how fine-grained the division of labour is. A market economy constrasts with both the simple and the complex forms of centrally administered economy in having many partial plans, which gives rise to the problem of coordinating them (Grossekettler 1989, p. 44). Eucken championed coordination through a competitive price system, which he took as the criterion of specific policies (Eucken 1982a, p. 116).

As Eucken knew, the character of an economic system does not hinge simply on whether or not property is private. Although fundamental to preserving individual freedom, a free state, and a free social order, private ownership is not enough. It prevailed among the Romans and in early medieval Europe, yet many small centrally directed economic units existed. Germany has had one law of property, the Civil Code, since 1900, but large elements of economic centralization appeared during the First World War and under the Nazis. The legalization of cartel agreements towards the end of the nineteenth century had a large effect. Problems of limiting property rights and the legal force of contracts have an entirely different significance according to the form of the market (Eucken 1950, p. 351n.). Freedom of contract, like private property, is essential to a competitive system; but it may be misused to destroy competition (Eucken 1950, pp. 85–7, 351n.; 1982a, pp. 124–5).

Although 'ordo' in Latin refers to a 'natural' order in contrast to 'ordinatio', an order imposed from outside, Ordoliberals do not believe that a natural order would emerge if the state merely refrained from positive action. The

spontaneity they valued was of operation, process, and coherent results; they did not expect it of the system or framework itself.[5] For that, the state has active responsibility (Grossekettler 1989, p. 43). Eucken rejected the old faith of classical economists in a 'natural' order under *laissez faire*. The notion of an economy developing as it is fated to do despite the legal system has an affinity with historicism and Marxian determinism (Eucken 1948/1989, p. 38; 1982b, p. 270; 1950, p. 315; Böhm et al., 1936). Ordoliberals also faulted classical *laissez-faire* liberalism for ignoring the power of private citizens over other private citizens and the possible private capture of state power. Ordoliberals favour giving private citizens every economic freedom except freedom to make up the rules of play and to shape the form of the market and the monetary system. These are the functions of a constitutional state (Grossekettler 1989, p. 58).

The economic order and a stable and effective state are parts of an overall order that must be constructed. Contrary to what may have been true in the Middle Ages, a state with unified and consistent will and enough power, under the rule of law, to carry out definite and exactly circumscribed tasks is indispensable to the immense division-of-labour process of the industrialized economy (Eucken 1952, pp. 329, 331–2, 338). Eucken's reason, presumably, is that a fine-grained division of labour and great interdependence among economic subsectors make coordination by properly working market forces all the more necessary.

Eucken distinguished, as already implied, not only among types of economic order but also between economic *order* and economic *process*, the ongoing transactions of economic agents. Under total *laissez faire* the state neither shapes the economic order nor intervenes in the economic process; in a centrally planned economy the state dominates both. Eucken wanted the state, while guaranteeing a market order, to abstain from direct intervention in the economic process. This distinction accords with the principle of the division of power, which aims at protecting the individual from state domination (Eucken 1948/1989, p. 45n; Eucken 1952, pp. 334, 336–7; Molsberger 1987).

Ordoliberals are skeptical of 'piecemeal engineering' in policy, knowing that amid the complexities of economic interdependence, separate measures are likely to work at cross-purposes. Implicitly referring to what later became known as the 'law of unintended consequences', Eucken warned that 'Every single measure of economic policy affects the whole economy and the whole economic process' (1950, p. 314). Like wise doctors, they seek remedies with the least damaging effect on the self-regulation of nature (Willgerodt and Peacock 1989, p. 9). All specific policies should be framed in view of the overall economic order aimed at (Böhm et al., 1936, p. 23; Eucken 1948/1989).[6]

Even such specifics as patent law and liability provisions should be framed for consistency with a competitive order (Eucken 1982a, pp. 120–21; Eucken 1950, p. 316). The purpose of liability for debts (and presumably for torts also) is to facilitate a natural selection of enterprises and managers and also to ensure that capital is handled with due care (Eucken 1982a, p. 126). Limited liability must not allow blunderers to escape punishment and shift their losses onto others (Eucken 1982a, p. 127). More broadly, authority to issue orders must not be separated from liability for the results, as often happens when, for example, government agencies issue orders to private business firms (Eucken 1982a, p. 127). Business concentration should not be fought by antitrust law while inadvertently promoted by company or tax law (Grossekettler 1989, pp. 47, 50).

Hitherto, policy has been made largely piecemeal and *ad hoc* partly because government officials (and business executives) tend to be experts in their own particular sectors, without adequate concern about other sectors and without adequate understanding of interdependence. Eucken likened a government practicing ad hoc policy to a man trying to assemble a machine by arbitrarily sticking together various castings, wheels, pipes, and other metal parts. Similarly, policies in Europe in the early years after the Second World War tried to curb imports and encourage exports while often regarding public finance, credit policy, and wage and price policy as unrelated to the balance of payments (Eucken 1948/1989, p. 39). Similarly again, labour can be fully employed without the economy's working well, as in early postwar Germany when Eucken was writing. This type of malcoordination was distinct from ordinary business depression (due to deficiency of aggregate demand). Macroeconomic policy is not enough (Eucken 1948/1989).[7]

Whatever interventions the state does undertake should be 'conformable' – in conformity with a market economy. They should work through or alongside the price mechanism (unlike, say, price ceilings and floors and exchange control, which subvert the market process). (The distinction between conformable and nonconformable interventions comes from Röpke 1942, pp. 258–64; cf. Grossekettler. Of course, 'conformable' does not mean 'desirable'; even a conformable intervention may be unwise.)

Not only do laws affect the economy, the economy – whether monopolized or competitive, concentrated or dispersed – affects power relations and the political and legal system. An economy of deliberate (as opposed to price-guided) allocations of resources tends to be centralized and therefore to undercut a supposed federal system of decentralizing government power. If policy were to curtail mobility and economic freedom, it is difficult, said Eucken, to see what basic rights could be guaranteed (1948/1989, p. 32 and passim; Eucken might have mentioned Hayek's *Road to Serfdom*, but happened not to).

While according the state an essential role in a healthy economic order, Eucken warned against an outmoded tendency to idealize it as Hegel and Schmoller[8] did. He skewered the commonplace notion that failure of the free economy requires state guidance of the economic process. People saying this overlook how difficult running an economy is and what the state is actually like. Responsibility for state action is so split nowadays, he observed, that nobody – certainly not the collective – is expected to bear moral responsibility. Concrete injustice is no longer felt as such. Eucken saw the dangerous idea widely accepted that the state is bound by no moral laws (Eucken 1982b, p. 270). Paradoxically, the more willing people are to acquiesce in the state's amorality, the more prepared they are to dump unlimited tasks onto it, as if it were the benevolent father of its citizens. People often answer worries about socialization and nationalization by saying that the state is controlled by parliament and ultimately by the people. But such extensive activity alters the character of the state and prevents effective control over the bureaucracy (Eucken 1982b, p. 272). (Invoking Eucken's view of the 'interdependence of systems', Böhm argued the impossibility of combining a democratic constitution with socialist planning; Streit 1994, p. 513.) Furthermore, the growing scope of the state's activity in the twentieth century has promoted yet obscured its loss of authority. The state has become dependent on particular power and interest groups scrambling for its favours, while the business community protects vested interests against competition with something like self-created law, crowding state law aside (Streit 1994, p. 511). Eucken saw all this as further signs of the pointillism (*Punktualität*) of today's thinking (Eucken 1952, pp. 327–32).

Ordoliberals urged the state to be wary of superfluous tasks. Heeding the principle of subsidiarity, a higher level of government should not take on a problem that a lower level or private arrangements could handle (Grossekettler 1989, pp. 48, 51). The constitution should force politicians to pursue limited policies in conformity with the economic order even despite contrary incentives. Ordoliberals value independent organizations, like central banks, whose tasks are clearly defined (Grossekettler 1989, pp. 56–57). Still, as Eucken warned, proposals for self-administering bodies or local administrations often fail to recognize that the economic life of one nation and many nations is one coherent whole. A steering mechanism is necessary (1982b, p. 274).

The Ordoliberals aimed at full or free competition (*vollständige Konkurrenz*). By this they did not mean perfect competition in the technical sense. They employed the more dynamic conception of unhindered market *processes*, with no one having power persistently to charge prices exceeding marginal costs (Grossekettler 1989, p. 45; Grossekettler prefers the term 'free competition').

But who should achieve this competitive order? Occupational and interest groups are scarcely competent or inclined to support it. Indeed, they have all too often made the state an instrument of their sectional designs (Eucken 1952, pp. 325–6). Neither syndicalism nor nationalization nor socialization is any solution to the monopoly problem (Eucken 1952, pp. 334–5; Eucken 1982b, p. 272).

The state must not merely refrain from itself erecting barriers to competition; it must also ensure that private pressure groups and 'law' created by business itself do not restrict the market (Grossekettler 1989, p. 41). 'All kinds of preventive competition should be ruled out: embargos of every kind, loyalty discounts, exclusive contracts and aggressive pricing to scare off outsiders – all of them intended to deter or destroy' (Eucken 1982a, p. 119). Whether price-undercutting should be permitted should not be decided by notions of 'fairness' or 'unfairness'; such judgements should be made in conformity with the economic constitution (Eucken 1950, p. 316). 'Unavoidable monopolies should be obliged to do business with willing customers and be publicly regulated' (Grossekettler 1989, p. 49).

Eucken's 'constitutive principles' went beyond those regarding private property, competition, and open markets.[9] Observing the historical record of inflation and of 'opportunistic, ephemeral interventionism', Eucken added the principles of 'monetary stability' and 'steadiness of policy' (Streit 1994, p. 511). The crises and depressions of the nineteenth and early twentieth centuries were no evidence that the price system as such had failed, for they were often due to failure of monetary policy. Eucken traced inflations and deflations mainly to the linkage of money and bank credit – to the transformation of banks into 'mints' (1982b, p. 273, pp. 117–18). Because 'a certain stability in the value of money' is so important for accurate economic calculations and for avoiding distortions in productive structure and income distribution, monetary policy takes on a special status in economic policy. Furthermore, the monetary system should function as automatically as possible, undominated by interest groups. Eucken was inclined to favour the Graham Plan for a commodity-reserve currency (Eucken 1982a, pp. 116–17; Grossekettler 1989, pp. 47, 49).

Eucken attributed a decline in the propensity to invest – still a topic of live worry when he was writing – not to shrinkage of opportunities in an economically satiated world but mainly to (i) distortions of input prices, particularly of capital goods and labour, relative to prices of output, and (ii) instability of policies regarding money, trade, taxation, and wages (Eucken 1982a, pp. 128–9). A sufficient volume of investment presupposes a certain consistency in economic policy (Eucken 1982a, p. 129; Jöhr, p. 275). Furthermore, uncertainty created by rapid policy changes gives impetus to mergers and concentration in industry (Eucken 1982a, p. 130).[10]

Eucken took a lively but critical interest in Keynes's writings on money-macro policy. He thought that Keynes did not understand the rationale of *laissez faire*. He would have scorned a supposed blend of Freiburg microeconomic policy and Keynesian macroeconomic management. He warned against confusing market-oriented policy for stability with Keynesian full-employment policy (Tuchtfeldt, pp. 77–8; Eucken 1952, pp. 27n., 230ff., 236, 244, 350, 361).

Heinz Grossekettler points out a resemblance between Eucken's Ordoliberalism and the Constitutional Choice approach of James M. Buchanan. The latter is more concerned with taming the power of politicians and the state ('leviathan'), however, while Ordoliberals emphasize taming private power (Grossekettler 1989, p. 66). In wanting to shape the economic order deliberately, Ordoliberals come close to what F. A. Hayek criticized as 'constructivism' (Grossekettler 1989, p. 59; Hayek 1989, pp. 22, 48, 50-51, 60, 66, passim, and earlier writings). Sharing key ideas with the 'Old Chicago School' of Henry Simons, they tried to work out 'a positive programme for a free and socially aware market competition' (Grossekettler 1989, p. 43, making reference to Henry Simons's 1934/1948 'positive program for laissez faire').[11] They worry about a self-destructive tendency inherent in competition. They are less optimistic than the (new) Chicago School about a tendency of competition to establish itself. The schools' difference in attitude may trace to different experiences – the Americans' with a large market, the Europeans' with small and protected markets (Grossekettler 1989, p. 59).

Simons and Eucken also resembled each other in considering themselves champions of liberalism, in urging restraints on private power (Simons being especially worried about labour-union power), and in insisting on stable money. Both were sceptical about limitations on liability. Both favoured redistribution through a progressive tax system somehow arranged so as not to endanger the propensity to invest (Grossekettler 1989, p. 50).

Eucken's work is also reminiscent of W.H. Hutt's proposals (1944) for postwar reconstruction of the British economy: the state was to take truly drastic action in efforts to make the real world resemble the imaginary world described in textbook chapters on pure competition. Eucken's views were less extreme. Even so, I conjecture that he would have modified his views, as Hutt apparently did, if he had lived to see further experiences with overambitious government and also to absorb the teachings of the Public Choice school.

Eucken did not base his policy views on legal dogmas or on any particular philosophy – not, for example, on notions of natural law or natural rights. Replacing 'random interventions and naive *ad hoc* experiments', policy should rest on principles having a sound theoretical and empirical foundation

(Grossekettler 1989, p. 46). Private property and other basic principles should be upheld because they have proved necessary for a competitive system (Eucken 1982a, p. 130). Presumably like most other economists, Eucken thus adopted a philosophical stance no more profound than a tacit and casual utilitarianism. For his purposes, this stance was appropriate.

Notes

1. Backhaus (1980, p. 3n) already noted that Eucken and his colleagues were harking back to an earlier tradition of close connection between legal and economic studies.
2. The founders of Ordoliberalism were committed Christians. Ludwig Erhard, who became German Economics Minister and later Chancellor, was a 'political entrepreneur' working with Ordoliberal ideas. The newspapers *Frankfurter Allgemeine Zeitung* and *Neue Zürcher Zeitung* were sympathetic to the school (Grossekettler 1989, pp. 57, 60, 67–8). As a member of the Federal parliament, Franz Böhm helped infuse Freiburg ideas into the postwar German economic order (Streit 1994, p. 513).
3. Although the idea of economic orders in recent writings traces to Eucken, Franz Böhm deserves credit for promoting the economic order as a test of law, especially economic law, in German-speaking territory (Backhaus 1987, p. 67).
4. Backhaus (1980, pp. 37–8) finds Eucken implying a further distinction – between a competitive exchange economy and a 'corporative' one dominated by cartels.
5. The Ordo school's distinction between spontaneous and planned economic orders – between catallaxy and planning – has a parallel in the distinction between spontaneous legal order [Recht] and legislation [Gesetz] (Backhaus 1987, p. 68).
6. Citing F.A. Hayek's (1969) as well as Eucken's insistence on the 'Systemgerechtigkeit' of individual policy measures, Backhaus (1987, p. 51n.) finds an interesting parallel with the German Constitutional Court's insistence that laws concerning the economy must ensure the capacity of the affected institutions and organs to fulfill their purposes.
7. More recently in several European countries, policy against unemployment provides another apparent example. Various uncoordinated individual measures have neutralized one another, and heavy unemployment persists (Coe and Snower, 1996).
8. The 1936 manifesto of Böhm et al. further criticizes Gustav Schmoller. He allegedly rejected making decisions on principle and instead wanted each issue of state intervention to be decided on its own separate merits. His Historical School forgot how to ask fundamental questions and became essentially incapable of transcending everyday experience. Schmoller did make several often-cited respectful remarks about theoretical research; but under his leadership, German political economists forgot how to apply and improve theory and carry out economic analysis. They were unable to help shape the economy because they did not understand economic interrelations (1936, pp. 20–22).
9. The Freiburg principles overlap, largely though not totally, with several put forth by Jürgen Backhaus. The legal order is an integral part of any economic order and conditions the latter's efficiency. Legal policy must be an integral component of any economic policy. Law must be recognized as a public good if its supply in adequate quantity and quality is to be assured. Policy for economic law must straightforwardly recognize that value judgements cannot be avoided. Two approaches to the theory of economic policy, although different, do not exclude one another. The pragmatic approach works with principles that aim at separating questions of analysis and value. The contractarian approach seeks to bring into the analysis value judgements capable of commanding consensus (Backhaus 1987, p. 46).
10. Eucken's insistence on steadiness of policy parallels F.A. Hayek's insight that every social system rests on acceptance of certain fundamental and reliably durable values. If people are to form reliable expectations, it is important that the normative bases of economic policy coincide with generally accepted social values (Backhaus 1987, pp. 51–2).

11. When I first read Simons's 'positive program', roughly ten years after it first appeared, my reaction was: 'I can see the positive program, but where is the *laissez faire*?' The program is activist indeed.

References

Backhaus, Jürgen (1980), *Öffentliche Unternehmen*, 2nd edn, Frankfurt: Haag and Herchen.

Backhaus, Jürgen (1987), *Mitbestimmung im Unternehmen*, Göttingen: Vandenhoeck and Ruprecht.

Böhm, F., W. Eucken and H. Großman-Doerth (1989), 'The Ordo Manifesto of 1936', in Alan Peacock and Hans Willgerodt (eds), *Germany's Social Market Economy: Origins and Evolution*, London: Macmillan for the Trade Policy Research Center, pp. 15–26.

Boudreaux, Donald J. (1994), 'Law and economics', in Peter J. Boettke (ed.), *The Elgar Companion to Austrian Economics*, Aldershot, UK and Brookfield, USA: Edward Elgar.

Coe, David T. and Dennis J. Snower (1996), 'Policy complementarities: the case for fundamental labor market reform', IMF working paper summarized in 'European job growth hinges on deeper labor market reform', *IMF Survey*, 25, 16 December, pp. 397–8.

Eucken, Walter (1950), *The Foundations of Economics*, translated by T.W. Hutchison, London: Hodge.

Eucken, Walter (1952), *Grundsätze der Wirtschaftspolitik*, Edith Eucken and K. Paul Hensel (eds), Bern: Francke, Tübingen: Mohr.

Eucken, Walter (1982a), 'A policy for establishing a system of free enterprise', in Ludwig-Erhard-Stiftung (ed.), translated by Derek Rutter, *Standard Texts on the Social Market Economy*, Stuttgart and New York: Gustav Fischer, pp. 115–31.

Eucken, Walter (1982b), 'The social question', in Ludwig-Erhard-Stiftung (ed.), translated by Derek Rutter, *Standard Texts on the Social Market Economy*, Stuttgart and New York: Gustav Fischer, pp. 267–75.

Eucken, Walter (1989), 'What kind of economic and social system?', translated from *Ordo* 1, 1948, in Alan Peacock and Hans Willgerodt (eds), *Germany's Social Market Economy: Origins and Evolution*, London: Macmillan for the Trade Policy Research Center, pp. 27–45.

Forman, Frank (1987), 'The Work of Rutledge Vining', paper read at the annual conference of the Southern Economic Association, 23 November, Washington, DC.

Grossekettler, Heinz G. (1989), 'On designing an economic order. The contributions of the Freiburg School', in Donald A. Walker (ed.), *Perspectives on the History of Economic Thought, Volume II, Twentieth-Century Economic Thought*, Aldershot, UK and Brookfield, USA: Edward Elgar, pp. 38–84.

Hayek, F.A. (1969), 'Rechtsordnung und Handelsordnung', in *Freiburger Studien: Gesammelte Aufsätze*, Tübingen: Mohr (Siebeck).

Hayek, F.A. (1989), *The Fatal Conceit*, W.W. Bartley III (ed.), Chicago: University of Chicago Press.

Hutt, W.H. (1944), *Plan for Reconstruction*, New York: Oxford University Press.

Jöhr, Walter Adolf (1950), 'Walter Euckens Lebenswerk', *Kyklos* 4, 257–78.

Molsberger, Josef (1987), 'Eucken, Walter', in *The New Palgrave*, II, 195, edited by John Eatwell et al., New York: Stockton Press.

Peacock, Alan and Hans Willgerodt (eds) (1989), *Germany's Social Market Economy: Origins and Evolution*, London: Macmillan for the Trade Policy Research Center.

Röpke, Wilhelm (1942), *Die Gesellschaftskrisis der Gegenwart*, 4th edn, Erlenbach-Zürich: Rentsch.

Simons, Henry C. (1934), 'A positive program for laissez faire', reprinted 1948 in his *Economic Policy for a Free Society*, Chicago: University of Chicago Press.

Streit, M.E. (1994), 'The Freiburg School of Law and Economics', in Peter J. Boettke (ed.), *The Elgar Companion to Austrian Economics*, Aldershot, UK and Brookfield, USA: Edward Elgar, pp. 508–15.

Tuchtfeldt, Egon (1982), 'Social market economy and demand management: Two experiments in economic policy', in Ludwig-Erhard-Stiftung (ed.), translated by Derek Rutter, *Standard*

Texts on the Social Market Economy, translated by Derek Rutter. Stuttgart and New York: Gustav Fischer, pp. 65–80.
Vining, Rutledge (1984), *On Appraising the Performance of an Economic System*, New York: Cambridge University Press.
Willgerodt, Hans and Alan Peacock (eds) (1989), 'German Liberalism and Economic Revival', in *Germany's Social Market Economy: Origins and Evolution*, London: Macmillan for the Trade Policy Research Center, pp. 1–14.

38 Otto von Gierke (1841–1921)

Jürgen G. Backhaus

Otto von Gierke was born in Stettin, the son of a Prussian official. He studied law at the University of Berlin and held professorships at the universities of Breslau (1872–84), Heidelberg (1884–87) and Berlin (1887 until his death). He is generally described as having formulated, on the basis of the writings of Jacob Grimm (1785–1863), a specific Germanist school of law, as opposed to the Romanists:

> At the beginning of Gierke's career, German legal scholarship was dominated by the Romanist school of Savigny; but Gierke began and remained a strong Germanist. The Germanists, like the Romanists, were historically minded; their research, however, did not take them back to the Roman empire, Justinian's code, and the reception of that code, but followed the path marked out by Jacob Grimm to the law of the ancient German *Mark* and the *Gemeinde* (local community) to feudal records, town charters, the rules of guilds in search of 'truly German' and legal principles. The first volume of Gierke's *Das deutsche Genossenschaftsrecht* (1868–1913) … was the first product of his self-imposed task of broadening the foundation for a German theory of associations by a detailed study of successive types of organizations in German history. (Lewis, 1968)

From an economic point of view, the emphasis should not be on the specific nationality of the German law. The emphasis of this empirical research is rather on the law as it has developed by itself, instead of the law as it has been imposed by church or state, both drawing on Roman law which is thereby transported through time without recognition of the immediate history and historically grown context where it is to be applied. The working hypothesis is that the law can be empirically found in the customs, charters, contracts and the like of identifiable associations of men, whether commercial, political, charitable, professional or other associations. In establishing contractual relationships one with the other, people actually create a new legal entity, such as a corporation. Here the view is distinct from the Roman tradition, which has to create the fiction of a legal identity being granted by the state through fiat. Hence the research is firmly embedded in identifiable economic practices: 'Gierke's concept of "social law" enables him to construe the internal rules of churches, trade unions, business corporations, etc., as independent of state determination and to put such bodies on an equal basis with human persons in claiming areas of freedom into which the state cannot intrude' (ibid., p. 180).

This specific approach to legal research was similar enough to economic research into history of the time for it to be merged with the economic research to form a powerful critique of legislation.

The German Civil Code

Economists have a long tradition of influencing legislation. Schmoller called his journal the 'Annals of Legislation, Administration and Political Economy in Germany'. Hence it is no surprise that economists would intervene in the debates about large legislative projects. Outstanding examples are the great codifications, such as the 'Code Civil de Napoléon' and the German Civil Code. The German Civil Code can actually even be considered an explicit attempt at efficient legislation. The Civil Code was, in its ultimate form, passed with the explicit input of the leading economists of their time in Germany, and based on explicit economic reasoning. Otto von Gierke played an important role in this process. How this came about is worth recounting briefly.

In 1848, when the first German democratic parliament convened in the Church of St Paul in Frankfurt, there were no fewer than 56 different legal systems governing bills of exchange. This exceeded by far the number of member states of the German Federation at that time. It is obvious that such splintering of the legal system stood in the way of the rapidly developing market economy, and there was therefore already an initiative of the Frankfurt parliament to pass a common German Civil Code and also a Commercial Code. However, the parliament had a short lifespan and it dissolved before it had even seriously started the task. There were new initiatives in 1866 and 1869 in the northern German Federation, the precursor of the Reich, but only after the unification of the German states as a confederation of principalities in 1871 could the task be resumed, and this happened with initiatives in 1873 in both the imperial parliament and the imperial council (the representation of the confederated princes). In 1874, the pre-commission of five members was established, and in the same year a commission of eleven was established under the chairmanship of Heinrich Eduard von Pape. This commission issued a report in 1887, and the 'act of introduction' with various drafts of by-laws was also issued, in 1887 and 1889, in order to provide for a broad discussion. However, the editor of the leading economics journal in Germany at the time, the *Jahrbücher für Gesetzgebung, Verwaltung und Volkwirtschaft im deutschen Reich* (Annals of Legislation, Administration and Political Economy in Germany), Gustav von Schmoller with his colleagues in the German Economic Association (Verein für Socialpolitik), agreed that the draft was impractical because it did not build on established economic practices and their respective legal counterparts, rather providing a deductively reasoned set of norms based on the Roman law tradition, and hence not

corresponding to the economic practices of a developed industrialized market economy. Entire issues of the journal were devoted to critiques of the draft act, with the articles by Gierke having the strongest influence. (See, for example, Gierke, 1888, 1889.) This unsolicited advice led to the establishment of a second commission in 1890 which provided a completely revised draft Code in 1895 which was duly passed, after stormy discussions in the imperial parliament, on 1 July 1896, with 222 votes in favour and 48 against, and in the imperial council on 14 July, and ratified by the emperor on 18 October. The Code took effect on 1 January 1900 and, with numerous small revisions, it has remained the same ever since. This episode illustrates again that efficient legislation is possible, although it also illustrates that, had the economics profession not intervened, legislation would probably have been imposed that would have burdened the German economy with high and persisting transactions costs. The Code would not have become the export article it actually became, and still is today.

Gierke's approach proved to be important well after the Civil Code had been passed. In building on his criticism of the first draft of the Code, he developed a full theory of German private law, consisting of the three pillars of the law of persons (1895), the law of things (1905) and the law of obligatory relations (1917). It was this coherent body of legal theory, based on empirical research, which allowed for consistent interpretations of the Code in the vicissitudes of the different economic and social environments of, first, the empire, second, the war economy of 1914–18, the challenge of the hyperinflation that it posed to private contracts, the corporatist economy established by the national socialists, the war economy again, and finally the reconstruction period in Germany after the currency reform of 1948.

Bibliography

Gierke, Otto von (1888, 1889), 'Der Entwurf eines bürgerlichen Gesetzbuches und das deutsche Recht' (The draft of a Civil Code and German Law), *Jahrbücher für Gesetzgebung, Verwaltung und Volkswirtschaft*, **12** (3,4), 842–904, 1179–1264; **13** (1), 183–322.

Gierke, Otto von (1895–1917), *Deutsches Privatrecht* [German Private Law], 3 vols, Munich and Leipzig: Duncker & Humblot; vol. 1, *Allgemeiner Teil und Personenrecht*, 1895; vol. 2, *Sachenrecht*, 1905; vol. 3, *Schuldrecht*, 1917.

Lewis, John D. (1968), 'Otto von Gierke', *International Encyclopedea of the Social Sciences*, vol. 6, New York: Macmillan/Free Press.

Pearson, Heath (1930), *Origins of Law and Economics: The Economist's New Science of Law 1830–1930*; reprint, Cambridge: Cambridge University Press, 1997.

39 Augusto Graziani (1865–1938)

Giampaolo Frezza and Francesco Parisi

Introduction

The work of the Italian scholar Augusto Graziani anticipates in many respects some insights of modern law and economics and institutional theory. His arguments focus with great clarity on the link between economics and legal studies. In his view, such a link is twofold. To identify and exemplify laws independent of their natural development and of any formal qualification of them – a particular need in the fundamental works under study here – Graziani illustrates the generic relationship between economic reasoning and legal choices (especially on the part of the lawmaker). Sometimes, in fact, the very existence of certain legal institutions is ontologically dictated by economic requirements: one example that can be cited even without deeper analysis is the institution of the state, which represents the modern constitutional entity and owes its origin to economics. The state was born of the need to draft public budgets in order to accomplish public goals. On the other hand, and more specifically, the author analyses some legal developments in terms of efficiency, and according to the strict criteria of the microeconomic method, especially in the areas of inheritance and contract. Interestingly, Graziani introduced the modern concept of 'efficient breach' of contract.

The economic foundation of law

We start our analysis of Graziani's work by considering his perspective on the 'economic foundation of law'. 'The economic foundation of law', a lecture to inaugurate the 1893–94 academic year at the University of Siena, was published in the university's *Yearbook* in 1894 and reproduced in its entirety in the monograph, *Economic Facts and Theories* (Graziani, 1912). Graziani addresses a series of identical considerations in his introduction to the manual: *Institutions of Economic Policy* (Graziani, 1917). A basic idea permeates the text: the close bond between legal regulatory choices and economic order. These relationships are specified regarding various spheres and types of law. The text spans constitutional law, administrative law, penal law, commercial law and, finally, private law.

Graziani believes it superfluous to insist on the economic bases of constitutional phenomena:

> The representative system of government, which now governs so many people in Europe, America, and some in Asia, whether parliamentary or purely constitutional, whether hinged on monarchical or republican institutions, principally owes its origin to economic reasons. It is the budget, the need to raise income and make expenditures, that is the primary stimulus for convening representative bodies of various classes and cities to institute legislation and public decisions. However, the approval of financial measures implies at least some effective control, however inapparent, over the power of the state. (1912)

Moreover, Graziani specifies that if representative phenomena have increased it is partly due to increases in public spending. Thus,

> [T]he rise in consolidated debt is an effect of democratic systems of government, though it would be more truthful to say that the extension of debt has, rather, resulted in the prevalence of representative institutions. This is not to say that it is a remote cause; debt extension is the result of general economic and financial conditions. Conversely, loss of sovereignty or the political semi-dependence of some states is a consequence of their economic condition … [and, in particular, of their economic instability]. (Ibid.)

Thus, Graziani submits that the political prevalence of a lower House in representative systems is for financial reasons. For, in comparison with the upper House, the lower chamber is almost exclusively concerned with state spending and income (the US Congress provides a historical example here). Although at first glance economic factors might not seem to have any influence on electoral systems, Graziani states,

> [I]t cannot be denied that, despite the perfection of electoral mechanisms, the wealthy, who often employ people who have no assets of their own and who look to the wealthy employers for their means of physical subsistence, benefit from the votes of these employees who have no recourse to the threat of dismissal. Even the basic rights of citizens, such as equality and freedom, are linked to economic circumstances, since the economically weaker may, unfortunately, only aspire to formal equality, which is of little relevance when more marked material inequalities predominate. (Ibid.)

In summary, Graziani believes that the foundations underlying institutions evade public constitutional law, unless they are of an economic order which often determines the very form of the organism and its action.

Similarly, respect for private property – which Graziani believes has been historically transformed from a collective to a capitalist concept – represents the effect of economic conditions which demand that personal ownership is necessary in order to guarantee the continuity of production necessary to meet man's fundamental needs.

State administration also proves to be a byproduct of economic necessities. It 'reduces the cost of providing certain services'. For example, the administra-

tion of military forces, whether voluntary or compulsory, is more cost efficient when consolidated through the state. Even the distribution of powers between the state and lesser political bodies, such as municipalities, is not immune to profound economic influence. The degree of authority exercised more or less by these lesser political consortia is associated with their revenues, their economic potential, and with the natural distribution of the classes, in relation to other factors. Therefore, the question of centralization and decentralization of bureaucratic or political agencies of various scope, cannot be decided on the basis of abstract criteria of administrative technique alone, but must also be resolved according to economic principles and economic relationships. The system of public works and benefits is a direct function of economic conditions. Public safety also correlates to economic factors, given that various social classes have differing degrees of potential to disturb the public order and safety as a function of the quantity and nature of the wealth at their disposal.

Graziani's view of criminal law is that human choices are the product of economic choices. In this respect, Graziani criticizes the traditional formulations of criminal law. Whether they are based on classical or positive lines, they fail to understand the economic nature of criminal behaviour. Graziani formulates a proposition according to which 'especially in periods of low wages, a correlation exists between subsistence prices and crimes against property, which tends to diminish when prices fall and to increase when prices rise' (ibid.). In other words, according to Graziani, 'facility of subsistence naturally removes the stimuli to commit crime against property which, in contrast, intensifies at times of insufficient means for survival' (ibid.). As to crime against people, this phenomenon is obviously related to economic factors, 'but it would not be inexplicable if a temporary benefit to a worker earning a subsistence wage – a benefit such as a reduction in the price of necessities – were to induce the worker to impulsive spending which, in its turn, might foment crime' (ibid.).

Graziani employs the same line of reasoning when considering the case of constitutional, administrative and private law. Some institutions of civil law – specifically the law governing inheritance and contracts – assume specific connotations of interdependence between economics and law. In Graziani's view, commercial law is the supreme area for the application of economics to the explanation of legal change: the transformation of the bill of exchange, for example, is reflected in the legal reforms in the German law of 1848, and was adopted by the Italian Code of Commerce. The dispositions concerning carry-over and financial transfers, commercial companies, transport, insurance, and less standardized trial periods and contract forms, are evidence of the change in commercial needs.

Relationships in civil law are an immediate source of economic law. On close inspection, according to Graziani, any initial assessment of the civil

code, and of all civil law in general, would seem to show that the highest order stands in arbitrary opposition to economic necessities. For example, civil law does not regulate the labour contract, but it does pay close attention to regulating assets. From this perspective, the governmental regulation of the institutions of ownership – that is, middle-class patrimonial interests, yet not those of wage-earners or of labour contracts in general – appears to be inefficient. However, these latter interests were emerging in the increasingly industrialized society in which Graziani lived and worked. Conversely, this situation is justified in precise relation to an efficiency criterion: considering the 'changeable' nature of the working relationship, which moves in relation to social substratum changes, and more rapidly so than in other legal sectors, it is better regulated by special laws 'themselves more easily modifiable'. Consequently,

> [L]egislation protecting factory workers, which promotes disability insurance and old age pensions, has come to be applied in countries seeking to introduce improvements in working conditions. The correlation between legal and economic institutions does not implicate the total adaptation to the latter, but is an indication that the continual violation of economic laws is impossible or harmful, and will inevitably spark serious and long-lasting reaction. (Ibid.)

On this premise, Graziani's examination of the principal institutions of civil law looks particularly at family law, and especially the three distinct stages of its historical evolution. It progresses from 'primitive promiscuity, to matriarchy, then patriarchal authority'. Graziani believes that the economic basis of marriage – that is, the origin of the family – is clearly manifest:

> The form of the *coemptio* is the obvious mark of institutions which considered the woman as a useful economic object to be bought by a man from her father in exchange for a price. The antithesis of the *coemptio* is the dowry system, which the woman brings as patrimonial enhancement should her husband, because of changed social conditions, have to procure sustenance for the family if he is unable to derive any pecuniary advantage from his wife's labours. (Ibid.)

Graziani notes that impediments to the marriage of certain people, and the caste systems of some countries, also originate in the different economic realities of those societies.

If, in addition, one analyses the patrimonial rules within the family, one finds that the economic origin and the development of such rules emerge with even more compelling clarity. Graziani is bitterly critical of labour and employment laws because they reconstruct the foundation of ownership as an institution. He maintains that it is important to reflect on the economic function of ownership prior to utilizing such category in an analytical context. Changes in the production system denote radical changes in the system

of ownership. Graziani further illustrates the economic basis of ownership considering the institution of legal servitudes, which nearly all jurists now regard as limitations on ownership imposed to maintain economic order. The expression 'for public purposes' is also an emanation of the economic nature of property rules since only in matters of public works, for growth believed to be useful to production, is an individual invested with the authority to transform immovable property into movable. Today, this is not necessarily detrimental because immovable property is no longer covered by rights of sovereignty, at least in Italy.

The institution of possession can only be explained with the use of economic categories. After a brilliant historical analysis of this institution, starting from its medieval connotations, Graziani addresses the important achievement of later scholarship. The rules for the protection of possessions are 'a necessary complement to the rules for the protection of property; it provides proof of ownership, which also benefits the non-owner. Wherever the system of individual ownership is acknowledged, possession must also be acknowledged, or the owner would constantly be required to provide proof which would be very difficult' (ibid.). In summary, possession was introduced by the lawmaker with ownership in mind. The practical value of this situation is not determined by the fact that it presents certain advantages, but by the relationship established between its advantages and disadvantages, and by the preponderance of the former over the latter. In this context, Graziani adds: 'Since the economic organism requires the permanent state of individual ownership, and this could not subsist without protection of ownership, it is clear that this same protection of ownership is an effect of the system of production and distribution of wealth' (ibid.).

The economic insight in Graziani's works, however, reaches its maturity in his study on laws governing inheritance and contract. In 'The economic nature of death and succession' (Graziani, 1890) and *Institutions of the Financial Sciences* (Graziani, 1929, 526ff.) the author criticizes and probes the theoretical justification behind succession taxes which were founded either on a presumed relationship with hereditary law, or on the legitimacy of the various ways of acquiring property. Graziani explains the economic effect of the alternative types of taxes. Inheritance taxes, from this perspective, must not be considered as a state's share of the inheritance, but are designed to integrate and diversify income tax.

Likewise Graziani justifies progressive taxation and higher rates for succession taxes in terms of a marginal rate of substitution:

> Death duties are paid by the taxpayer because, while his personal wealth increases, he senses that the value of public services grows at the same time. In other words, the amount for the public spending that each one of us is willing to

> pay depends on the relative urgency and intensity of his individual needs, and on the amount of wealth possessed. … An individual compares the cost of goods with the benefit of satisfying his needs. In determining cost, he compares the fraction required to obtain his satisfaction of desires (benefit) with the total quantity of wealth at his disposal. That is, the larger total quantity of wealth possessed, the lower the value attributed to each fraction. This is because ultimate utility diminishes; in fact, the more goods at an individual's disposal, the more needs he satisfies, since the last one he satisfies, where the utility is the dependent factor, is of lesser importance. An inheritance increases the wealth of the heir so the heir attributes less value to each fraction of his wealth and, in his progressive estimation in equal values of the utility of public services, he must provide the state with a larger quota which has the same value for him as the quota previously paid. Given that, after the inheritance brings an increase in assets, a larger expenditure for public purposes (tax) represents the same sacrifice or the same utility as a smaller quantity paid before the inheritance. Therefore, the correspondence between individual evaluations and taxes must be graduated. So, the simple principle is that the ultimate degree of utility explains the tax level. It could be observed that the ultimate utility of the wealth inherited is also a function of that already possessed by the heir, and which most tax systems do not take into account. But since this is a special tax, and not general income tax, it objectively regards the quantity of wealth as the basis for the taxation, the case of every property tax.

Some tax systems generally allow for rates of taxation to be graded depending on how close the family relationship between the deceased and the heir. Graziani considers those criteria of taxation: observing that in the event of devolution to descendants who were already advantaged during the father's life by the sums now left to them legally on his death, these descendants see their patrimony only partially increased, in some cases reduced to nothing or even to a loss. This is because they now lack the income generated by the father's work. But the economy of more distant relatives is distinct from the economy of the deceased, since a larger patrimonial advantage can be presumed which leads to a reduction of the level of final assets across the board. It is not because the state regards the right to hereditary collateral as less legitimate than that of the descendants whom it taxes more heavily, but because it considers their patrimonial relations, and by this means state law takes into account the different situations and the different expectations of the parties involved.

In Graziani's view,

> [T]he succession tax constitutes a supplement to income tax. Diversification of rates prevents overburdening on property income. Income tax differentiation accentuates the greater burden on established income but, when liabilities are moderate, this serves to tax income that would otherwise escape taxation. The recent legislative tendency is toward the progressive approach, but this also depends on how other taxes are formulated as a whole. Laws must adjudicate the

> overall ratio between the wealth of the taxpayer and the tax rate. But one objection might be that the deceased no doubt paid direct taxes on the income of the assets he bequeathed and, therefore, death duties imply dual taxation. But since the matter regards individuals, this payment is not of a dual nature. However, it does guarantee that any extraordinary increase in the private wealth of an heir through inheritance does not escape taxation. It is true that the heir in his turn will pay income tax and that he will pay indirect duties on the same wealth. But this is payment imposed by the lawmaker since death duties are a sort of partial patrimonial tax designed, as we said before, to integrate, compensate and balance. (Graziani, 1929)

More generally, Graziani considers the entire law of succession, suggesting its coherence with the laws of economics: 'Wealth induces greater savings. Individuals are motivated to accumulate more wealth than they need to pay averted present or predicted future needs. Naturally private, deeply-rooted ownership leads one to draw up a will and testament'. The freedom to make a will is limited only by requirements of an economic order. The perpetual dispositions and substitutions are explained by the need for free circulation and use of wealth. Limitations and prescriptions established by the law for estates or future interests of long or perpetual duration are justified by a simple economic logic. Even the legal quota, according to the author, must be studied in relation to economic phenomena. It 'may be a reaction to the majority and a means to prevent any significant concentration or centralization of wealth since the obligation to distribute a share of the inheritance equally among descendants results in a redistribution of assets as opposed to a concentration of assets'.

Finally, in the essay 'The economic foundation of law', (Graziani, 1912), he observes how civil obligations, especially those arising from binding contracts, have evolved according to the economic needs of the time. Specifically, Graziani clearly anticipates, many of the modern theories of economic analysis of contract law, both in the execution phase of a contract and in regard to its breach:

> A contract is always the result of an economic calculation on the part of the contracting parties who estimate that the advantage they derive from the thing obtained in exchange or from the services of others is higher than the cost of their performance. If I agree to sell an object to Tom for Lit. 10,000 and he accepts, it means that Tom attributes more utility to the object than to Lit. 10,000 and that, vice versa, I attribute a higher subjective value to the Lit. 10,000 than to the object. The same reasoning may be applied to all contracts whether they are for the purchase of a thing (*do ut des*), or for the acquisition of a service (*facio ut des* or *facio ut facias*). This is because the obligation may be said to arise only when the contracting parties make an inverse assessment of their respective utility of the wealth or services exchanged. (Graziani, 1912)

This calculation, in Graziani's view, constitutes the very substance of the contract. However, the emergence of a legal obligation lies in compulsion designed to prevent violations of the agreement:

> The parties, in fact, foresee the possibility of non-fulfilment of the contract. They can formulate provisions in the contract in advance, or they can rely on the protection of the law. The legal obligation in the contract could be matched by the protection that law provides to the promises of the contracting parties. This objection is not irrefutable because, primarily, one must look for the need for legal enforcement. If no party ever defaulted on his commitments, there would be no reason for legal enforcement or legal sanctions. But, there are three reasons why he might well default on his commitments: first, the contracting parties would aim at obtaining the other parties' services without performing themselves; second, the contracting party might not fulfil his obligations because of changed circumstances; third, he might no longer consider it convenient to carry out the original agreement. (Ibid.)

Graziani considers each possible situation in turn:

> In the first case, the other party would derive no benefit and, if the law were not to guarantee the performance of contracts, fulfilment of contracts would be the exception rather than the rule. Therefore, legal enforcement is an economic necessity. In the second case, the same analysis applies. In the third case the reasoning is that the value ratios change continually, rendering an agreement less profitable or detrimental even though it may have been useful in various respects when it was concluded. If the utility no longer exists for both parties, the contract would cease to exist automatically. What would be the justification of the Code's stance on bilateral contracts, if not the sanction of this economic principle? If legal enforcement were to be applied strictly, the contractor who did not receive the services promised would have to limit himself to asking the courts to intervene to enforce the agreement's execution. However, in reality, the party who did not receive the services contracted for has a choice as to remedy. He may ask the court to enforce the contract, or try to resolve it himself, or he may refuse to perform his obligations under the contract. While this demonstrates the correlative nature of contract obligations, it also allows for an escape from a contract where, because of changed circumstances or convenience, a contract is no longer beneficial. For reasons of public interest, the lawmaker may not recognize the validity of some agreements, such as those on gambling credit. However, in these cases, the economic reasons stand intact. (Ibid.)

In this sense, Graziani anticipates the common law concept of 'efficient breach' of the contract as recently elaborated by legal scholars. Still referring to the contract, Graziani considers the economics of typical categories of contract:

> Special contract forms are established, as everyone knows, with the aim of guaranteeing the specific desires of the parties. Idiosyncratic provisions are the easiest, though not always the most convenient, expression of that desire. Individual

> contracts, moreover, are more manifestly dependent on economic relations. For example, the possibility of emphyteusis redemption is the consequence of the tendency to remove all obstacles to full freedom of disposing of lands and capital, which constitutes a hinge of the present system. The emphyteutic institution itself brought marked advantages when the density of population was low with respect to land or when, in order to stimulate cultivation, soil conditions were such that long-term concessions had to be made and nearly all ownership rights were conferred upon the farmer. The same transformation of the lessee's right from the simple *jus ad rem*, which is totally personal, to *jus in re* precluding the owner as a legal person, was born of the need to guarantee the farmer against sudden dismissal such as to impede or reduce the intensity of cultivation and production in general. (Ibid.)

References

Graziano, Augusto (1890), 'The economic nature of death and succession', in *Studi Sensi*, Woodland Hills, CA: Siena.

Graziani, Augusto (1912), 'The economic foundation of law', in *Economic Facts and Theories*, Woodland Hills, CA: Siena.

Graziani, Augusto (1917), 'Introduction', in *Institutions of Economic Policy*, Milan, Turin, Rome: Fratelli Bocca Editori, pp. 26–50.

Graziano, Augusto (1929), *Institutions of the Financial Sciences*, 3rd edn, Turin: UTET.

40 Robert Lee Hale (1884–1969) – legal economist

Nicholas Mercuro, Steven G. Medema and Warren J. Samuels

Introduction[1]

Robert Lee Hale was born in New York in 1884. His secondary school education was undertaken in New York and Connecticut, and thereafter he spent one year studying in Germany. He returned to the United States and entered Harvard College, where he studied economics, graduating with a BA in 1906 and an AM in 1907. While studying at Harvard, he served as an assistant to one of the leading figures in orthodox economics, Frank W. Taussig. The next year he entered Harvard Law School, receiving an LLB in 1909. After working at a Chicago law firm and then as a clerk in the legal department of AT&T in New York, Hale enrolled at Columbia University, where he earned a PhD in economics in 1918. As a graduate student at Columbia, he regularly taught courses in the Economics Department. In 1919, the year after he earned his PhD, Hale was invited by Dean Harlan Fiske Stone to teach at Columbia Law School and, as a consequence, Hale was subsequently (in 1922) awarded a joint lectureship in the Economics Department and at the Columbia Law School. He was formally invited to join the law school on a full-time basis in 1928 as an assistant professor of legal economics. He was made full professor in 1935 and retired as professor emeritus in 1949. Hale died on 31 August 1969.

Hale can perhaps best be described as a legal realist who drew upon the emerging tradition of institutional economics (Duxbury, 1995, pp. 107–8). While Richard Posner contends that the legal realists have had little influence on the contemporary mainstream law and economics movement, he does allow that Hale 'anticipated some of the discoveries … of law and economics' as we know it today (Posner, 1995, p. 3). Hale's emphasis on (i) the mutual impact of legal and economic processes and (ii) the terms and concepts used in furthering such an understanding, combined with his unique integration of economics and law, is reflected throughout his work.

The origins of Hale's ideas

While it is difficult to establish conclusively the true origins of an individual's ideas, certain facets of Hale's experiences appear instructive. It is known

that Hale grew up in a very politically conscious family environment during a period of time in which fundamental socioeconomic questions and political policy issues were being widely debated.

As a graduate student in economics at Columbia University, Hale's ideas were partly shaped by three prominent Columbia professors.[2] He attended the seminars of E.R.A. Seligman, who was then a professor of economics and a proponent of the ideas set forth by the German Historical school. Like many young American economists of his day, Seligman had gone abroad and studied in Germany. Upon their return, these scholars tried to convince both the public and their conservative, *laissez faire*-oriented colleagues that there was much to be gained by giving the state an enhanced role in the management of economic affairs and, in particular, that the way to achieve a more efficient allocation of resources was through collective action. Seligman, along with others (especially Richard T. Ely), formed the American Economic Association in 1885 as part of an intellectual rebellion against the self-satisfied domination of orthodox economics (Seligman, 1963, p. 615).

Hale also attended courses given by John Dewey, a professor of philosophy and member of the American pragmatic school of philosophy. Like the other pragmatic philosophers, Dewey recognized the uncertainty inherent in understanding and looked to philosophical methods for a means of establishing the meaning of concepts and beliefs. The pragmatists argued that the analysis of social phenomena should be conducted within systems of relationships among individuals in their empirical settings, necessitating the replacement of a priori abstract reasoning with empirical studies. This led them to the contention that 'true' ideas are those to which responsible investigators would assent after thorough examination – that is, after considering what conceivable effects of a practical kind a theory or object holds. Thus only those hypotheses that contributed to organizing data garnered through sense perceptions related to the real world (that is, those that held practical significance) and did so in a progressive and unifying manner, were taken to be legitimate. In short, an idea was right if it had 'fruitful' consequences. The pragmatist emphasis on the uncertainty inherent in understanding served to provide an epistemological foundation and a social philosophy upon which to erect the basic tenets of institutional economic thought.

Finally, Hale also attended the courses of Thomas Reed Powell.[3] Powell received his LLB from Harvard Law School in 1904 and his PhD in political science from Columbia University in 1913. During his tenure at Columbia, where he was the Ruggles Professor of Constitutional Law, and later at Harvard Law School, where he moved to take up the Langdell Professorship in 1925, Powell's teachings and writings evidenced a blending of law and political science. While a prolific writer (of some 400 articles, reviews and

comments), it is generally accepted that Powell probably had his most significant impact through his teaching.

Like the realists, Powell often criticized the courts, teaching his students that judicial decisions were influenced, even if not always consciously, by judges' social, economic and psychological biases (Braeman, 1988, p. 145). As a professor of constitutional law and a critic of the court, Powell was sympathetic to legal realism, yet he avoided identification with the realist movement, believing that, while social and economic factors exercise a potent influence in human affairs, so too do ideas. For Powell, the realists came too close to perceiving law as some judicial whim or fiat, a position that he felt had nihilistic implications. He asserted that, while the social sciences have much to offer legal analysis, they could never supply definitive answers to the difficult policy issues before the court. Here his Harvard Langdellian legal education still held sway, allowing Powell to argue that (i) the case method of teaching law was *the* way to develop the ability to think and (ii) judicial opinions should satisfy the mind of the trained observer – the law school professor.

The major focus of Powell's interest was the study of the constitutional problems arising out of the federal system and, in addressing these problems, he was always working to explicate the determinative role of judges' personal values and attitudes. He admonished the judges to stop taking refuge behind abstractions and to instead examine the real-world practical consequences of their legal decisions. As described by the dean of Harvard Law School, Irwin Griswald, in 1956, 'In his writings and teachings, Powell laid down the intellectual foundations for the developments in constitutional law which have been seen in the past twenty years' (quoted in Braeman, 1988, p. 146).

Beyond these three professors, Hale was also influenced by the work of Wesley Newcomb Hohfeld. Like John R. Commons, a leading institutionalist economist of the time, Hale was influenced by Hohfeld's articulation of 'fundamental legal conceptions', which, in Hohfeld's mind, were the 'lowest common denominators of law' (Cotterrell, 1989, p. 88). Hohfeld's model of fundamental legal concepts was stated in terms of jural opposites and correlatives that identified one individual's position or condition *vis-à-vis* that of another with respect to (i) right–no-right, (ii) privilege–duty, (iii) power–disability and (iv) immunity–liability (Hohfeld, 1923). For Hale, the Hohfeldian scheme of opposites provided a structure from which to describe the legal relations between persons with respect to advantage and disadvantage and thus with a mutually coercive capacity to act. Hale's analysis of the legal relations between (i) individuals and their employers and (ii) individuals and the state implicitly adopted the Hohfeldian framework (see Hale, 1922a, p. 212; 1927b, p. 523).

The ideas and writings of Oliver Wendell Holmes Jr also had an influence upon the work of Hale. Hale fully concurred with Holmes's famous dissent in

Lochner v. *New York*, where the majority of the US Supreme Court struck down a New York State statute that set the maximum number of working hours for confectionery and bakery workers at ten hours per day or 60 hours per week. The Supreme Court held that the statute violated the due process clause of the 14th Amendment and stated:

> The [legislative] act must have a … direct relation, as a means to an end, and the end itself must be appropriate and legitimate, before an act can be held to be valid which interferes with the general right of an individual to be free in his person and in his power to contract in relation to his own labor. (*Lochner* v. *New York* 198 US 45 (1905) at 57–8)

The case began what has come to be known as 'the *Lochner* era', a period in time (from 1905 to the early 1930s) during which the judiciary relied heavily upon the due process clause of the Constitution to invalidate what were held to be unduly restrictive (and hence arbitrary and capricious) economic regulations. Hale supported Holmes's dissent in *Lochner* (a dissent Holmes later restated in *Tyson & Bro* v. *Banton*[4]), praising it as a 'literary masterpiece' (Hale, 1934a, p. 415). Holmes's famous dissent constituted a critique and rejection of classical economics – specifically, the so-called Spencerian 'natural' phenomena of capitalist economies. Essentially, Holmes's pro-regulation argument, which Hale advocated, was that the legislature may forbid or restrict any business when it has sufficient force of public opinion behind it (*Tyson & Bro* v. *Banton*, 446).

One year before Holmes's dissent in *Lochner*, Thorstein Veblen, the founder of American institutional economics, published *The Theory of Business Enterprise* (1904), which contained a critique of the classical principles of natural liberty and freedom of contract – a critique that had a profound influence on Hale's thinking. Veblen identifies the price system as the leading economic institution in the so-called 'pecuniary economy'. Many contributing factors, most of which are inherent in the economic system (including 'businesslike technicians', 'labour organizations', 'technological advances', and so on) led to economic tensions among the interrelated forces of production and profit. These tensions resulted in the establishment of a class system, a system where the productive class comprised those who were socially productive, and a leisure class, those who came to depend upon acquisition. And, as described by John F. Bell (1967, pp. 548–9) these tensions manifested themselves in 'custodians of absentee-credit' who were certain to engage in 'capitalistic sabotage'; and, further, 'the struggle for economic advantage for their own vested interests' where both labour and business technicians would engage in a 'conscious withdrawal of efficiency'. 'Cultural lags' brought on by technological change altered the institutions and human behaviour resulting in an ever-present class conflict (Srivastava, 1965, pp. 474–

5). Thus, as Veblen described it, it is the very factors within the institution of the price system that lead business to experience fewer intervals of short-term depressions that ultimately gave way to more chronic stagnation. For Veblen, the pecuniary aspects of life are all-pervading and become an integral part of understanding the perpetual unfolding and evolution of society.

Given these influences, it is not surprising that Hale's work represented a challenge to legal and economic orthodoxy – in particular, the dominant tradition of *laissez-faire* capitalism. It was in the merging of these legal and economic influences that his own work achieved its distinctive flavour.

Hale's course, book and research

His course

In the mid-1920s, Hale began experimenting with and developing what may best be termed a legal-economics course for the law school. In 1935, the course – 'Legal factors in economic society' – was formally incorporated into the curriculum at Columbia. The core materials for the course comprised a carefully selected group of articles, legal cases and text that was prepared by Hale. During his tenure at the law school, Hale produced five unpublished editions of the course materials, the longest of which, the third edition in 1940, ran to almost 800 pages.

His book

The evolution of Hale's brand of legal economics is most fully spelled out in his classic book entitled *Freedom Through Law*, which was published in 1952, approximately three years after his retirement from Columbia. As described in the book's Preface, the text of the book 'follows a line of thought which I have been pursuing ever since my graduation from the Harvard Law School in 1910' (Hale, 1952, pp. x–xi). Indeed, much of the book is drawn from the many articles he had previously published and from the *Course Materials for Legal Factors* (Hale, 1935–47).

His research

In this limited overview of Hale's research interests, we shall briefly review the three major elements of his research: (i) his paradigm of the economy as a system of mutual coercion/compulsion, (ii) in that context, his idea of 'private government', and (iii) his writings on the regulation of public utilities.

Hale's paradigm of the economy as a system of mutual coercion/compulsion

In order to appreciate Hale's theory, one must be familiar with his use of certain terms – in particular, the non-pejorative use of the words 'coercion' or 'compulsion'.[5] The most important elements of his theory are the following:

1. *voluntary freedom*, meaning completely autonomous and unconstrained by others with regard to one's behaviour/choices;
2. *volitional freedom*, meaning that behaviour/choices are circumstantially limited by virtue of other individuals' behaviour/choices;
3. *coercion*, meaning the impact of the behaviour/choices of other(s) (an individual or a group) in limiting one's freedom – that is, reducing one's voluntary choice to volitional choice – which also includes the (un)conditional withholding of alternatives;
4. *power*, meaning the capacity to coerce (which may be concentrated or diffused); and
5. *government*, meaning those with the capacity to coerce, that is, to impose constraints on behaviour/choice.

For Hale, voluntary freedom connotes 'the absence of any coercion which keeps you from doing the one [thing] or compels you to do the other' (Hale, 1923a, p. 475). In this context, the concept of voluntary freedom is not a norm by which real-world restrictions on freedom are to be judged; rather, since real-world restrictions are inevitable, the focus is on volitional freedom, which enables real-world restrictions to be seen and analysed. Clearly, for Hale coercion is the means by which voluntary freedom is transformed into volitional freedom by restricting one's behaviour/choice. The restrictions upon choice are of two types: complete and conditional.

One class of actions which constitute a restriction on choice includes those actions that unconditionally constrain the range of alternatives: where the individual or group being constrained has no recourse or control against those coercive impacts. The second class is composed of those actions which conditionally constrain the range of alternatives: where the individual or group being constrained has some control by being able to alter the structure of alternative consequences and tradeoffs. That is, the coercive effects of others' behaviour/choices which are imposed conditionally may be avoided by the affected party by the latter's agreeing to the conditions imposed (implying withholding or the threat of withholding).

In Hale's terms, individuals and groups in a society are subject to a complex structure of governance, a system in which power or the capacity to exert impact – that is, coercion – exists throughout. As a consequence, owing to the proximity of individuals and the corresponding exposure of each to others' behaviour/choices, there is little voluntary freedom in a society. An individual's actual freedom is limited by the aggregate impact of the behaviour/choices of others and therefore is typically volitional and *not* voluntary. From this process emerge solutions to the resource allocation and income distribution questions – solutions which must be conceived of and fundamentally understood in terms of the economic system of mutual coercion.

Much of Hale's writing developed and expanded upon the theme that virtually all legal transactions, including those involving individuals, corporate bodies or the state, constitute and channel compulsion or coercion and thereby reinforce the economic bargaining power of one party *vis-à-vis* another throughout the government and the economy. In Hale's paradigm, legal and economic processes were viewed as inseparable. He described the economy as a structure of coercive power arrangements and relationships, which necessitated an understanding of the formation and structure of the underlying distribution of that economic power. As such, the economy was seen as a system of power operating through a system of coercion, and thus the economic freedom expressed by the courts of the day was merely freedom to engage in economic coercion. Hale did not view this coercion as something to be condemned, but, rather, as a basic fact of economic life. As such, he argued that if, for example, income is in fact the fruit of coercion, abetted actively or passively by government, it cannot be said that overt coercive redistributions of income by government are inherently wrong (Dorfman, 1969, pp. 162–3).

Following Hohfeld's perspective on rights, Hale viewed nearly every statute with economic implications as impacting negatively upon someone's liberty or property. Likewise, following Holmes, Hale believed that the presence of choice within an economy does not necessarily entail the absence of duress. And, following Veblen, Hale on numerous occasions critiqued the concepts of 'natural' liberty and 'freedom' of contract. Given this, he believed that it was essential for the courts to undertake an intelligent balancing of the gains and losses brought about by the particular statutes brought before them – a process which, he said, requires 'a realistic understanding of the economic effect of the legislation' (Dorfman 1969, p. 163). While Hale believed that ethical judgments must ultimately be the basis upon which the court's decisions are made, he felt that the judicial application of economic principles was necessary in order to ascertain the economic consequences – allocative and distributive – of the legislation whose constitutionality the court was asked to evaluate (Hale, 1924, pp. 189–225; 1927a, pp. 131–42).

Private government As noted above, for Hale, the concept of government involved the capacity to coerce; that is, to impose constraints on the behaviour/choices and thus the volitional freedom of others. Thus government is at once both a private and a public phenomenon. In this context, *private government* is the bargaining power that each individual exercises while engaging in the economy as a system of mutual coercion. Government is intrinsically involved in all economic affairs – in the public sphere of course, but no less in the private sphere. Thus Hale writes often of 'the all-pervading role of privately instituted government in the economic sphere' (Samuels, 1973,

p. 296). To Hale, what makes private government 'government' is that it has the same effects on volitional freedom as does 'public', or official, government. In both cases, one's alternative courses of action are subject to the coercive impacts of others. That is, it is through the holding of private property (and other relevant factors) that we are able to exercise private governing power by setting the terms according to which we will allow someone else to acquire or enjoy that which we currently own and control. Duxbury (citing Hale's papers) describes this invisible government, not as a single, coherent unit, but as a cluster of different groups and persons who hold sway in different fields (Duxbury, 1990, pp. 434–5). Nowhere is the essence of this idea better captured than in Edmond Cahn's review of Hale's *Freedom Through Law*, where Cahn wrote:

> [T]hose who own economic goods exercise a kind of governmental power. Being entitled to restrain their property or part with it as they choose, the owners like petty sovereigns can dictate terms and conditions their neighbors must perform in order to have access to the property. In this sense every lawful economic power becomes a type of political power, and every economic inequality poses a question of political inequality. Property so viewed is 'private government'. (Cahn, 1953, p. 14)

For Hale, the so-called 'authoritarian' or 'coercive' aspects of government (conventionally perceived) are everywhere at work and hence are relevant as much within the private sphere as they are alleged to be within the public sphere, where their presence is often castigated.[6] Hale stated the matter clearly:

> We live … under two governments, 'economic' and 'political'. In many matters of everyday life our liberty is restricted by requirements laid down by those who have superior economic power. These stronger persons are not called rulers, or governors, nor are their dictates known as laws or ordinances, however great the pressure which forces obedience. The sway of economic superiors is not thought to be 'government' at all, nor is 'liberty' thought to be curtailed by it. (Quoted in Duxbury, 1990, p. 435)

Finally, the positive facet of Hale's position is simply that volitional freedom is as much a function of private mutual coercion – private government – as it is a function of governmental coercion – public government. In the normative realm, Hale promoted a rough equality of power, coercion and freedom. He recognizes that private and public governing power may be abused and stifle what some may think essential, merely to gratify the whims of those who exert it. Given this, Hale argues that 'the individual liberty of the governed often demands some sort of protection against abuses of private government power, analogous to the safeguards which our constitutional

system furnishes against the abuse of official government. Such safeguards only the official government itself can furnish' (Samuels, 1973, p. 300).

Public utility regulation Hale wrote extensively on the legal and economic theory of rate-base valuation, as well as on the regulation of rate structure and level, and his writings were instrumental in the adoption by the courts of the 'prudent investment' doctrine of valuation for public utilities (Dorfman, 1969, p. 161). His research centred on the 'fair value fallacy' in regulatory law as articulated in *Smyth* v. *Ames* (169 U.S. 466, 1898), where the Supreme Court held that the regulated rates of a public authority were valid if, and only if, they allowed a public service company to earn 'a fair rate of return' on the 'fair value' of its property.

The inherent circularity of the court's reasoning prompted Hale to observe that: 'We can thus find what rates are reasonable by first acertaining the value of the property, and we can ascertain the value of the property only by first finding what rates are reasonable' (Hale, 1939a, p. 517). Hale showed that the court, in relying on the state's power of eminent domain and the Fifth Amendment's just compensation requirement, employed the fair value principle in assessing whether the rates set by public authorities were constitutionally justified. At the same time, the court essentially invoked the state's police power to allow public authorities to reduce rates in the interests of the consumers – which were heretofore 'correct'. As observed by Duxbury, this conflating of the two separate powers resulted in the court allowing the state to both 'constrict the economic initiatives of public utility owners who endeavored to yield as high a profit as feasible from their services', and, at the same time, 'to appropriate the private property of those owners in the form of "excess value" accrued from setting exorbitant rates and prices' (Duxbury, 1990, pp. 426–7).

Having concluded that rates of return cannot be properly calculated independently of value (since value tends to be the capitalization of returns), Hale argued that regulated rates ought to approximate, as nearly as possible, the original costs. However, he recognized that existing conditions – for example, the element of risk, the needed incentive to promote efficient service for consumers, and competition from non-regulated industry for investor funds – might require an allowed rate of return in excess of that justified solely under the original cost (Dorfman, 1969, pp. 161–2).

Hale's impact on law and economics

Hale's ideas and influence are reflected in two distinct movements in modern-day legal jurisprudence.[7] First, there is no doubt that many of Hale's ideas have been influential among those contributing to the critical legal studies movement (Duxbury, 1990). Second, and for more immediate purposes here,

Hale's ideas are prominently evidenced in the core set of ideas that at present comprise American institutional law and economics.[8]

The institutional approach draws no distinction between jurisprudential, legislative, bureaucratic or regulatory treatments. All are seen as particular parts or manifestations of the interrelation of government and the economy, or of legal and economic processes. Instead of concentrating on a unidirectional sequence in which the law or legal structure governs behaviour or conduct in the mixed market economy, which in turn drives economic performance, following Hale, institutional law and economics emphasizes the interpenetration and interrelations between government and the economy: that law or legal structure, behaviour or conduct in the mixed market economy, and economic performance are dependent upon each other through a process of circular and cumulative causation. To various degrees, the influence of Hale's writings is evident within the six major elements of contemporary institutional law and economics.

The evolutionary nature of law and economy

While not eschewing static analysis or denying its value, the role of legal change in affecting the course of this evolution makes the institutional approach to law and economics inherently evolutionary.

Continuity versus change

The recognition of the evolutionary nature of legal–economic relations brings to the fore a second fundamental theme of institutional law and economics: the ever-present tension between continuity and change in legal–economic relations. The evolutionary path of the legal–economic system is derivative of the legal–economic policy *choices* that are made over time. Consistent with Hale, continuity and change are seen as the outcome of the policy-making process – the mutual coercion inherent in the interaction between the groups supporting the respective forces for continuity and change and the power that each can bring to bear on this process. The ongoing choice process within the legal–economic arena determines both the institutional structures that obtain at any given point in time and whether the status quo institutional structures, or some other, will prevail in the future: that is, whether there will be change and, if so, how much.

Mutual interdependence, conflict and the problem of order

The fundamental problem addressed in institutional law and economics is that of order, which is defined as the reconciling of freedom and control, or autonomy and coordination, including hierarchy and equality, with continuity and change. The existence of conflicting interests necessitates both a process (or processes) for deciding between these competing interests and a method (or methods) for determining how these conflicts are to be resolved.

Rights, power and government

Law is fundamentally a matter of rights creation and recreation. Consistent with Hale, government is seen to play a central and inevitable role within this process, for rights are not rights because they are pre-existing, but rather are rights because they are protected by government. The positive, descriptive nature of the institutional approach is concerned with the rights (re)creation process and the impact of this process on legal–economic decision making and activity.

The problematic nature of efficiency

One of the hallmarks of institutional law and economics is the rejection of the Chicago and general neoclassical emphasis on the determination of *the* efficient resolution of legal disputes. Like Hale, the institutionalists believe that more is involved. Whereas the more neoclassical approaches to law and economics seek unique, determinate, optimal equilibrium solutions, institutional law and economics finds such solutions to be question-begging and instead concentrates on identifying and analysing the processes by which the various legal structures, conduct and economic performance are worked out. The institutionalists do not reject efficiency as an important variable in legal–economic analysis, but, rather, maintain that efficiency alone cannot, and should not, determine the assignment of rights.

Towards a comparative institutional analysis

The driving force behind institutional law and economics is the need to come to grips with the interrelations between legal and economic processes. Three efforts, each of which are very much reflective of Hale's approach, are central to this process: (i) models of legal–political and economic interaction must be developed; (ii) objective, positive, empirical studies of government as both a dependent and independent variable, and of economic activity as both an input and an output of political–legal processes, must be undertaken; and (iii) efforts must be made to wed both theoretical and empirical analyses towards a self-consciously objective, positive comprehension of law and economics. Such analysis will serve the twin purposes of deepening the understanding of legal and economic processes and their interrelations, and of providing a more sound basis upon which to predict the potential consequences of legal–economic change.

Conclusion

Thus, consistent with the ideas, the recurring themes and the formulations of the legal realists and institutional economists of the day, Hale's work was very much a challenge to and critique of the dominant tradition of *laissez-faire* capitalism. Hale expressed concern over the distribution and structure of

freedom, coercion and opportunity, held an antipathy towards the unquestioned acceptance of orthodoxy, and maintained an affection for pluralist democracy (Samuels, 1973, p. 272). As compared to the more dominant view, Hale had a very different conception of the interrelations between the economy and the law which led him to emphasize the need for the development of economic and legal theory that would help make clear interconnections between these two domains (Hale, 1923b).

Hale's writings

Books

Hale, Robert Lee (1935–47), *Course Materials for Legal Factors in Economic Society* (1st edn 2 vols, 1935; 2nd edn 3 vols, 1937; 3rd edn 2 vols, 1940; 4th edn 1946; 5th edn 1947).
Hale, Robert Lee (1952), *Freedom Through Law: Public Control of Private Governing Power*, New York: Columbia University Press.

Articles and chapters

Hale, Robert Lee (1917), 'The concentration of wealth: discussion', *American Economic Review*, **7**, 174–5.
Hale, Robert Lee (1920), 'Law making by unofficial minorities', *Columbia Law Review*, **20**, 451–6.
Hale, Robert Lee (1921), 'The "physical value" fallacy in rate cases', *Yale Law Journal*, **30**, 710–31.
Hale, Robert Lee (1922a), 'Rate making and the revision of the property concept', *Columbia Law Review*, **22**, 209–16.
Hale, Robert Lee (1922b), 'Political and economic review', *American Bar Association Journal*, **8**, 752–3.
Hale, Robert Lee (1923a), 'Coercion and distribution in a supposedly non-coercive state', *Political Science Quarterly*, **38**, 470–94.
Hale, Robert Lee (1923b), 'Economic considerations in the restatement and clarification of the law', *Proceedings of the Academy of Political Science*, **10**, 50–54.
Hale, Robert Lee (1924), 'Economic theory and the statesman', in R.G. Tugwell (ed.), *The Trend of Economics*, New York: Knopf, pp. 189–225.
Hale, Robert Lee (1925), 'Labor legislation as an enlargement of industrial liberty', *American Labor Legislative Review*, **15**, 155–60.
Hale, Robert Lee (1925–26), 'Pseudo-protection of property in rate cases', *Michigan Law Review*, **24**, 166–70.
Hale, Robert Lee (1927a), 'Economics and Law', in W.F. Ogburn and A. Goldenweiser (eds), *The Social Sciences and Their Interrelations*, London: Allen & Unwin, pp. 131–42.
Hale, Robert Lee (1927b), 'Value and vested rights', *Columbia Law Review*, **27**, 523–9.
Hale, Robert Lee (1932), 'Labor law – Anglo American', in *Encyclopedia of Social Sciences*, vol 8, New York: Macmillan, pp. 667–72.
Hale, Robert Lee (1934a), 'The constitution and the price system: some reflections on *Nebbia* v. *New York*', *Columbia Law Review*, **34**, 401–25.
Hale, Robert Lee (1934b), 'The new supreme court test of confiscatory rates', *Journal of Land & Public Utility Economics*, **10**, 307–13.
Hale, Robert Lee (1935a), 'Force and the state: a comparison of "political" and "economic" compulsion', *Columbia Law Review*, **35**, 149–201.
Hale, Robert Lee (1935b), 'Unconstitutional conditions and constitutional rights', *Columbia Law Review*, **35**, 321–59.
Hale, Robert Lee (1939a), 'The "fair value" merry-go-round – 1898–1938: a forty year journey

from rates-based-on-value to value-based-on-rates', *Illinois Law Review of Northwestern University*, **33**, 517–31.
Hale, Robert Lee (1939b), 'Our equivocal constitutional guaranties', *Columbia Law Review*, **39**, 563–94.
Hale, Robert Lee (1940), 'Commissions, rates and policies', *Harvard Law Review*, **53**, 1103–44.
Hale, Robert Lee (1942), 'Does the ghost of *Smyth* v. *Ames* still walk?', *Harvard Law Review*, **55**, 1116–40.
Hale, Robert Lee (1943), 'Bargaining, duress and economic liberty', *Columbia Law Review*, **43**, 603–28.
Hale, Robert Lee (1944a), 'The Supreme Court and the contract clause', *Harvard Law Review*, **57** (Clause I, II & III, 512–57, 621–74, 852–92).
Hale, Robert Lee (1944b), 'Utility regulation in the light of the Hope natural gas case', *Columbia Law Review*, **44**, 488–530.
Hale, Robert Lee (1946), 'Prima facie torts, combination and non-feasance', *Columbia Law Review*, **46**, 196–218.
Hale, Robert Lee (1959), 'Book review: *The Idea of Freedom: A Dialectical Examination of the Concept of Freedom*, by M.J. Adler', *Columbia Law Review*, **59**, 821–37.

Notes

1. This overview draws heavily on Samuels (1973) and Duxbury (1990).
2. The influence of these three individuals on Hale's work is documented in Duxbury (1990, p. 423).
3. This discussion of Powell draws on Braeman (1988) and Urofsky (1989).
4. *Tyson & Bro* v. *Banton*, 273 US 418 (1927), where a majority of the US Supreme Court held unconstitutional a New York statute limiting the resale price of theatre tickets to 50 cents in excess of the actual box office price.
5. Hale struggled with using the term 'coercion', considering at times 'compulsion', 'pressure', 'force', 'influence', 'duress' and even 'oppression'. Taken non-pejoratively, coercion, compulsion and force capture the essence of the idea being conveyed.
6. Duxbury (1990, p. 435, n. 127), quoting Hale, has written that 'To the conventional eye … governing power is invisible save when exerted by public officials, wearing the authentic trappings of the political state'.
7. In addition to the themes outlined below, Hale influenced John Maurice Clark's work on social control, as evidenced in Clark (1926 [1939]), and Samuels's (1992) analysis of the economic role of government.
8. More extensive surveys of American institutional law and economics can be found in Mercuro and Medema (1997), which also surveys the various other schools of thought now comprising the field of law and economics, and Medema et al. (1999).

References

Bell, John F. (1967), 'Institutional economics', in *A History of Economic Thought*, New York: Ronald Press, pp. 539–71.
Braeman, John (1988), 'Thomas Reed Powell on the Roosevelt Court', *Constitutional Commentary*, **5** (1), 143–7.
Cahn, Edmond (1953), 'Book review: *Freedom Through Law*', *New York Times*, 18 January, p. 14.
Clark, John Maurice (1926), *Social Control of Business*, Chicago: University of Chicago Press; 2nd edn, New York: McGraw-Hill, 1939.
Cotterrell, Roger (1989), *The Politics of Jurisprudence*, London: Butterworths.
Dorfman, Joseph (1961), *The Economic Mind in American Civilization, Volumes IV and V: 1918– 1933*, New York, Viking Press; reprinted New York: Kelley, 1969.
Duxbury, Neil (1990), 'Robert Hale and the economy of legal force', *Modern Law Review*, **53**, July, 421–44.

Duxbury, Neil (1995), *Patterns of American Jurisprudence*, Oxford: Oxford University Press.
Hohfeld, Wesley Newcomb (1923), *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, ed., W.W. Cook, New Haven, CT: Yale University Press.
Medema, Steven G., Nicholas Mercuro and Warren J. Samuels (1999), 'Institutional law and economics', in B. Bouckaert and G. de Geest (eds), *Encyclopedia of Law and Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 418–55.
Mercuro, Nicholas and Steven G. Medema (1997), *Economics and the Law: From Posner to Post-Modernism*, Princeton, NJ: Princeton University Press.
Posner, Richard A. (1995), *Overcoming Law*, Cambridge, MA: Harvard University Press.
Samuels, Warren J. (1973), 'The economy as a system of power and its legal basis: the legal economics of Robert Hale', *University of Miami Law Review*, **27**, Spring & Summer, 261–371.
Samuels, Warren J. (1992), *Essays on the Economic Role of Government*, Vol. 1, New York: New York University Press.
Seligman, Ben B. (1963), *Main Currents in Modern Economics*, New York: Free Press of Glencoe.
Srivastava, Sik (1965), *History of Economic Thought*, Delhi: Atma Ram & Sons.
Urofsky, Melvin I. (1989), '"Dear Teacher": the correspondence of William O. Douglas and Thomas Reed Powell', *Law and History Review*, **7**(2), 331–86.
Veblen, Thorstein B. (1904), *The Theory of Business Enterprise*, New York: Charles Scribner's Sons.

41 Friedrich August von Hayek (1899–1992)

Ludwig Van den Hauwe

Biographical note

Friedrich August von Hayek, a central figure in twentieth-century economics and a representative of the Austrian tradition, 1974 Nobel laureate in economics, was born on 8 May 1899, in Vienna, then the capital of the Austro-Hungarian empire. Following military service as an artillery officer in the First World War, Hayek entered the University of Vienna, where he attended the lectures of Friedrich von Wieser and obtained doctorates in jurisprudence (1921) and political science (1923). After spending a year in New York (1923–24), Hayek returned to Vienna where he joined the famous *Privatseminar* conducted by Ludwig von Mises. In 1927 Hayek became the first director of the Austrian Institute for Business Cycle Research. On an invitation from Lionel Robbins, he delivered four lectures entitled 'Prices and production' at the London School of Economics in 1931 and subsequently accepted the Tooke Chair. He was a vigorous participant in the heated debates that raged in England during the 1930s concerning monetary, capital and business cycle theories. Hayek was to become the only intellectual opponent of John Maynard Keynes (see Caldwell, 1995). As an outgrowth of his participation in the debate over the possibility of economic calculation under socialism (Hayek, 1948 [1980], 119–208), the focus of Hayek's research shifted during the late 1930s and early 1940s to the role of knowledge and discovery in market processes, and to the methodological underpinnings of the Austrian tradition, particularly subjectivism and methodological individualism. In 1950, Hayek moved to the United States, joining the Committee on Social Thought at the University of Chicago. His research there engaged the broader issues of social, political and legal philosophy. He associated with such figures as Frank Knight, Milton Friedman, Aaron Director and, somewhat later, George Stigler. He returned to Europe in 1962, with appointments at the University of Freiburg, West Germany, and then, in 1969, at the University of Salzburg, Austria. However, in 1977, Hayek moved back to Freiburg, where he died on 23 March 1992.

Hayek was a prolific author not only in the field of economics but also in the fields of political philosophy, psychology, epistemology and legal theory. However, his contributions to social and political philosophy and to legal theory emerged, to a significant degree, as extensions of his scholarship in the field of economics and its methodological foundations. The present entry is

concerned mainly with Hayek's contributions in the fields of legal theory and constitutional economics.[1]

Epistemological foundations

In retrospect and with the benefit of hindsight, it is apparent that Hayek's approach to the social sciences is rooted in a physiologically derived epistemological basis, the foundation of which he worked out during the 1920s. It was further developed in the time thereafter and finally published in 1952 as *The Sensory Order*. Although the book has for a long time been neglected, several commentators have recently recognized its importance (Bouillon, 1991; Streit, 1993; Caldwell, 1997). Hayek himself had already pointed out that his work on *The Sensory Order* had greatly helped him in developing his conception of evolution and of a spontaneous order and in analysing the methods and limits of our endeavours to explain complex phenomena (Hayek, 1979, pp. 199–200, fn 26). One of the central arguments of *The Sensory Order* is that our perception of the world around us is theory guided or conjectural, in the sense that it is informed by a pre-existing system of classification – or set of classificatory dispositions – which is itself the product of a kind of 'learning', the outcome of an evolutionary process that can be said to reflect the accumulated 'experience' of the species. Thus Hayek's views as expounded in *The Sensory Order* come close to some of the tenets of what later became known as 'evolutionary epistemology'. In fact, Hayek's work can be interpreted as a systematic elaboration of the consequences that follow from an evolutionary epistemology for the issue of socioeconomic–political organization. His main subject is the social dimension of the knowledge problem, the problem of social learning; that is, the nature of the process by which knowledge is accumulated and utilized in society.

Hayek's concept of perception as classification has a systematic counterpart in his concepts of rules and rule-following behaviour (see also Hayek, 1967, pp. 43–65). Both our perception and our behavioural responses to situations are a matter of classification. Both are abstract in the sense that we respond not to the unique properties, but to typical features of situations with certain kinds of actions. In both realms, learning is a matter of reclassification. If, at the level of our cognitive apparatus, the existing classification system generates expectations which are disappointed, there will be a tendency for the mind to reclassify experience. The mind will rearrange sensory experiences into new configurations that allow better predictions to be made about reality. Those expectations that are 'fit' tend to survive while those that are 'unfit' tend to be weeded out. This selection process clearly has a counterpart operating at the social level. As will be seen, the role of the judge consists in upholding those rules that will 'maximize' the match-

ing of expectations. Rules in effect draw the demarcation line between 'legitimate' expectations and 'illegitimate' expectations, thus defining the kind of expectations that can be expected to enjoy social protection. On all levels, expectations have a tendency towards coherence and coordination. The criterion of fitness is confirmation of expectations as indicated by the success of our actions. Expectations more consistent with social reality give a competitive advantage to the individuals holding them. Rules are thus valued for instrumental reasons.

To conclude this introductory section about Hayek's epistemology, *The Sensory Order* constituted Hayek's decisive step away from standard economic theorizing. It marked Hayek's further movement away from the equilibrium and rationality constructs used by economists to understand the market order. It paved the way for his rejection of *homo economicus* and ultimately also for the constitutional reorientation of his work, that is, his proposal that we examine the role of various institutions in assisting the creation, discovery, use, conveyance and conservation of knowledge. The mind of the individual described in *The Sensory Order* is a complex adaptive self-organizing neural order. The study of such complex orders led Hayek to criticize particular uses of the equilibrium construct in economics and to question quantitative economics. The newly acquired philosophy also permeates his mature work on legal theory. As he wrote in volume 3 of *Law, Legislation and Liberty*, Hayek (1979), in a section entitled 'The evolution of self-maintaining complex structures':

> These changes in structure are brought about by their elements possessing such regularities of conduct, or such capacities to follow rules, that the result of their individual actions will be to restore the order of the whole if it is disturbed by external influences. …
>
> There is now, in particular, no justification for believing that the search for quantitative relationships, which proved so effective for accounting for the interdependence of two or three different variables, can be of much help in the explanation of the self-maintaining structures that exist only because of their self-maintaining attributes. … In particular, in order to explain the economic aspects of large social systems, we have to account for the course of a flowing stream, constantly adapting itself as a whole to changes in circumstances of which each participant can know only a small fraction, and not for a hypothetical state of equilibrium determined by a set of ascertainable data. And the numerical measurements with which the majority of economists are still occupied today may be of interest as historical facts … With the functions of the system these magnitudes have evidently very little to do. (pp. 158–9)

Spontaneous order and the theory of cultural evolution

The problem of social learning has two aspects which, although they cannot be sharply separated, are nevertheless distinct (see also Hayek, 1960, ch. II, esp. p. 27). First there is the cross-sectional problem of using and communi-

cating the fractional knowledge that is dispersed among the individual contemporaries in a society.

This aspect is the subject of Hayek's theory of the spontaneous order of the market which is best known for its emphasis on the capacity of markets to utilize dispersed knowledge (Hayek, 1948 [1980], pp. 77–91). On account of his work about the social role of prices as carriers of information, allowing the specialized knowledge of each individual to be fully incorporated in decisions concerning resource allocation, Hayek has been considered a precursor of the economics of information (Laffont, 1989, p. 68; Landsburg, 2002, p. 306). Markets are polycentric orders. They can be seen as continuing, open-ended processes of trial-and-error elimination, processes in which constantly a multiplicity of independent trials, of conjectural problem solutions are tried out and selected through the choices of market participants. Through the interaction of experimental exploration and competitive selection, markets can be expected to generate a cumulative growth of problem-solving knowledge. Hayek's notion of competition as a 'discovery procedure' (see Hayek, 1978, pp. 179–90) alludes to this role of the market as an evolutionary learning process.

Hayek's theory of the spontaneous order of the market delivers the insight that, if we want to generate in society any particular order of a certain degree of complexity, we should look for general rules of conduct which, if followed by individuals, would tend to induce that order to form spontaneously. Such a spontaneous or polycentric order rests on a 'division of knowledge' which is analogous with the division of labour in classical economic theory. Because each individual makes use of his or her specific knowledge in deciding how to act, spontaneous orders embody a totality of knowledge that is not known to any single mind. Via the insight that a spontaneous order utilizes much more knowledge than can possibly be made accessible to any central agent or agency, the theory gives us reasons for doubting the ability of governments to achieve complex feats of social organization by deliberate planning. At the same time the theory amounts to an instrumental justification of a particular type of rules. Spontaneous orders emerge out of the interaction of a multiplicity of elements which, in their responses to their particular environment, are governed by certain general rules. The individuals themselves may be unable to articulate the rules they follow. However, the rules of just conduct on which spontaneous orders rely exhibit certain structural characteristics that make them conceptually distinct from rules of organization. They are negative, purpose-independent, abstract, universal and permanent (Van den Hauwe, 1998, p. 101). The character of these rules will permit the inference only of the general features of the overall pattern. The particular content of the resulting order will always be dependent on the specific circumstances to which the elements respond and consequently will be unpredictable. Further-

more, it should be noted that the specification of the criteria that rules must meet in order to allow for the formation of a spontaneous social order remains, equally, very general. In fact, Hayek's account offers only a very general schema which has to be filled in in detail.

The theory of the market, or the catallaxy, is only one part of Hayek's idea of an evolutionary process of collective learning. The second aspect is concerned with the accumulation and growth of knowledge over time, that is, the intertemporal problem of profiting from experiences that previous generations have had. With regard to the intertemporal dimension of the use-of-knowledge problem, Hayek has advanced a theory of cultural evolution. At the core of this theory is the notion that 'the various institutions and habits, tools and methods of doing things, which … constitute our inherited civilization' (Hayek, 1960, p. 62) have passed 'the slow test of time' (Hayek, 1967, p. 111) and can, therefore, be expected to embody the experience of generations. They are, as Hayek argues, the 'product of long experimentation in the past' (Hayek, 1978, p. 136) and 'embody the experience of many more trials and errors than any individual mind could acquire' (Hayek, 1967, p. 88). What is distinctive in Hayek's theory is his account of social institutions and rules of conduct as 'bearers of knowledge' (Kukathas, 1989, p. 220).

The rules and institutions that define the frameworks within which social interactions take place embody knowledge of which we are otherwise unaware because they are themselves the outcome of a process of competitive selection. The abstract frameworks which contribute to the formation of spontaneous orders – basically the rules of the law of property, tort and contract (see Hayek, 1976, p. 109) – are thus themselves conceptualized as more or less unintended products of an evolutionary process. Thus Hayek sees common law both as a codification of previously unarticulated rules of conduct and as providing a framework within which spontaneous orders can form.

As Vanberg points out (Vanberg, 1994, *passim*), Hayek's theory of cultural evolution becomes disputable where it seems to argue that because of our 'incurable ignorance' we ought necessarily to rely largely on unquestioned traditional rules instead of attempting to choose rationally or to construct the system of rules that we want to follow.

The question of how the different kinds of rules differ in their nature – rules of conduct versus organizational rules – must be distinguished from the question of how they originate – whether they 'spontaneously evolve' or are 'deliberately designed'. In a particular sociohistorical situation there may exist a *de facto* correlation between the two aspects, but this need not be so. The two dimensions are conceptually distinct. This fact was acknowledged by Hayek (see Hayek, 1973, pp. 45–6).

The evolution of common law and the role of the judge

A central theme in Hayek's system of thought refers to the interplay between the order of rules and the order of actions. Hayek distinguishes between the legal framework, upon which the relative certainty of expectations is founded, and the system of market exchanges within that framework, in which there is no certainty of expectations. Because of the stability of the legal framework, agents can rely on expectations regarding the typical form or pattern of economic interactions. In contract law, for example, there are criteria for a valid contract regardless of the price or nature of the goods exchanged. But within this overall pattern equilibrium at the level of legal institutions, there is a disequilibrium or continual change in the economic variables. Agents will continually change their plans in accordance with new facts about both the external world and other agents. In fact, the stable legal framework, in facilitating such changes, ensures maximum market coordination. In other words, maximum coordination – the highest possible degree of coordination – does not necessarily mean a state of full or exact coordination. The very process of coordinating must involve a certain amount of (adaptive) discoordination.

Moreover, the stability of the legal framework is not absolute. As will be seen, the rules themselves evolve. Hayek's account of the role of the common law judge is very illuminating in this respect. In Hayek's theory of the common law and the role of the judge, the emphasis is on the coordination of individual activities through a process of systematic mutual adjustment of expectations (Hayek, 1973, p. 86). The function of the judge is to ensure a maximal coincidence of (legitimate) expectations, that is, to create a situation in which the chance to form correct expectations is as great as possible. But the chance of as many expectations as possible being fulfilled will be best enhanced if some expectations are allowed to be systematically disappointed.

Thus the judges, by upholding those rules which make it more likely that expectations will match and not conflict, are consciously trying to give greater internal coherence to the law. However, they do not need to know anything about the nature of the resulting overall order which they serve, beyond the fact that the rules are meant to assist the individuals in successfully forming expectations in a wide range of circumstances. They are unintentionally playing a part in the formation of a spontaneous order: a system of rules of conduct conducive to the efficient operation of the order of actions which rests on it. The body of the common law constitutes a spontaneous order, which evolves as an unintended consequence of the following of meta rules. One of the most fundamental meta rules is that, when deciding a difficult case, the judge's task is to try to make the law as a whole a little more coherent: he is required to think only about the internal logic of the law. Hayek quotes a famous statement by the great eighteenth-century judge Lord

Mansfield, who stressed that the common law 'does not consist of particular cases, but of general principles, which are illustrated and explained by those cases', thus highlighting the fact that a law based on precedent is more rather than less abstract than one expressed in verbal rules (ibid., p. 86).

The judge assists in the process of selection of rules. In fact, three distinct evolutionary mechanisms are involved in Hayek's account of the *modus operandi* of the judiciary. If there were no variation, evolution could not get started as a result of selection. At first sight, however, there seems to be little room for a *variation or mutation mechanism* in Hayek's account of the judge's task. The judge will merely assist in 'the process of articulation of pre-existing rules' (ibid., p. 78). The judge 'is committed to upholding the principles on which the existing order is based' (ibid., p. 120). He discovers the rules 'presumed to have guided expectations in many similar situations in the past' (ibid., p. 86). The judge 'is not a creator of a new order but a servant endeavouring to maintain and improve the functioning of an existing order' (ibid., p. 119). 'The task of the judge will be to tell [the parties in the dispute] what ought to have guided their expectations … because this was the established custom which they ought to have known' (ibid., p. 87). Thus the emphasis is laid on the fact that judges adjudicate particular cases by means of custom and precedent. This is what is meant by *stare decisis*, which can be said to account for the *transmission or replication mechanism* in the evolution of the law.

How then does variation arise? 'Experience will often prove that in new situations rules which have come to be accepted lead to conflicting expectations' (ibid., p. 115). And further: 'Since new situations in which the established rules are not adequate will constantly arise, the task of preventing conflict and enhancing the compatibility of actions … is of necessity a never-ending one, requiring … the formulation of new rules necessary for the preservation of the order of actions' (ibid., p. 119). Thus variation is generated. However, 'This will in some measure always be an experimental process, since the judge … will never be able to foresee all the consequences of the rule he lays down, and will often fail in his endeavour to reduce the sources of conflicts of expectations' (ibid., p. 102); and: 'The judge may err' (ibid., p. 119).

By what mechanism are errors, that is, unfit rules, eliminated?

> [It is] only by their effects on that order of actions, effects which will be discovered only by trial and error, that the adequacy or inadequacy of the rules can be judged. (Ibid., p. 102)

> Like most other intellectual tasks, that of the judge is … one of testing hypotheses at which he has arrived by processes only in part conscious. … he must stand by his decision only if he can rationally defend it against all objections that can be raised against it. (Ibid., p. 120)

> As in all other fields advance is here achieved by our moving within an existing system of thought and endeavouring by a process of piecemeal tinkering, or 'immanent criticism', to make the whole more consistent both internally as well as with the facts to which the rules are applied. (Ibid., p. 118)

Thus a learning process of trial-and-error elimination accounts for the *selection mechanism*.

Law and legislation

Hayek clearly recognized that 'the spontaneous process of growth may lead into an impasse from which it cannot extricate itself by its own forces' (ibid., p. 88). It is therefore acknowledged that 'grown' law may require correction by legislation. It seems that legislation may be required both to generate novelty – that is, legislation works as a mutation mechanism – and to eliminate errors in past developments – that is, it also works as a selection mechanism.

The insight that spontaneous growth will not necessarily operate to generate efficient results is corroborated by the game-theoretic analysis of invisible hand processes. The institutions that develop may be suboptimal in nature and they need not be efficient to persist. For example, in impure coordination games, the players may get stuck in a self-enforcing Nash equilibrium that is not Pareto efficient and thus suboptimal self-sustaining conventions may evolve (Van den Hauwe, 1998, p. 111).

The rule of law

Hayek undertakes to show that the operation of the market leads not only to the creation of an order, but also to a great increase of the return which men receive from their efforts. 'The game of catallaxy' is 'a wealth-creating game (and not what game theory calls a zero-sum game), that is, one that leads to an increase of the stream of goods and of the prospects of all participants to satisfy their needs' (Hayek, 1976, p. 115). However, the tendency of the market to promote welfare is subject to some qualifications. Hayek accepts that the state should provide a safety net of social security provision for the very poor, that it should finance the supply of certain public goods, and that it should impose regulations to control negative externalities. Since, on Hayek's argument, the monopoly on coercion is to be centralized in the hands of government, that is, essentially a monocentric organization, it is crucial that this immense power should not be misused. Government must therefore itself be constrained by general rules, or by what Hayek calls the 'Rule of Law'.

Hayek's doctrine of the Rule of Law is a meta-legal doctrine – a political ideal – a set of standards against which we can judge any laws, regardless of their particular content. The guiding idea is that interventions should, as far as possible, take the form of laying down general rules, which apply equally

to everyone and which must be known and certain. In *The Constitution of Liberty* (Hayek, 1960), Hayek traces the history of the Rule of Law tradition. Moreover, he sets himself the task of answering the question of what liberalism means 'when applied to the concrete problems of our time' (ibid., p. 3). He examines the problems, among others, of welfare statism, labour unions, taxation and transfer payments, money, housing and town planning, agriculture and natural resources, education and research, and 'neighbourhood effects'. This book, in which Hayek developed the ethical, anthropological, legal and economic bases of a liberal economic and social order, is by some considered his *magnum opus*.

In conclusion, it should be clearly understood that through the theory of polycentric or spontaneous orders Hayek grounds the political ideal of the Rule of Law 'positively', that is, on an empirical–scientific basis and not on arbitrary metaphysics.

Proposal for constitutional reform

Hayek became increasingly concerned about the problem of controlling the growth of government. The problem was raised as early as in his *The Road to Serfdom* (Hayek, 1944). This concern finally led him to question democracy as it now exists: the kind of democracy based solely on the principle of majority rule, which has gradually been transformed into a new kind of despotism and under which legislative and administrative powers are increasingly confused. On one hand, omnipotent democracy is quite strong since the individual can hardly escape from its far-reaching coercive power; on the other, however, it is rather weak in that it easily succumbs to the pressure of special-interest groups. Hayek had been particularly impressed by Mancur Olson's (Olson, 1965 [1971]) description of the mechanism of the process of government by coalitions of organized interests and of the tendency towards a persistent exploitation of unorganized and unorganizable groups by organized group interests (Hayek, 1979, pp. 97, 143). Through this analysis, Hayek came to propose a model of an ideal constitution along the following lines (ibid., pp. 105–27).

The task of stating the general rules of just conduct, the rules of action that are followed only to preserve the social order and not to achieve specific targets, would be entrusted to a legislative assembly, the composition of which is different from the governmental assembly entrusted with the task of government. The coercive powers of the latter assembly would be limited by the rules of justice laid down by the first. Thus a true separation of powers would be achieved. The basic clause of such a constitution would have to state that in normal times men could be restrained from doing what they wished, or coerced to do particular things, only in accordance with the recognized rules of just conduct designed to define and protect the individual

domain of each. Practical difficulties are worked out through a special constitutional court.

An application of Hayekian law and economics: the comparative analysis of alternative monetary and banking regimes

It is Hayek's emphasis on the theme of the interrelation between the system of rules and its systematic outcome at the level of the order of actions that qualifies him as a law and economics theorist (see also Hayek, 1967, pp. 66–81). Hayek views the economic problem of society as essentially a coordination problem:

> The essential problem remains that of whether the plans of different individuals will tally and will accordingly all stand a chance of being successful, or whether the present situation carries the seed of inevitable disappointment to some, which will make it necessary for them to change their plans. (Hayek, 1941, p. 22)

Thus in a sense the theme of the interplay between institutional structure and economic order was, at least implicitly, already touched upon in Hayek's early economic work on business cycle theory. Business cycles are instances of how an economy can suffer coordination failures on economy-wide scale. In the so-called 'Austrian' theory of the trade cycle (Hayek, 1935 [1967]), the boom is a self-reversing process set into motion by monetary expansion brought about by the central bank, which is essentially a governmental bank. If the banks were truly competitive, the business cycle would never get under way.

It had long been recognized that the Austrian theory of the business cycle embraces a constitutional perspective with respect to monetary problems: the search is not so much for a proper policy but for an appropriate monetary framework. The cycle depends on the elasticity of bank credit, that is, the characteristic of a developed financial system that allows the supply of money credit to differ from the supply of credit based on real saving.[2] Economists may study different institutional arrangements in order to determine which type of institution is most likely to minimize the tendency for the market rate of interest to be reduced below the natural rate. Monetary and banking arrangements will acquire particular significance in this respect.

As Vera C. Smith reminds us:

> Any attempt to make a final evaluation of the relative merits of alternative systems of banking must look primarily to the tendencies they manifest towards instability, or more particularly to the amount of causal influence they exert in cyclical fluctuations. Most modern theories of the trade cycle seek the originating force of booms and depressions in credit expansions and contractions with the banks as the engineering agencies. (Smith, 1936 [1990], pp. 192–3)

Hayek, however, never endorsed free banking in his early economic writings, despite his general emphasis on the coordinating properties of market competition. Hayek believed that the impulse initiating unsustainable cyclical booms was often the failure of the market rate of interest to rise with the equilibrium or natural rate when the demand for loanable funds increased. To explain why the market rate failed to rise, Hayek elaborated Thomas Joplin's argument as to how commercial banks responded to an increase in loan demand by varying only the quantity of loans and not the price (Hayek, 1935 [1967], pp. 15–17); in other words, the argument is built on the assumption that the short-run supply of bank loans is perfectly elastic at the prevailing rate of interest (Hayek, 1933 [1966], pp. 171–3; White, 1999b).

Therefore a useful central bank, Hayek advised, 'will have to act persistently against the trend of the movement of credit in the country, to contract the credit basis when the superstructure tends to expand and to expand the former when the latter tends to contract' (Hayek, 1937 [1989], pp. 89–90).

Hayek's later 'The denationalisation of money' (Hayek, 1978 [1991]) constituted a radical policy departure from his earlier support for central banking. Hayek, following up on an earlier suggestion by Benjamin Klein (1974), now envisaged a market in which all issuers, public and private, would offer non-redeemable currencies, each currency constituting its own monetary standard. Each private issuer would pledge to maintain purchasing-power stability in terms of a particular basket of goods, but this pledge would not take the form of an enforceable redemption contract. Thus, and consistent with his early scepticism toward free banking, Hayek did not suggest free competition among banks offering wholly or fractionally backed liabilities redeemable for a commodity money.

Although Hayek's proposal with regard to the denationalisation of money has not been exempt from serious criticism,[3] it has led to a variety of proposals for fundamental financial and monetary reform. The development of a contemporaneous free banking school in economics has thus been stimulated (White, 1984 [1995]; Selgin, 1988; Salin, 1998). The advocates of fractional reserve free banking believe that one of the desirable macroeconomic characteristics this system would exhibit is the avoidance, or at least the limitation of Hayekian cycles. Monetary economists Greenfield and Yeager (1983) on the other hand have proposed a multi-commodity standard involving a separation of the medium of account from the medium of redemption, a system which, as these authors believe, would foster price level stability and the avoidance of monetary disequilibrium. More recently the search for institutional alternatives in the domain of money and banking has also led to the development of law-based macroeconomics, a project which culminates in a consistent argument for a return to a 100 per cent reserve requirement in banking (Jesús Huerta de Soto, 1998).[4]

Notes

1. For a more complete intellectual biography of F.A. Hayek, see Hennecke (2000).
2. Several excellent accounts of the theory of the business cycle are available; reference is here made to Garrison (2001) and Skousen (1990). The theory was originally conceived in Mises (1912 [1981], pp. 377–404). Mises's achievement resulted from the integration of different strands of thought: he essentially combined Böhm-Bawerk's capital theory and Wicksell's interest rate mechanism with his own sequence analysis of monetary forces. The theory highlights how changes in the market for loanable funds lead to capital restructuring, and how changes emanating from some sources can lead to unsustainable capital structures that will inevitably require a later correction. An artificial boom is an instance in which the change in the interest rate signal as a result of credit expansion and the change in resource availabilities are at odds with one another. With seemingly favourable credit conditions, long-term investment projects are being initiated at the same time that the resources needed to see them through to completion are being consumed. As the market guides these projects into their intermediate and late stages, the underlying economic realities become increasingly clear. The sequence of malinvestment and overconsumption, followed by crisis, liquidation and unemployment characterizes the intertemporal disequilibrium that is summarily described as a business cycle.
3. More particularly, and in light of the general problem of time-inconsistency identified by macroeconomists the question can be raised of whether the keeping of such a nonenforceable pledge would be consistent with profit maximization. A profit-maximizing fiat-type issuer could choose to hyperinflate its own brand of money, and would do so if staying in business promised less than the one-shot profit available from an unanticipated hyperinflation (White 1999c, 227 ff.). Moreover, the feasibility of private fiat-type money may be doubtful in view of the money regression theorem; on the latter, see Mises (1912 [1981], pp. 129–46) and Selgin (1994, pp. 809–11).
4. Jesús Huerta de Soto (1998) basically demonstrates that Hayek's business cycle theory can be integrated with his legal theory, his constitutional economics and his spontaneous order paradigm. This accomplishment raises doubts as to the merits of the thesis of an author like Witt (1997) who argues that the distinction between Hayek's early investigations into price theory, capital theory, and theory of the business cycle prior to and during his time at the London School of Economics (Hayek I) and, thereafter, his work on social philosophy, spontaneous economic order, and societal evolution (Hayek II) is one between two different and basically incompatible research programmes pursued consecutively by Hayek. It is nevertheless correct that Hayek himself never reconsidered the theory of business cycles from the perspective of his later work. The comparative institutional approach to the study of business cycles is also embraced in Van den Hauwe (2005).

Further reading

Works by F.A. Hayek

Hayek, F.A. (1933 [1966]), *Monetary Theory and the Trade Cycle*, New York: Augustus M. Kelley.
Hayek, F.A. (1935 [1967]), *Prices and Production*, 2nd edn (revised and enlarged), New York: Augustus M. Kelly.
Hayek, F.A. (1937 [1989]), *Monetary Nationalism and International Stability*, Fairfield: Augustus M. Kelley.
Hayek, F.A. (1941), *The Pure Theory of Capital*, London: Routledge & Kegan Paul.
Hayek, F.A. (1944), *The Road to Serfdom*, London: Routledge & Kegan Paul.
Hayek, F.A. (1948 [1980]), *Individualism and Economic Order*, Chicago and London: University of Chicago Press.
Hayek, F.A. (1952), *The Sensory Order*, Chicago: University of Chicago Press.
Hayek, F.A. (1952 [1979]), *The Counter-Revolution of Science*, Indianapolis, IN: Liberty Press.
Hayek, F.A. (1960), *The Constitution of Liberty*, London: Routledge.

Hayek, F.A. (1967), *Studies in Philosophy, Politics and Economics*, London: Routledge.
Hayek, F.A. (1973), *Rules and Order*, vol. 1 of *Law, Legislation and Liberty*, London: Routledge & Kegan Paul.
Hayek, F.A. (1976), *The Mirage of Social Justice*, vol. 2 of *Law, Legislation and Liberty*, London: Routledge & Kegan Paul.
Hayek, F.A. (1978), *New Studies in Philosophy, Politics, Economics and the History of Ideas*, London: Routledge & Kegan Paul.
Hayek, F.A. (1978 [1991]), 'The denationalisation of money – the argument refined', published originally as Hobart Paper Special No. 70, 2nd (extended) edn, IEA (1978), now in *Economic Freedom* (1991), London: Institute for Economic Affairs, pp. 125–235.
Hayek, F.A. (1979), *The Political Order of a Free People*, vol. 3 of *Law, Legislation and Liberty*, London: Routledge & Kegan Paul.
Hayek, F.A. (1988), *The Fatal Conceit – The Errors of Socialism*, London: Routledge.

Selected general reference works and articles related to themes explored in Hayek's thought

Barry, Norman P. (1979), *Hayek's Social and Economic Philosophy*, London: Macmillan.
Birner, Jack and Rudy van Zijp (eds) (1994), *Hayek, Co-ordination and Evolution – His Legacy in Philosophy, Politics, Economics and the History of Ideas*, London: Routledge.
Bouillon, Hardy (1991), *Ordnung, Evolution und Erkenntnis* [Order, Evolution and Knowledge], Tubingen: Mohr (Paul Siebeck).
Butler, Eamonn (1985), *Hayek, His Contribution to the Political and Economic Thought of Our Time*, New York: Universe Books.
Caldwell, B. (1995) (ed.), *Contra Keynes and Cambridge: Essays, Correspondence, The Collected Works of F.A. Hayek Vol. 9*, Chicago: University of Chicago Press.
Caldwell, B. (1997), 'Hayek and socialism', *Journal of Economic Literature*, **35** (4), pp. 1856–90.
Dostaler, Gilles and Diane Ethier (1989), *Friedrich Hayek, Philosophie, économie et politique* [Friedrich Hayek, Philosophy, Economics and Politics], Paris: Economica.
Fleetwood, Steve (1995), *Hayek's Political Economy: The Socio-economics of Order*, London: Routledge.
Frei, Christoph and Robert Nef (eds) (1994), *Contending with Hayek*, Bern: Peter Lang.
Garrison, R.W. (2001), *Time and Money – The Macroeconomics of Capital Structure*, London: Routledge.
Gray, John (1984), *Hayek on Liberty*, Oxford: Basil Blackwell.
Greenfield, R.L. and L.B. Yeager (1983), 'A laissez-faire approach to monetary stability', *Journal of Money, Credit, and Banking*, **15**, 302–15, reprinted in L.B. Yeager (1997), *The Fluttering Veil-Essays on Monetary Disequilibrium*, Indianapolis: Liberty Fund, pp. 363–81.
Hennecke, H.J. (2000), *Friedrich August von Hayek – Die Tradition der Freiheit* [Friedrich August von Hayek – The Tradition of Liberty], Düsseldorf: Verlag Wirtschaft und Finanzen.
Huerta de Soto, J. (1998), *Dinero, crédito bancario y ciclos económicos* [Money, Bank Credit and Economic Cycles], Madrid: Unión Editorial.
Klein, B. (1974), 'The competitive supply of money', *Journal of Money, Credit, and Banking*, **6** (November) 423–53.
Kley, Roland (1994), *Hayek's Social and Political Thought*, Oxford: Clarendon Press.
Kotterman-van de Vosse, I. (1994), *De visie van Hayek, Een pleidooi voor persoonlijke vrijheid* [Hayek's Vision, A Plea for Personal Liberty], Zwolle: Tjeenk Willink.
Kresge, Stephen and Leif Wenar (eds) (1994), *Hayek on Hayek, An Autobiographical Dialogue*, London: Routledge.
Kukathas, Chandran (1989), *Hayek and Modern Liberalism*, Oxford: Clarendon Press.
Laffont, J.-J. (1989), *The Economics of Uncertainty and Information*, Cambridge, MA: MIT Press.
Landsburg, S.E. (2002), *Price Theory and Applications*, 5th edn, Mason, OH: South-Western.
Mises, L. von (1912 [1981]), *The Theory of Money and Credit*, Indianapolis: Liberty Classics.

Nemo, Philippe (1988), *La Société de droit selon F.A. Hayek* [The Lawful Society According to F.A. Hayek], Paris: Presses Universitaires de France.
Nishiyama, Chiaki and Kurt R. Leube (eds) (1984), *The Essence of Hayek*, Stanford, CA: Hoover Institution Press.
Olson, M. (1965 [1971]), *The Logic of Collective Action: Public Goods and the Theory of Groups*, New York: Schocken Books.
Salin, P. (1998), 'Free banking and fractional reserves: a comment', *The Quarterly Journal of Austrian Economics*, **1** (3) (Fall), 61–5.
Selgin, G. (1988), *The Theory of Free Banking*, Totowa, NJ: Rowman & Littlefield.
Selgin, G. (1994), 'On ensuring the acceptability of a new fiat money', *Journal of Money, Credit, and Banking*, **26** (4), 808–26.
Shearmur, Jeremy (1996), *Hayek and After*, London: Routledge.
Skousen, M. (1990), *The Structure of Production*, New York: New York University Press.
Smith, V.C. (1936 [1990]), *The Rationale of Central Banking and the Free Banking Alternative*, Indianapolis, IN: Liberty Fund.
Streit, Manfred (1993), 'Cognition, competition, and catallaxy: in memory of F.A.v. Hayek', *Constitutional Political Economy*, **4** (2), 223–62.
Van den Hauwe, L. (1998), 'Evolution and the production of rules: some preliminary remarks', *European Journal of Law and Economics*, **5** (1), 81–117.
Van den Hauwe, L. (2005), *Foundations of Business Cycle Research*, PhD dissertation, Université Paris IX Dauphine.
Vanberg, Viktor J. (1994), *Rules and Choice in Economics*, London: Routledge.
White, L.H. (1984 [1995]), *Free Banking in Britain*, 2nd edn, London: Institute for Economic Affairs.
White, L.H. (1999a), 'Hayek's monetary theory and policy: a critical reconstruction', *Journal of Money, Credit, and Banking*, **31** (1), 109–20.
White, L.H. (1999b), 'Why didn't Hayek favor laissez faire in banking?', *History of Political Economy*, **31** (4), 753–69.
Witt, U. (1997), 'The Hayekian puzzle: spontaneous order and the business cycle', *Scottish Journal of Political Economy*, **44**, 44–58.

42 Theodor Herzl (1860–1904)

Peter R. Senn

Theodor Herzl was a unique figure in the institutional history of law and economics. He was a pioneer in the application of law and economics. Although neither an economist nor a specialist in law, he combined both fields as he thought about something unique in the history of economics.

He envisaged a new state that was to have distinctive legal and economic systems. Both were to be based on Jewish views about them. The state he proposed was built around the idea that Jews would have a place where they could conduct their lives in accordance with their principles. His writings and leadership led to the only example of the establishment of a country from such a foundation, Israel.

His programme was laid out in a booklet, *Der Judenstaat* (henceforth, *The Jewish State*) published in 1896. A gifted writer, he kept a diary, of which there are many editions. He published much about his vision, including *Old–New Land* (*Altneuland*), a novel designed to inspire faith in the Zionist ideal. *The Jewish State* was the main publication in which he laid out his ideas about law and economics.

Herzl's background

Herzl's contribution and his place in the history of law and economics can only be understood in the light of his background and the intellectual context of his time. His work embodies much of the best of the German and Austrian approach to law and economics. Some facts about Herzl's life help to understand what he did.[1] In all of his works there are strong signs of the influence of both his Viennese and educational background.

Herzl was born in Pesth, now part of Budapest, Hungary. His parents, Jacob and Jeannette Herzl, had moved there four years earlier. The family was well-to-do. Like most of the Jews of Budapest, Germany and the German language played significant roles in their lives. Herzl's mother, particularly, was important in imbuing him with the German cultural heritage.

He had a good conventional early education. Although he was born in Budapest, his Hungarian background was overshadowed by his middle-class, orthodox and cosmopolitan education. When he was eighteen, the family moved to Vienna, where Herzl entered the university. He earned a Doctorate in Law from the Faculty of Law and Political Science (*Facultät der Rechts – und Staatswissenschaften*) of the University of Vienna in 1884. In common

with the German universities of its time, an important function of the faculty of law was as a school for state officials. Vienna was then the capital of the Austro-Hungarian Empire, a multicultural conglomerate.

His courses often dealt with the real problems of governance. Many elements of Herzl's thought can be traced back to the cultural and intellectual milieu of the Austria of the time. Especially strong are the influences of the Habsburg monarchy and its administrative model. That Herzl had no qualms about a corporate body acting authoritatively for the good of all is shown in his discussion of the topic in the section 'Negotiorum Gestio' of *The Jewish State*.

Herzl's background made him a humanist. His humanistic vision is demonstrated by the last sentence of his conclusion: 'And whatever we attempt there to accomplish for our own welfare, will react powerfully and beneficially for the good of humanity' (1946, p. 157). Mark E. Blum has documented the positive influence of Austrian humanism on another aspect of his work (Blum, 1996, p. 175). Austrian humanism looks for the implications of the subject matter on the human condition.

One trait of Herzl's law and economics was based on the Austrian humanist viewpoint about formal cause. This was his allowance for the realization of his model in the light of future uncertainties. As Blum puts it, 'His plan is always one that has stages of development' (ibid., p. 182). This, and his desire to avoid charges of utopianism, are the main reasons why Herzl treated many of the details of his plans with a rather lofty disdain.

From the time when he was expected to accept adult responsibilities as a Jew (Bar Mitzvah) he occupied himself increasingly with the Jewish question. Finally, he devoted his life to the cause of Zionism. His work in law and economics was predicated on his analysis of anti-Semitism. Both his legal and economic proposals had moral and ethical reasons.

The historical context

The time was important. It was more than 50 years before the beginnings of the subdiscipline of law and economics as it later developed. Herzl's inspiration came at a high point in the life of the now classical German Historical school. It was also a time of change in the kinds of economic analysis used by the Germans and Austrians. Carl Menger (1840–1921), the founder of what is today called the Austrian school, was breaking away from his teacher, Gustav von Schmoller (1838–1917). If anachronistic misunderstandings are to be avoided, it is essential to appreciate the approach to the study of law and economics as it had been cultivated in Germany and Austria.

Adam Smith's (1723–90) recognition that there were profound and close relationships between law and economics was expanded and developed by German scholars, notably Wilhelm Georg Friedrich Roscher (1817–94). Adolph

Heinrich Gotthilf Wagner (1835–1917) and Schmoller. Their writings on law and economics were generally centred on situations where the law affected economic activity, such as labour, taxes, tariffs, the value of money and the role of the state. The German studies were mostly descriptive, often theoretical and almost never primarily analytical in the sense that we now understand the term. They had relatively little use for abstract economic theory. They knew, but deplored, the methods of the famous English economists. Although they used marginal analysis, it was in verbal, not mathematical terms. They were pioneers in the economic application and development of statistics.

The German writings were marked by high scholarship standards. The main tools for economic analysis were solid logic, sound judgement and common sense, nearly always firmly based on contemporary and historical data. The use of specifically economic tools, such as comparative cost, was sparing. Often their employment of other tools such as diminishing returns and supply and demand curves was more implicit than explicit.

Distinctions between the application of economic analysis to the workings of the legal system and how the legal system influences the performance of the economic system were relatively infrequent. Schmoller's work, the most influential of its time, emphasized the latter approach, although sometimes studies combined both approaches. Law and economics were more integrated, in the sense that economic studies frequently included consideration of the legal side of the problem.

A conspicuous characteristic of the classical literature on law and economics was that it was almost always based on explicit value judgements about the social effects of the laws and their economic consequences. Most of the German and Austrian writings that were not purely historical were concerned with social problems. Important secondary themes were the effects of the law on economic development and the role of the state in social welfare.

Faith that changing the law would improve welfare was widespread. Almost all the important German and Austrian economists were optimistic about the possibilities of solving social problems with actions from above. They differed in their approaches. Some, like Wagner and Anton Menger (1841–1906), advocated state socialism of various kinds. Others, like Schmoller and Carl Menger, Anton's brother, emphasized changing the institutional preconditions to achieve goals in the public interest. Both Anton and Carl Menger were professors at the University of Vienna when Herzl studied there.

Another characteristic of classical law and economics was the active involvement of economists in attempts to influence public policy. In England, the Anti-Corn Law League in the 1840s led by John Bright (1811–99) and Richard Cobden (1804–65) resulted in the repeal of those restrictive laws. In

Germany, economists were major figures in the organization and operation of the Association for Social Policy (*Verein für Sozialpolitik*) established to influence public opinion. Similar organizations, although none so influential and most more narrowly focused, were formed in France and the United States.

The older style in thinking about social reform was in sharp contrast to the law and economics which developed after the Second World War. In the main, it started from a different approach to social problems. Modern studies in law and economics usually derive their proposals for changes in either the law or economic institutions from an economic analysis. The classical tradition was the other way around. Economists defined the social problem first and then gave the economic, moral and ethical arguments for changing the law.

All of these influences are to be found in Herzl's work.

What Herzl did

Herzl was well aware of the requirement that the new Jewish state be economically viable. As a result, he devoted a great deal of attention to its economic structure. He recognized that what we today call an adequate infrastructure (social overhead capital, roads, railroads and so on) would be needed if the state was to succeed. Here he broke new ground.

The vast bulk of the literature on the problems of industrialization and development assumed an existing state with an infrastructure and a functioning legal and economic system. Herzl could not. Although he had to start from scratch, he found the right approach. Nicholas Balabkins points out that 'Herzl argued that such public investments should not be undertaken in a haphazard manner and in fact insisted that there should be a blueprint for these public investments' (Balabkins, 1996, p. 170).

The economic requirements of the new state as Herzl imagined them, coupled with his own training, required nothing less than a strong hand to guide developments. The direction initially was to come from two corporate bodies, the Society of Jews and the Jewish Company and, afterwards, the state. He outlined the legal structure of the first two and guessed about the last. The Jewish Company furnishes a good example of his method of combining law and economics. He starts by defining its principal function which was to be to 'convert into cash all vested interest left by departing Jews' (Herzl, 1946, p. 98). 'The Jewish Company will be the liquidating agent of the business interests of departing Jews, and will organize commerce and trade in the new country' (ibid., p. 93). The Jewish Company was to be the financial arm of Zionism. He then decides upon an appropriate legal structure. It is to be 'founded as a joint stock company subject to English jurisdiction, framed according to English laws, and under the protection of England' (ibid., p. 98).

Herzl touched upon several kinds of law: civil, criminal, international and commercial. He did not analyse them, but used his knowledge about them for his own purposes. He was concerned with Jewish law which was different from both Continental code law and English law. The differences stemmed, not so much from the religious nature of the Jewish law, but from the fact that the fundamental concepts which are the basis of the law, justice, welfare, equity and the like, were defined very differently. Here, again, he was a pioneer in his recognition of still unresolved problems that stem from different premises about the law and the relationships of different kinds of legal systems to the economic order.

Herzl wanted a rule of law but he was not sure about what kind of legal system would best suit the new state. In discussing immigration he says that it will be 'inaugurated with absolute conformity to law' (ibid., p. 83). During a transition period the settlers 'must act on the principle that every emigrant Jew is to be judged according to the laws of the country which he has left. But they must try to bring about a unification of these various laws to form a modern system of legislation based on the best portions of previous systems' (ibid., p. 147).

Property rights defined in law are fundamental for a modern market economy. Herzl recognized their importance. In discussing the departure of the colonists, Herzl insisted that everything be 'carried out with due consideration for acquired rights, and with absolute conformity to law' (ibid., p. 150). He is concerned that no property held by Jews should be lost (ibid., p. 79). He is also concerned that the emigrants should not defraud those in the countries they leave: 'Every just private claim originating in the abandoned countries will be heard more readily in the Jewish State than anywhere else' (ibid., p. 148).

Herzl's recognition and acceptance of many of the principles of the market economy are very evident. He has numerous uses for them, for example in selling land by auction, the distribution of the products of the clothing industry, labour and the creation of demand. He shows some understanding of how they function. But no invisible hand was to be at work. Markets were to be organized so that they furthered Herzl's notion of what an economy should be. He recognized some relationships between risk and profit. 'Financial morality consists in the correlation of risk and profit' (ibid., pp. 100–101). He is not very alert to the conflicts that arise between some kinds of central planning and the workings of a free market.

Herzl did not want to be an economic innovator, but he became one. There is little that is original in the specific innovations he proposed. Innovation refers to change, originality refers to the quality or fact of being original. Many innovations are not original. His method was close to the concept of entrepreneurship described by Joseph Alois Schumpeter (1883–1950). Change

is brought about by new combinations of techniques, practices and ideas. The innovations and adaptations in the legal and economic institutions he proposed were all devised in the light of his understanding of what was required for his dream to come true. His aim was a Jewish state, a homeland for Jews. It was not to be a theocracy. What is original is how he put together the institutions he proposed in order to achieve his aims.

One way was his attempt to link the problems he saw with solutions employing appropriate incentives. He discussed many of the subjects of law and economics, often, as Backhaus says, in a modern way (1996, p. 128). His approach is primarily institutional, but with a greater emphasis on incentive-compatible behaviour than is found in most older institutional analyses. Voluntarism was to play a role. Incentives to follow the plans were to be more those of the carrot than the stick. He found no conflict between his efforts to combine the carrot of incentive-compatible behaviour with his belief that social problems can be solved by actions from above and the stick that is implied. He desired modern forms of social organization in the new country. For example, he wanted to encourage 'large department stores which provide all necessaries of life' (Herzl, 1946, p. 135). His awareness of the need to develop human capital is shown when he proposes that vocational education should be an integral part of the educational system.

Statistics, in the older sense of measurement, play a big role in Herzl's plans. They were to be used for the purpose of making sure that the plans were based on reality, for example in providing for the settlers and the new towns. His legal training, along with his later experience as a journalist, made him acutely aware of the complexities that his plans faced. This awareness found many expressions in his work on law and economics.

There are dozens of passages in which Herzl demonstrates his methods of using law and economics for his purpose. He recognized the potential difficulties of securing territory for the new state. He always assumed it would have to be purchased. In his discussion of the possibilities for obtaining it he often combines political and economic bargaining concepts. He considered the ways the benefits of the new state would come to those who occupied the land on which it was to be founded.

To fully appreciate Herzl's view of economics in the new state, one must understand two other premises. The new state was to utilize the most advanced technology. This, coupled with entrepreneurship, would ensure its success. His confidence in the spirit of enterprise and technological solutions appears in many places besides his discussions of worker education, building and agriculture: 'The word "impossible" has ceased to exist in the vocabulary of technical science' (ibid., p. 151). He was one of the first to recognize the practical influence of technological change on both the law and economic institutions.

Herzl's broad approach required that he consider many economic elements outside the scope of modern law and economics. Among these are incentives to work and its organization, philanthropy, the evolution of social organizations, housing and town planning. Some topics he treated lightly. Women are scarcely mentioned. Of course he was not right about everything. He badly misjudged the Arab reaction to the Israeli occupation. Because so many interests had to be taken into account, Israel took a much longer time to come into being than he expected. Many of his ideas were ignored, for example those about Israel's flag, the seven-hour day, the rigid hierarchical organization of labour, and the barter system.

Herzl and the new law and economics

After the 1950s, developments in economics, combined with growing attention from the legal community, changed the whole idea of the economic analysis of the law. Sharper distinctions than those found in Herzl are now made between the application of economic analysis to the workings of the legal system and the way the legal system influences the performance of the economic system.

Herzl was not critically self-conscious as regards methodology, an element also often missing in law and economics today. His focus on human welfare as well-being, whether measured by production or consumption, is sharply different from that of all but a few of the institutionalists of today. Herzl follows the older tradition in his treatment of welfare. The new law and economics almost always defines it in much more limited, often mathematical, economic terms than did the classical writers. He realized that there were systematic relationships between legal institutions and the character of economic life. He used his understanding of the interrelations between legal and economic processes in developing his vision. For Herzl, the historical roots of the problem went a long way towards defining it. His work was normative and presumptive, in contrast with the law and economics of today, but absolutely necessary for his purposes. He appears to have reached his economic decisions without establishing any presumption in favour of or against any particular solution or any general category of solution. He never appeared to consider whether public spending, public goods or other social policies were necessarily inadequate, excessive or exploitative.

The classical study of law and economics could accommodate a much wider range of phenomena than the new. A case in point is Jewish law. Although Herzl says little about it, his booklet is full of its spirit. Jewish *Halakahah* (law) is characterized by its emphasis upon human duties. It does not subscribe to the ethical principle of distributive justice. Responsibilities to the poor, not income redistribution, are what the *Torah* (teaching) prescribes. Much of modern law and economics is devoted to problems of equity

in income distribution and is based on political and legal systems that emphasize human rights and legal remedies. These characteristics limit its application to Jewish and other legal systems, which stress different kinds of ethical imperatives.

In Jewish law, judges are given much more latitude to take into account circumstances that do not easily fit the tight analytical mode required by the new law and economics. They can respect fair dealing and good faith. They can deny legal force to deceitful legal agreements, violations of good morals and misuse of rights. Modern law and economics has not yet been able to include moral and ethical precepts like these in its analytical tool box.

Herzl had to put the law and economics of his time into a mould that would fit the special conditions of a highly diverse group of Jews. When his work is examined in the light of the new law and economics, it is apparent that there are limits to the application of the latter, despite the great progress it has made in many respects.

Herzl and the history of economic thought

There are several reasons why Herzl does not fit into the usual histories of economics. He was not an economist. His focus was on one religious group, unlike conventional economics, which has always been more concerned with classes and nations. He was an early institutionalist of a rather special kind. In modern terms, he was an applied economist in the property rights theory tradition.

His practical outlook and the rejection of profit as the principal economic motivator are but two of the traits he shares with most institutionalists. For him, motives that influence economic behaviour were not necessarily economic. Group behaviour, not price, was a central theme. Most of his economic generalizations were relative to time and place. Custom, habit and law were all important factors in the organization of group life.

Herzl's work is a solid example of quality classical work in law and economics. It is also much more than that. It is also as Jürgen Backhaus has pointed out, 'a remarkable exercise in applied law and economics' (1996, p. 127). His attempt to connect the problems he saw with solutions using compatible incentives is worthy of the best law and economics studies of today.

History has shown that Herzl was right about many things. Israel is in existence. Classical law and economics had its uses. Should not the modern practitioner know and use the appropriate methods of both the old and new? Is it not possible that the older ways of doing law and economics might have modern applications? Could it be that rejecting the classical approach to law and economics might be a bit premature?

Note

1. There is a huge literature on both Herzl and Zionism. For eight early writings about him, see Herzl (1946, p. 158). For a sample of others, see Jozsef Patai (1946), Ludwig Lewisohn (1955), Oscar K. Rabinowicz (1958), Norman Kotker (1972), Desmond Stewart (1974), Amos Elon (1975), Ernst Pawel (1989), Steven Beller (1991) and Jacques Kornberg (1993). The *Herzl Yearbook*s (New York: Herzl Press, 1958–96) have much about him. There have long been bibliographies on Israel and Zionism. For an older one, see Iva Cohen (1970). For an annotated list of books on Zionism currently recommended by some distinguished scholars, see the Internet file: ftp://shamash.nysernet.org/israel/lists/scj-faq/reading-lists/zionism.

References

Backhaus, Jürgen (1996), 'On building a sovereign state from scratch: the law and economics of Theodor Herzl's programmatic work', *European Journal of Law and Economics*, **3** (2), June, 127–30.

Balabkins, Nicholas (1996), 'Providing infrastructure for a future state: reading Herzl's *Der Judenstaat* in 1996', *European Journal of Law and Economics*, **3** (2), June, 167–74.

Beller, Steven (1991), *Herzl*, London: Weidenfeld & Nicolson.

Blum, Mark E. (1996), 'The influence of Austrian humanism in Theodor Herzl's vision of a Jewish State', *European Journal of Law and Economics*, **3** (2), June 175–92.

Cohen, Iva (1970), *Israel: A Bibliography*, New York: Anti-Defamation League of B'nai B'rith.

Elon, Amos (1975), *Herzl*, New York: Holt, Rinehart & Winston.

Herzl, Theodor (1936), *Der Judenstaat: Elfte Auflage*, Berlin: Jüdischer Verlag. (The original German title is *Der Judenstaat; Versuch einer modernen Lösung der Judenfrage*; reprinted as *Der Judenstaat*, Zurich: Manne, 1988. I worked from the 1936 reprint.)

Herzl, Theodor (1946), *The Jewish State: An Attempt at a Modern Solution of the Jewish Question*, with introduction by Louis Lipsky, New York: American Zionist Emergency Council (biography based on the work of Alex Bein); reprinted as *The Jewish State*, New York: Dover Publications, 1988.

Herzl, Theodor (1960), *Old–New Land ('Altneuland')*, trans. from the original German, with revised notes, by Lotta Levensohn and a new preface by Emanuel Neumann, 2nd edn with revised footnotes, New York: Bloch; originally published in 1902.

Kornberg, Jacques (1993), *Theodor Herzl: From Assimilation to Zionism*, Bloomington, IN: Indiana University Press.

Kotker, Norman (1972), *Herzl, the King*, New York: Scribner.

Lewisohn, Ludwig (ed.) (1955), *Theodor Herzl: A Portrait for this Age*, ed. and with an introduction by Ludwig Lewisohn, Preface by David Ben-Gurion, Cleveland: World Publishing Co.

Patai, Jozsef (1946), *Star over Jordan; The Life of Theodor Herzl*, trans. from the Hungarian by Francis Magyar, New York: Philosophical Library.

Pawel, Ernst (1989), *The Labyrinth of Exile: A Life of Theodor Herzl*, New York: Farrar, Straus, Giroux.

Rabinowicz, Oscar K. (1958), *Herzl, Architect of the Balfour Declaration*, New York: Herzl Press.

Stewart, Desmond (1974), *Theodor Herzl*, Garden City, NY: Doubleday & Company.

43 Rudolf von Jhering (1818–92) and the economics of justice

J.L.M. Elders

Rudolf von Jhering was a German legal scholar who departed from the dominant legal science of his time. His first writings were still influenced by the conceptualist jurisprudence in his country, the so-called '*Begriffsjurisprudenz*'. In his main work, *Der Zweck im Recht*, however, published in two volumes between 1877 and 1883 and translated into English under the title Law as a Means to an End, von Jhering developed a social utilitarian principle, maintaining that purpose in law is as important as cause in the physical world. While, according to the nineteenth-century historic school of jurisprudence, law has to be regarded as a mainly irrational code of conduct, law was found, not made, and legislation was less important than custom, von Jhering stressed the fact that man, in order to survive, needs assimilation.

In his famous study about the spirit of Roman law (*Geist des römischen Rechtes*) von Jhering defined Roman law as the system of disciplined ego(t)ism, a concept that was to be developed further in his last work, as we shall see. The main thesis in his last study is that purpose has to be regarded as the source of all law. Therefore, the legal system has to deal with social reality. In this connection we can also mention the proposition of H.S.A. Hart in his *The Concept of Law* (Hart, 1961), where he says: 'Given survival as an aim, law should have a specific content' (p. 189).

Von Jhering based his theory upon the principle of psychic causation: human conduct is ruled by purpose and there is no true human act without meaning. Given this as a starting point, law can be defined as a complex of rules of conduct, maintained by a political authority by means of external coercion in order to secure the essential conditions of life. In German: 'Recht ist der Inbegriff der mittels äußeren Zwanges durch die Staatsgewalt gesicherten Lebensbedingungen der Gesellschaft'. The goal of every legal order thus has to be the protection of the conditions of life of man in society.

Von Jhering was influenced by Darwinism and for this reason he saw ego(t)ism as one of the main motives of mankind. Therefore, taking into account individual selfishness, we do need motives to socialize individual selfish striving in order to make it possible to barter on an equal footing. These motives, according to von Jhering, are reward and constraint or punishment. Without reward or at least the prospect of reward, there would not be a

single motive to act, while, without constraint, there would not be an incentive to keep one's promises if to do so would be against what one saw as one's best interests.

While, in nature, predominance goes to the strongest without exception, apart from the care for the young ones, human society will gradually develop a restriction of prevailing power of the strong towards their fellow men and this might be the result of well-understood self-interest. So the captured enemy will no longer be killed but will be made subservient. Eventually arbitrariness will be ended by legal rules. Social relationships wherein individual members can enjoy their subjective rights have to be regulated by legal rules.

The final development of this process has taken place with the different universal declarations of human rights, a little more than a half-century after the publication of von Jhering's work. One of the main characteristics of von Jhering's theory is, as we shall see, the search for incentives to reconcile self-interest with the interests of other people and of society as a whole. Going more deeply into the subject, we shall see that von Jhering takes as a starting point that human beings in general do act out of self-interest. Thus he is asking how society can survive if its members just strive for their individual interest without taking into account the legitimate interests of other people. The answer has to be that human egoism can be made subservient to the general interest of society by engaging and paying the individual the remuneration he is asking. So the egoist will develop a reasonable interest in the realization of the ends of his fellow men if he were to be paid for it.

If human conduct which responds to basic human needs was not rewarded in one way or another, whether by gaining pleasure, profit or social esteem, nature could not achieve its aims for want of incentives to act in conformity with these basic human needs. Only by reconciling self-interest and the interests of others can a foundation be laid for barter, commerce and social intercourse. Any private association, firm or company has to be based on the common interest of its members, partners or shareholders. Also any bilateral agreement presupposes the satisfaction of expectations created by mutual promises.

Therefore, when the law enforces a promise, it does so in order to secure the promisee's reasonable expectations of due performance. Von Jhering takes as a simple example the case of a mill owner who wants to extend his factory and for that reason needs a piece of his neighbour's land and so starts bargaining. If the price to be paid by the mill owner does not exceed a profitable return for him and if the landowner is adequately compensated for the loss of his plot, an agreement is likely to be entered into. Von Jhering stresses that well-understood self-interest of the partners in a bilateral contract is the only sound motive for its formation. Lack of interest on the part of one of the parties, on the contrary, may make a contract voidable. However, von Jhering does not overlook altruistic aspects in human life and conduct.

For instance, if a bilateral agreement implies a strong disproportion between the price to be paid and the end to be achieved, while this would be in conformity with the declared will of the parties concerned, we will have to look for other motives to explain the agreement. Von Jhering therefore makes a clear distinction between egoistic and altruistic motives because human life implies not only personal life but also the survival of the human species. Egoistic motives are physical maintenance, economic maintenance and the assertion of one's individual rights. On the contrary, social motives look beyond these to the interests of social life and the survival of humanity or posterity.

Important for our subject is the way in which von Jhering explains economic maintenance. In human life, he says, past, present and future are all linked up together. When one has to take care of one's future, past experiences will nearly always play an important role. And this will often produce a strong motive to ownership to secure future expectations. For instance, as land grows scarcer as a result of the increase in population, a system of individual property rights tends to develop, as has been showed by Richard A. Posner in his *The Economics of Justice* (Posner, 1983). Property in law has been defined as the closest relation between man and the thing owned, to the exclusion of others.

But the price of property has to be paid, and this price must in general be earned by labour. Von Jhering regards labour as a kind of merchandise in conformity with the prevailing view of his time. Labour will be offered in exchange for money because barter is the basis of all economy. The classic definition of contract, notably the meeting of the will of the parties concerned as to true, full and free consent, is criticized by von Jhering in so far as the economic motives remain undisclosed in this description. The interests of the parties which seem to coincide at the moment of the formation of the contract, may afterwards turn out to be incompatible. In von Jhering's time the hard-and-fast rule '*pacta servanda sunt*' or 'lawfully made contracts do bind the parties' held without exception. Therefore he stressed the importance of the economic purpose of the parties and the necessity of harmonizing their mutual interests.

In modern law we could say that von Jhering had already defended the priority of substance over form, and in this respect he was ahead of his time. In most West European countries, the law of contract in the second half of the twentieth century has shown a development towards fair dealing. A contract in modern legal thinking is a defeasible concept. To allow the enforceability of a contract, several positive conditions have to be fulfilled. Increased recognition of inequalities of bargaining power, knowledge and understanding has led to the legal protection of persons in a weaker position in order to ensure a reasonable fairness between both parties.

The most striking application of the ideas of von Jhering as to the nature of an agreement is to be found in the doctrine of the disappearance of the foundation of a contract which in German civil law is called the doctrine 'des Wegfalls der Geschäftsgrundlage'. For further reading about this subject, see 'Principles of equity in German civil law' (Diederichsen and Gursky, 1973). This doctrine applies to the economic background of bilateral contracts and has developed equitable remedies in the case of lack of consideration.

The halting of individual arbitrariness in business and economic intercourse is one of the most striking results of the state of law. Although our penal codes are not always successful in the struggle for the prevention of crime, there can be no doubt that, without the possibility of prosecuting and sentencing crime, public safety would decrease and the conditions for free negotiations and free bargaining would soon fail to exist. But of no less importance is the possibility of enforcement in civil law. The law of contract in general upholds the agreement and the possibility of suing a party in the case of breach of contract may not only further the motives of both parties for due performance but also serve the economic interests of the community at large.

Economic need will urge man to cooperate with his fellow men but, as von Jhering rightly observes, society cannot survive on benevolence. The basis of every economy, it has to be reiterated, is barter and mutuality while each performance demands a service in return. Even where public offices are without remuneration, they offer power, which can be used or abused, and in most cases they generate social esteem which in von Jhering's eyes can be regarded as moral reward. The principle of mutuality is based on the idea of requital, to be defined as equivalence between good and bad, payment and service in return. In practising barter, the original idea of reciprocity may be expressed clearly but the general acceptance of money as legal currency has made possible an open and free market economy based on reciprocity. Even in cases of temporal shortage of money, different credit systems can fill the gap between performance and the moment of payment. But barter prices are not necessarily the equivalent of the value of goods. Von Jhering, however, considered this an economic problem which could not be solved by law.

The fixing of the price to be paid may result from the bargaining power of the offerer and von Jhering admits that, in this respect, problems may arise because the parties are not always in an equal situation. In von Jhering's time, however, the legal order could not cope with the demands of equity, as it can in ours. The protection of the public, for instance, against economic pressure by means of a new statute law, such as a contract of service act, a rent act or an unfair contract terms act, has only taken definite form in the twentieth century.

No doubt there remain cases in international trade where the weaker party has no recourse but to accept the conditions put by the stronger offerer, but

this does not mean that human need could not be met in a satisfying way by free marketing, given the legal possibilities in cases of unacceptable results. In von Jhering's time, as we have seen, legal corrections in this respect were not at all adequate and therefore, resulting from different social movements and trade unionism, the civil law system in most West European countries has been brought up to date in the course of a half-century. On the other hand, in socialist countries, where the system of free marketing was replaced by centralistic political leadership, the protection of the individual became even less adequate than it was in von Jhering's time. The breakdown of the communist system has made it clear for all to see that, without the motive of economic self-interest, the production of goods sufficient to meet the needs of citizens falls short.

Imputation on the single ground of benevolence fails to do justice to relevant differences between people and will not promote activity. As von Jhering put it, in the long run, justice will be realized better by retaliation in which the principle of reward and punishment is upheld. Therefore, since in modern Western law the drawbacks of ultimate freedom of enterprise and market have been mitigated by adequate statute law, and the protection of the weak and the poor has taken a definite form in a social charter, von Jhering's ideas still hold firm.

While profit can be regarded as a reward of free enterprise, the duty of care towards one's fellow men has been gradually developed in a complicated system of legal rules and judge-made law. Contracts inconsistent with public order or morals are in principle null and void. In case of duress, undue influence, error or fraud, contracts are in general voidable and damages may be granted to the injured party. In the unfair contract term acts, which exist in various countries, there are several legal possibilities or remedies for freeing a victim from the burden and consequences of unfair terms imposed upon him. Finally, there are many legal provisions that seek to protect citizens against the dangers of monopolistic economic power by prescribing free competition and forbidding price fixing.

As we have already seen, von Jhering was fully aware that prices in business do not always correspond to the value of the performance, but value was regarded by him as a strict economic category and for this reason he considered free competition to be the best remedy to secure performances that are worth the money. In theory, weighing up their interests, people should be able to refrain from dealing with monopolistic offerers, but in practice most people will not weigh present against future wants, with the result that they prefer instant satisfaction.

On the other hand, if abuse of economic power should enable a man to gain a very large profit, moral considerations might fail to stop him. Therefore, as von Jhering rightly put it, the only way to prevent this egoistic threat is

recourse to the law. In classic Roman law, an equitable remedy called '*laesio enormis*' had already been introduced. According to this, a sale which turned out to be very disadvantageous for one of the parties because the price of an object was more than twice its value could be rescinded unless the seller consented to adjust the price.

A later development is the so-called '*iustum pretium* doctrine', whereby there should be no disproportion between price and value. However, this doctrine has not been enacted in most civil law systems. On the one hand, it is incompatible with party autonomy and freedom of contract, while, on the other, several legal remedies have been introduced by which, in the event of error, fraud, duress or abuse of power, the injured party may rescind a contract. Apart from this, modern civil law has introduced the possibility of adjusting the terms of a contract if there is a significant change in circumstances that could not have been foreseen when the contract was drawn up.

Summing up, in bilateral contracts there has to be some payment, but it need not be for the full amount. In case of frustration, the courts have a discretionary power to order compensation or to adjust the costs or benefits in cases of total or partial loss of payment. This is what von Jhering thought to be a demand of justice, without which society cannot survive. In his eyes, survival was a value that surpasses individual interest.

Another elaboration of von Jhering's principle of reward and punishment is the way he deals with the idea of requital or retaliation. In his eyes, this concept has been used mainly to indicate punishment or revenge. Claims for reward on account of special achievements barely exist, at least in the Western world. The orders of chivalry to which distinguished persons are admitted by way of honour or reward have to be regarded as a favour, not as the honouring of a right. Of course, social esteem might be seen as a form of moral reward, but the legal order does not recognize it, with one important exception. Harm to one's reputation may give rise to a claim for damages. However, nowadays reward for special achievements in business may be found in profit sharing and high salaries. In so far, von Jhering's principle of reward has found considerable recognition.

Much attention is paid to the importance of constraint in the way von Jhering deals with the bilateral juristic act, which is interpreted by him against the background of barter. Every agreement in which performance and service in return do not coincide in time may make enforcement necessary in the case where one of the parties fails to perform duly or on time. If one of the parties to a contract breaks the obligation imposed upon him by their mutual agreement, a new obligation will arise of right. The justification for enforcement by constraint is to be found in the character of a corresponding offer and acceptance. Von Jhering regarded a mutual promise in dealing as a motor for economic development and this the more so because by means of

bilateral agreement people may draw on their future economic capacities to create new possibilities for the present which might lead to an enormous extension of economic growth.

As we have seen, the words of the offerer with regard to future performance replace immediate settlement, and the expectation of future due performance is secured by the possibility of legal enforcement in the case of breach of contract. Von Jhering points out that legal scholars in the past had great difficulty in explaining and justifying the binding character of mutual agreements. His explanation is that, in former times, the subjective nature of the agreement was overstressed in so far as there was a demand for agreement in every aspect. In this view, the possibilities for failing to uphold an obligation would be almost endless and therefore, as Anson in his *Law of Contract* (Anson, 1982) clearly pointed out, the difficulties inherent in contracts in mere subjective terms have led to the formulation of an objective theory of contract which places more emphasis upon the legal expectations that have been aroused by the conduct of the parties and less on true consensus in every case.

In this respect, von Jhering taught that the purpose of the deal as such is more important than the individual will. If the will should prevail, one could hardly accept that it sought to bind itself for the future with no possibility of withdrawal. In addition, there is usually some kind of ambiguity between two persons during the formation of the contract because their interests will not always coincide. Therefore the core of civil law has to be found in a set of rules that can be interpreted in accordance with objective criteria.

Common survival has to be guaranteed without sacrificing the individual. The choice of the legislator has to be determined by the principle of the best orientation and this means that the law of contract has to promote barter in the most satisfactory way for all. This point can also be illustrated by von Jhering's conception of property. He opposes the classical doctrine of property as the right to use or misuse a thing at will to the total exclusion of others. Von Jhering defended the view that individual property should not frustrate the legitimate interests of society. As to taxation, a system is proposed wherein tax is related to income, but von Jhering seems not to agree with taxation as a political means for the redistribution of wealth.

Looking back on the doctrine elaborated so far, it must be admitted that important incentives such as reward and punishment or constraint are not the only factors by which human behaviour will be affected. There are circumstances in which reward or punishment do not work. In cases of unlimited resources, there would be no need to earn a living or to cooperate with other people. And someone who has the opportunity to enrich himself unjustly at the expense of someone else without the law interfering needs a different incentive to act fairly. This incentive is called morality.

But, at this point, the joint action of economics and the law seems to be powerless – with one exception, the survival of the next generation. Caring for the next generation might further the application of what John Rawls has called 'the just savings principle' (Rawls, 1973, pp. 284–93), in which moral views and economic self-interest are combined. One day, the working generation of the present will become dependent on the next generation and, given the growing life expectancy, the importance of care for each new generation is clear.

This aspect of a moral code can already be found in von Jhering's work. Nowadays von Jhering's thinking about the relation between economics and the law might at first sight seem self-evident. But the roots of more recent thinking about the connection between a prosperous economic development and reasonable legal standards of conduct are already to be found in the writings of Rudolf von Jhering, and more particularly in his principle that law has to be a means to an end.

Summary

Comparing von Jhering's study with Jeremy Bentham's *An Introduction to the Principles of Morals and Legislation* (1780), it might seem that von Jhering did nothing more than use Bentham's 'sovereign motives, pain and pleasure' to found his theory, but this impression would be wrong. Although utilitarian, von Jhering has proved his mastership by reconciling the goals of maximizing wealth and at the same time protecting human autonomy by means of legal institutions and safeguarding the interests of society and future generations.

References

Diederichsen, Uwe and Karl Heinz Gursky (1973), 'Principles of equity in German civil law', in R.A. Newman (ed.), *Equity in the World's Legal Systems*, Brussels: Etablissements Emile Bruylant.

Guest, A.G. (1982), in *Anson's Law of Contracts*, 25th edn, Oxford: Clarendon, pp. 3–6.

Hart, H.S.A. (1961), *The Concept of Law*, Oxford: Oxford University Press.

Posner, Richard A. (1983), *The Economics of Justice*, Cambridge, MA: Harvard University Press.

Rawls, John (1973), *A Theory of Justice*, Oxford: Oxford University Press.

Von Jhering, Rudolf (1877), *Der Zweck im Recht* [Law as a Means to an End], vol 1.

Von Jhering, Rudolf (1924), *Geist des römischen Rechtes*.

44 Franz Klein (1854–1926)

Peter Lewisch

I

Franz Klein, Viennese law professor and Austrian Minister of Justice, is best known as the drafter of the Austrian Civil Procedural Code (ZPO), a codification that has been praised (by Wolf, epilogue to *Zeit- und Geistesströmungen im Prozesse*, see Klein, 1958) as 'the most influential modern civil procedural codification altogether'. Apart from his rich legal work, both doctrinal and legislative, Klein also engaged in politics towards the end of his life, participating in the St Germain peace negotiations and, however unsuccessfully, leading a political party in Austria's 1919 general elections. Since historical details regarding the personality of Franz Klein are well documented in the literature (see in particular the contributions in *Franz Klein Forschungsband*, Hofmeister, 1988), this survey concentrates on those aspects of his legal work that are of interest from a law and economics perspective, most notably his reform of the Austrian civil procedure.

II

Klein was 'Viennese' throughout his life. Born in Vienna, as the son of an artisan, he enrolled in law at the University of Vienna (1872–76), achieved his PhD (*Habilitation*) in civil procedural law (thesis: 'Die schuldhafte Parteihandlung – Eine Untersuchung aus dem Civilprocessrechte', published 1885) in 1883, and subsequently assumed a position in the university administration (*Kanzleidirektor*). This – in itself quite uninteresting – position granted him enough leisure for scientific work. In 1891, Klein was made an associate professor of civil procedural law at the University of Vienna; in the same year his *venia legendi* was extended to Roman law.

In 1890–91, Klein published a series of articles in the *Juristische Blätter* (also published in 1891 as a separate volume under the title *Pro Futuro*) that explicitly addressed shortcomings of the old, then still valid, Procedural Code. He related these shortcomings to deficiencies in the institutional design of the old Code and, on the basis of such analysis, proposed avenues of institutional reform, culminating in relatively detailed suggestions for legal change. In the light of his general high academic reputation, but in particular because of his publications, Klein was offered a position as *Ministerialsekretär* (secretary) in the Austrian Ministry of Justice with the specific task of elaborating and pursuing the reform of the Austrian Civil Procedural Code.

Klein held this position from 1891 to 1897. Unlike many other (also Austrian) examples, where large committees of experts tailored new codifications, the Austrian Civil Procedural Code was basically the sole work of Klein. He had clear ideas of the institutional goals to be accomplished, he drafted the Code himself, and he observed the Code subsequently in practice, commenting on various occasions on the success (and on the problems) emerging with the new Code in actual operation. Although during the years that followed various German universities offered chairs to Klein, he had to turn down all such requests because of his legislative work for the Austrian Ministry of Justice, where he was made a leading state official (*Sektionschef*) in 1895. In 1898, the ZPO was finally enacted.

Klein pursued his legislative work in other areas and twice assumed the position of Minister of Justice (1905–8 and for a short period in 1916). After 1908, he intensified his academic work. Whereas he was highly successful and influential with respect to his theoretical and legislative work, his political efforts failed. Originally proposed as the leader of the Austrian delegation to the St Germain peace negotiations, he was eventually admitted only as its legal counsel and engaged in fruitless disputes with his own Austrian colleagues there. His political ambitions in the Austrian general elections of 1919 for a middle-class democratic party were disappointed; Klein failed to gain a seat in the parliament by only 80 votes. After suffering several strokes over a number of years, he died in 1926.

III

Klein's legislative work was all-encompassing, covering, *inter alia*, the legal right of construction (*Baurechtsgesetz*, 1912), which he considered in relation to state funding of home building, labour law (*Handlungsgehilfengesetz*, 1910), insurance law and company law. Under Klein's ministership (1906), Austria enacted the Act on Limited Liability Companies which, however, was mainly a copy of the pertinent German legislation. Klein contributed to an amendment to the Austrian Civil Code (ABGB) and, moreover, to the then acute problem of agricultural indebtedness with respect to legal constraints of mortgages. He published on legal aspects of commercial papers and served as an expert for the League of Nations on the international standardization of cheque law. With respect to criminal law, beginning in 1905, the ministry enacted several regulations for the protection of minors both in criminal proceedings and in criminal enforcement. He himself pursued the reform of the Austrian Criminal Code, but did not accomplish the enactment of a proposed amendment. He also commented on cartel law and, in his famous 1904 speech, argued against a general prohibition of cartels (Klein, 1927, Briefe I, pp. 29ff.). His most important legislative work, however, remains the reform of the Austrian civil procedure.

IV

Klein's major claim was that a Civil Procedural Code had to be designed such that it would allow for (i) speedy proceedings, (ii) cheap proceedings and (iii) a determination of facts according to actual truth (for an economic analysis of the 'optimal procedure', see Tullock, 1980). It is with respect to these major claims that Klein drafted his new Procedural Code.

The foundations of Klein's legal philosophy and political agenda were eclectic (see Stampfer, 1998). One can identify, though, within his legal thinking, two quite heterogeneous elements. On the one hand, he held dear, as did many of his colleagues at the time, the ideal of the 'welfare state'. Whereas he was born in the days of libertarianism and maintained a deep appreciation for the ideals of the constitutional libertarian state, he was convinced that the state had to play an active role in the provision of elementary services, among them the provision of secular justice through a court system. These judicial services, according to Klein, had to be supplied effectively by the state regardless of the social and financial position of plaintiff and defendant. The intellectual sources of this somewhat 'organic' concept were twofold: first, Klein was influenced by his academic teacher, Anton Menger, whose collective approach to the law was inspired by turn-of-the-century socialism; second, another academic teacher, Emil Steinbach, represented the monarchic, religious and 'organic' (and still socioeconomic) approach to the law. Both influences shaped Klein's idea of seeing the civil procedure as one component within the modern welfare state (for details, see Sprung, 1979). It is also with respect to this general idea that the new Code emphasized the legal responsibility of the judge for the revelation of factual truth and for a lawful ruling. On the other hand, as regards his concrete work as the drafter of the Civil Procedural Code, Klein was not interested in theoretical principles, but in workable solutions. He is therefore identified as a pragmatic reformer, whose main goal was the improvement of the 'practical side of the civil procedure' (Kralik, 1988, p. 89). As such, Klein showed a clear understanding of how legal rules operate. With respect to his analysis of the working properties of legal rules, his methodological approach was individualistic: when discussing the avenues of legal reform, he analysed shortcomings in the actual practice of the old Civil Procedural Code and argued for remedying these deficiencies by deliberate change of the incentives embedded in the legal framework – a change in the legal constraints would generate a change in results.

V

Klein's Civil Procedural Code was a piece of institutional reform. He was well aware of the basic truth in any kind of institutional improvement, namely that there was no feasible way of reform other than starting from the status

quo. This status quo, the previous Austrian Civil Procedural Code, was one strongly characterized by adversarial elements, although not of the Anglo-Saxon type. This old Civil Procedural Code was in its substance still the *Allgemeine Gerichtsordnung* of 1781, amended on several occasions since. In its core provisions, it still reflected the old principles of non-public, non-oral, non-immediate proceedings that granted the parties most of the control over the pursuit of the lawsuit, the judge depending in most of his decisions on their initiative. In particular, the old Code lacked those 'inquisitorial' elements that are characteristic of the current one. Klein criticized the working properties of the old Code for leading to proceedings that were inefficient and slow. In his view, the principles that he advocated (speedy, cheap and truth revealing) were not only goals of a 'good Civil Procedural Code' as such, but also necessary avenues of reform with respect to the shortcomings under the old legislation. One necessary requirement of the 'inquisitorial' type of proceeding was, of course, the oral and 'immediate' (*unmittelbare*) court hearing. This hearing was fully developed in the Austrian Procedural Code only through Klein's reform. The determination of factual truth was, however, of particular importance to Klein. He was determined to accomplish this institutional task by incorporating 'inquisitorial elements' into the still adversarial civil procedure; inquisitorial elements that would grant the judge under the ZPO a far more prominent position than under the German Code.

The above-mentioned elements within Klein's legislative goal where thus interrelated: the oral and immediate hearing contributed both to the revelation of factual truth and to result-oriented speedy hearings. In order to further accelerate the speed of the proceedings, Klein also (in part unsuccessfully, however) reduced competences and therewith related dilatory pleas. He also provided relatively short terms for procedural acts (such as for the defendant's answer to the complaint). Among his, albeit not original, contributions (see Böhm, 1998) was the introduction of the *Neuerungsverbot* in the ZPO (prohibition on presenting new facts to the court in appellate proceedings). It is this *Neuerungsverbot* that also provides a clear difference from the German procedure. The driving force behind this institutional device was Klein's attempt to speed up appellate proceedings by forcing the parties to present the *relevant* factual information at first instance. At the same time, the *Neuerungsverbot* was meant to avoid, with respect to proceedings at first instance, the disadvantages of the (old) principle of contingent claims (*Eventualmaxime*), whereby the parties were forced, at the very beginning of the procedure, to present all theoretically possible factual arguments that either fostered their own claim or weakened the opponent's position. The *Neuerungsverbot* has proved successful: it reduced the duration of appellate hearings without prolonging proceedings at first instance. In his writings (1900, pp. 7ff.) Klein examined empirically the success of his procedural

reform in terms of duration of hearings: he compared the duration of proceedings under the previous and the new procedural regimes, demonstrating the actual acceleration of lawsuits, as brought about by his reform.

VI

As mentioned above, Klein was convinced that, in order to provide for a speedy, cheap and, in particular, truth-revealing procedure, he had to strengthen the 'inquisitorial' elements in the Procedural Code. The pertinent Austrian literature emphasizes (and praises) Klein's concern for a correct determination of facts, in particular by the '*materielle Prozeßleitungspflicht des Richters*' ('obligation of the judge to direct the proceedings in substance'). Its explicit goal was, as stated in the literature (Kralik, 1988, p. 92), the determination of facts according to the factual truth.

Klein's basic approach is well reflected in his own writings: he argues (Klein, 1891, p. 12) that the civil procedure deals with given and objectively existent facts; with facts that, by and large, are something static, namely the result of an already completed development, ('*mit einer bestimmten vollendeten Entwicklung … also wieder mit etwas mindestens derzeit Festem*'). Hence it was his understanding that there existed a factual truth 'out there' to be revealed by the appropriate proceedings (very much along these lines, Klein's ideas were centred around 'the determination of true and false in the proceedings': Klein, 1900, p. 40).

In the light of this, Klein argued that an oral, immediate and 'inquisitorial' type of procedure would guarantee a thorough assessment of all merits of the case (Klein, 1958, p. 26):

> For a thorough hearing is best provided by the intensive judicial guidance which not rarely is successful in distilling within a short time the relevant facts from the entire bulk of claims, so that the entire attention of all individuals concerned may be concentrated on those few decisive points.

On some other occasion, Klein (1900, p. 6) reports that, as a result of the new type of procedure, the facts of the disputed case are clearly reproduced in the presence of all parties concerned and legally determined without delay ('*das Tatsächliche des Streitfalles anschaulich reproduziert und daraufhin ohne Unterbrechnung zugleich festgestellt wird*'). He continues (ibid., p. 9) by arguing that factual truth and law can more easily be perceived in this light ('*in diesem Lichte man nicht bloß Wahrheit und Recht viel leichter erkennt*').

On the other hand, Klein argues (1891, p. 13) that, if the process by which truth is revealed lies entirely within the parties, the degree of such revelation will be 'quantitatively reduced' ('*quantitative Verkürzung*'). In the context of the further development of this argument, he also considers the legal evaluation of the facts thus presented. In this respect, he makes the criticism that the

adversarial system would grant the parties the right to reduce or change (obviously by collusion) the facts of the case, thereby also altering the pertinent legal interpretation. If this was so, the parties would be given the right 'to lead the judge to an erroneous or only relatively correct, or imprecise judgment' ('*den Richter zu einem irrigen oder doch nur relativ richtigen, bzw. ungenauen Urteil zu leiten und zu verhalten*': ibid.). Therefore it is the obligation and duty of the judge to search *ex officio* for the truth with respect to disputed facts (ibid., p. 18). The judge himself or herself should clarify disputed facts (ibid., p. 20).

In our view, a fair interpretation of Klein's approach can be made by comparing it to Gordon Tullock's argument for inquisitorial proceedings: the concepts share the idea that litigation typically involves a clear factual truth 'out there', involving a party who is right (and therefore should win) and a party who is wrong (and therefore should lose) and that in adversarial proceedings resources are invested to mislead (see Parker and Lewisch, 1998).

VII

Centred around this basic idea, Klein amended the Austrian civil procedure in several respects. On the one hand, he introduced the (however unsanctioned) obligation of the parties to produce 'true' claims and statements at court. On the other hand, the concern for the revelation of factual truth in the proceedings found its clearest expression in the so-called '*materielle Prozeßleitungspflicht*' (substantive obligation of the judge to direct the proceedings). This discretionary power of the judge is threefold. First, it contains an obligation to direct (advise) the parties (this obligation would, for example, require the judge to ask a party for clarification before turning down a motion). Second, the judge also has the duty to clarify open factual issues: he may oblige the parties to complete their statements or clarify existing 'inner contradictions' (so that the parties' obligation to produce truthful statements at court is ultimately observed). Third, and most important, the judge enjoys *ex officio* competences to clarify and determine the facts of the case (ZPO, section 183): he may oblige the parties to appear personally at court for purposes of informal discussion of the case; he may oblige the parties to produce pieces of evidence that they have in their hands, namely objects of inspection or documents, provided that the parties have made reference to these documents; he may order the production of existing files (deposited at other courts) or of public documents (deposited with a public notary or public authority) provided that either party has made reference to such documents; he may order the appearance of witnesses at court, and he may require the opinion of experts or order inspection in the presence of the parties. These quite far-reaching competences are limited with respect to (i) documents and (ii) witnesses: if both parties object, the judge may not by himself require this evidence.

The 'inquisitorial' elements in the institutional reform of the then existing procedural law must also be seen in the light of feasible alternatives. In this respect, Klein devoted particular attention to the discussion of 'mutual assistance' of the parties, in particular the legal obligation to provide assistance to the opponent under the adversarial system: he explicitly addressed the legal devices of interrogatories (discovery by interrogatories) as existing under the Anglo-Saxon system. He discussed this issue along lines congenial to contemporary law and economics, in terms of the 'monopolistic position' a party may enjoy with respect to certain facts of the case or the respective evidence ('*Monopolisierung gewisser Sachverhaltsbestandteile und Beweisstücke*'). Under the old Austrian Procedural Code, there was hardly any legal possibility of gathering evidence that was located with the opponent (or a third party). Klein, in turn, introduced a relatively complex, though in substance limited, system of mutual assistance in the new Code.

Most illustrative of the direction of Klein's reform was his conviction of the infeasibility of copying the Anglo-Saxon system and of incorporating it into the then existing Austrian procedure. He considered the concept of mutual assistance of discovery by interrogatories too alien to the Austrian tradition for it to be accepted by the legal community. The infeasibility of reform along these lines necessarily restricted the parties' legal opportunities to search for and to reveal factual truth in the civil lawsuit. In this respect, the 'inquisitorial' elements in the Austrian proceedings provide, to a certain extent, a substitute for the impossibility of a full incorporation of the Anglo-Saxon system.

VIII

If, however, the position of the judge under the ZPO is so prominent, what about selective incentives that would avoid shirking and would induce judges to exercise their rights effectively so that they in fact substantially direct the proceedings? The extent of the involvement of the judge in the pursuit of factual truth does play an important role in the relative attractiveness of the 'inquisitorial' system. This fact has been well recognized by the Austrian doctrinal literature: Kralik (1988, p. 92) argues that the energetic and decisive participation by the judge with respect to the gathering of material and evidence is one of the core issues of the entire procedural reform introduced by Klein. But also Klein himself was aware that the true success of his reform depended on actual judicial involvement in the case. To a certain extent, (however small) there exist direct legal incentives (the 'stick') for the judge: a violation of the judge's duty to direct the parties (pursuant to ZPO, section 182) would allow for successful appeal by that party not properly directed by the judge. However, the judge's pursuit of the case in terms of revelation of factual truth is not subject to direct selective incentives. Klein argued that the

judge's successful direction and management of the case could not have been properly fixed by strict procedural rules: rather, quite the opposite: the more rigorous these procedural rules are, the more passive and guided by mere formalities the judge will be. Despite the necessary dependency on the judge's temperament and individuality, Klein was convinced that the judges would use the opportunity for initiative and personal involvement, as granted under the new Code. Interestingly enough, he became personally involved in the 'instruction' of judges regarding the new Code: he himself gave a series of lectures and published short articles commenting on practical issues in order to familiarize the judges with the Code, its opportunities, duties and, in particular, its reliance on 'an energetic judge in order to accomplish the goals sought'. Whereas Klein argued that judges had to alter their behaviour in order to make the new Code effective, he mentioned on various occasions that, empirically, such change in behaviour was already to be observed (for example, Klein, 1900, p. 57).

IX

Klein's economic thinking has found its expression not only in the incentive-oriented tailoring of the Procedural Code, but also on various other occasions. He addressed, *inter alia*, the relation between macroeconomic growth and the frequency of lawsuits, arguing that an economic upturn would decrease the number of lawsuits (Klein, 1958, p. 28). He also discussed the influence of the interest rate on litigation (ibid., p. 20): if the interest rate was high, then the defendant might be tempted to delay the proceedings on grounds of 'cheap credit'; conversely, if interest rates were low, the *ex lege* interest rate of 4 or 5 per cent, as provided in private law, might even be attractive for the plaintiff in terms of regular investment (for a thorough assessment of the role of interest rates on litigation in a law and economics perspective, see Adams, 1981).

References

Adams, Michael (1981), *Ökonomische Analyse des Zivilprozesses* [Economic Analysis of Civil Procedure], Königstein: Athenäum-Verlag.

Böhm, Peter (1998), 'Was will das Neuerungsverbot? Hintergrund, Funktion, Einfluß auf das Prozeßverhalten in 1. Instanz' ['What is the "Neuerungsverbot" for? Background, function, influence on litigation in the first instance'] in *Bundesministerium für Jusitz*, P. Lewisch and W. Rechberger (eds), *100 Jahre ZPO – Ökonomische Analyse des Zivilprozesses* [100 years of Civil Procedural Code – Economic Analysis of Civil Procedure], Vienna: Manz'sche Verlags- und Universitätsbuchhandlung, pp. 238–48.

Bundesministerium für Justiz, Peter Lewisch and Walter Rechberger (eds) (1998): *100 Jahre ZPO – Ökonomische Analyse des Zivilprozesses*, with contributions by Martin Schneider, Peter Roth, Gerhard Clemenz, Klaus Gugler, Alex Stomper, Bettina Stomper, Walter Rechberger, Peter Lewisch, Erwin Weissel, Paul Oberhammer, Robert Fucik, Jeffrey Parker, Michael Adams, Hans-Bernhard Schäfer, Vienna: Manz'sche Verlags- und Universitäts-buchhandlung.

Fasching, Hans (1988), 'Die Weiterentwicklung des österreichischen Zivilprozeßrechts im Lichte der Ideen Franz Kleins' ['The further Development of the Austrian Civil Procedural Law in the Light of the Ideas by Franz Klein'], in H. Hofmeister (ed.), *Franz Klein Forschungsband*, Vienna: Manz'sche Verlags- und Universitätsbuchhandlung, pp. 97–117.

Hofmeister, Herbert (ed.) (1988), *Franz Klein Forschungsband*, with contributions by Rainer Sprung, Peter Reindl, Winfried Kralik, Hans Fasching, Walter Rechberger, Peter Mayr, Jürgen Damrau, Franz Baltzarek, Fritz Fellner, Peter Böhm, Herbert Hofmeister and Egmont Foregger, Vienna: Manz'sche Verlags- und Universitätsbuchhandlung.

Klein, Franz (1891), *Pro Futuro – Betrachtungen über Probleme der Civilproceßreform in Österreich* [Pro Futuro – Observations on Problems of the Reform of the Civil Procedure in Austria], Leipzig and Vienna: Deuticke.

Klein, Franz (1900), *Vorlesungen über die Praxis des Civilprocesses* [Lectures on the Civil Procedure in Practice], Vienna: Manz'sche Verlags- und Universitätsbuchhandlung.

Klein, Franz (1927), *Reden, Vorträge, Aufsätze, Briefe* [Speeches, Lectures, Articles, Letters], ed. J. Friedländer and O. Friedländer, 2 vols, Vienna: Manz'sche Verlags- und Universitätsbuchhandlung.

Klein, Franz (1958), *Zeit- und Geistesströmungen im Prozesse* [Influences of Prevailing Trends and the General Spirit on the Procedure], Frankfurt am Main: Klostermann, first published 1901.

Kralik, Winfried (1988), 'Die Verwirklichung der Ideen Franz Kleins in der Zivilprozeßordnung von 1895' ['The realization of Franz Klein's ideas in the civil procedural code of 1895'], in H. Hofmeister (ed.), *Forschungsband Franz Klein*, Vienna: Manz'sche Verlags- und Universitätsbuchhandlung, pp. 89–96.

Lewisch, Peter (1998), 'Ökonomische Analyse des Zivilprozesses' ['Economic analysis of the civil procedure'] in P. Mayr (ed.), *100 Jahre österreichische Zivilprozeßgesetze* [100 Years of Austrian Civil Procedural Laws], Vienna: Verlag Österreich, pp. 155–78.

Parker, Jeffrey and Peter Lewisch (1998), 'Materiele Wahrheitsfindung im Zivilprozeß' ['Material truth in civil procedure'], in *Bundesministerium für Jusitz*, P. Lewisch and W. Rechberger (eds), *100 Jahre ZPO – Ökonomische Analyse des Zivilprozesses*, Vienna: Manz'sche Verlags- und Universitätsbuchhandlung, pp. 203–23.

Sprung, Rainer (1977), 'Die Grundlagen des österreichischen Zivilprozeßrechts' ['The foundations of the Austrian civil procedural law'], *Zeitschrift für Zivilprozeßrecht*, **90**, 380.

Sprung, Rainer (1979), 'Die Ausgangspositionen österreichischer Zivilprozessualistik und ihr Einfluß auf das deutsche Recht' ['The starting positions of the Austrian civil procedural doctrine and its influence on German law'], *Zeitschrift für Zivilprozeßrecht*, **92**, 4.

Sprung, Rainer (1988), 'Der Lebensweg Franz Kleins' ['The life of Franz Kleins'], in H. Hofmeister (ed.), *Forschungsband Franz Klein*, Vienna: Manz'sche Verlags- und Universitätsbuchhandlung, pp. 13–78.

Stampfer, Michael (1998), 'Die Zivilprozeßordnung von 1898 vor dem Hintergrund zeitgenössischer sozialer Rechtsgestaltung' ['The Civil Procedural Code of 1898 against the background of contemporary social modelling of the law'], in *Bundesministerium für Jusitz*, P. Lewisch and W. Rechberger (eds), *100 Jahre ZPO – Ökonomische Analyse des Zivilprozesses*, Vienna: Manz'sche Verlags- und Universitätsbuchhandlung, pp. 69–93.

Tullock, Gordon (1980), *Trials on Trial*, New York: Columbia University Press.

45 Etienne Laspeyres (1834–1913)

Wolfgang Drechsler

Introduction

(Ernst Louis) Etienne Laspeyres was Professor *ordinarius* of Economics and Statistics or 'State Sciences' and cameralistics in Basle, Riga, Dorpat, Karlsruhe and, finally, for 26 years in Gießen. Laspeyres, who held doctorates both in law and in economics, was the scion of a Huguenot family of originally Portuguese descent, which had settled in Berlin in the seventeenth century, which has confused pronunciation of his name to this day (see Rinne 1981, p. 196). He himself (probably) pronounced it as if it were German, which comes close to the Portuguese sound 'Lass-pay-ress' (see Meyers 1905, p. 209).

Statistics

Laspeyres is mainly known today for the index number formula for determining the price increase, which he developed in 1871 and which, usually combined with Paasche's, is still in use all over the world (Laspeyres, 1871; the best short summary in Rinne 1983, p. 661; more extensively 1981, pp. 206–209; especially accessible, 1984, pp. 56–61). Other than that, he may count as a father of business administration as an academic-professional discipline in Germany (see Rinne 1983, p. 660; 1981, p. 200), and as one of the main unifiers of economics and statistics by 'developing ideas which are today by and large nationally and internationally reality: quantification and operationalization of economics; expansion of official statistics; cooperation of official statistics and economic research; and integration of the economist and the statistician in one person' (Rinne 1983, p. 660; see Laspeyres 1875, pp. 10, 17–18).

One may even go so far as to say that Laspeyres was one of the leading economists to establish statistics as a key branch of economics as a scholarly discipline (see Stieda 1921, p. 221). For Laspeyres, this was a realism issue – it is, after all, the realist method that needs statistics, not the mathematical-modelling one, where data only disturb the model (see Lexis 1910, p. 244). Laspeyres' main point was that any critique of standard economics would have to be founded on clear results (1875, pp. 8–10). What should be noticed in the law and economics context especially, and what is hardly ever noted when the Laspeyres index number formula is mentioned, is that Laspeyres' point with the formula, the reason why he critiqued an earlier model and

substituted it with his own, is that by doing so, one obtains better statistics on the real situation.

Law and economics being as it is a form of economic realism – a concern with what results a law really has, rather than is supposed to have – Laspeyres' first key contribution to the economics of his time, especially to the emerging specifically continental school of *Kathedersozialismus* and hence to the law and economics tradition that this school forms as well, is the strong emphasis on statistics, correctly gathered and correctly processed, as a necessary basis for any statement concerned with reality.

The History of the Economic Thought of the Netherlanders

The History of the Economic Thought of the Netherlanders and their Literature during the Time of the Republic is arguably Laspeyres' main contribution to economic literature. The work covers the economic literature and history of ideas of the Netherlands from 1600 to 1785, i.e. the time of the greatest extension of Dutch economic and political power, as well as the time of its decline.

A special feature, which was criticized when the book appeared (Vissering 1863, pp. 736–8), is the use, recommended by Roscher (Laspeyres 1863, p. ix), not only of the scholarly but also of the popular and incidental economic literature, such as broadsheets and pamphlets, in which Laspeyres locates the actual strength of Dutch economic theory. This kind of literature had hitherto been unknown, or unused, even in the Netherlands itself (Vissering 1863, pp. 734–5; De Bosch Kemper 1863, pp. 278, 282). Laspeyres' style is colourful and judging, comparable to Karl Bücher or more recently David Landes, and thus, in spite of the generally dry topic, is very readable. Today, especially as a bibliographical guide to its topic, the *History* is still not superseded. Outstanding in two meanings of the word (unusual and excellent) is the *History*'s interweaving of the history of economics and of economic history, partially by way of the use of incidental economic literature. That, together with Laspeyres' decidedly realist – yet not cynical – general attitude, already forms some sort of law and economics framework.

At least in this period of his life, Laspeyres was an avowed, almost militant Smithian who measured all earlier theorists against how similar they are to the 'great Scot' (Laspeyres 1863, *passim*; cf. 1865a; for more moderate later views, see 1875, pp. 6–7). However, his subject and especially his method lead, or almost force, Laspeyres in the end to more or less categorize free trade not as right or wrong as such, but as a figure of argumentation, the use of which – just as that of protectionism – primarily stems from the respective interest of the discussants and the respective circumstances. Thus, neither free trade nor protectionism can be either good or bad for every country and/or every point in time; rather, they are fully context-dependent (see Laspeyres

1863, pp. 87–8, 134–6, 160–62). As an example, high prices are generally bad policy for the country as such (rather than for special interests; a continuous differentiation in Laspeyres, who has a very keen *cui bono* sense in economic policy):

> In the Netherlands, however, high prices were, for special reasons, more easily defensible than in other countries. The small nation was almost throughout producing, and in fact for other countries, from which it was separated by a great advantage in trade and industry. All its inhabitants could see … themselves as producers; thus, the entire surplus of the high prices for the products sold to other countries was to their benefit. (1863, pp. 87–8)

This is similar to, but a further and qualitatively significant development of, Roscher's thesis regarding protectionism, which already counters Smith's, and which is seen as one of Roscher's key law and economics insights (see Streissler on Roscher, Chapter 5 this volume). Another, more concrete example of a law and economics approach in Laspeyres is his summation of what made the Netherlands and their trade so prosperous during their time of flourish. Here, we find – among more traditional items – such reasons as

i) a good justice system, protecting gains 'against the desires of one's fellow citizens',
ii) harsh penalties for criminal bankruptcy,
iii) good insurance companies, distributing risk more evenly, and
iv) contract culture based on a strong *pacta sunt servanda* ethos, which makes small investments into incorporated companies possible and therefore utilizes all the capital available for trade financing. (Laspeyres 1863, pp. 120–23)

Taxation

In addition to his two most important works, one might also mention in the present context a few other publications whose law and economics aspect is more incidental, but which fill the picture of Laspeyres' interest in realism and the effects of regulations in this regard. First, there are two key essays in the highly important, standard-setting *Deutsche Staats-Wörterbuch*, that on 'Staatswirthschaft' (1867) and, in effect, an auxiliary one on 'Staatsmonopole' (1865b).

The very long former article includes the entire general-theoretical treatment of taxes (see 1867, 105 n. *); it therefore basically forms a comprehensive article on public finance. The parts of this entry that deal with state income or revenues (pp. 94–146) could form an excellent basis for an introduction class even today. The segment on taxes (pp. 105–46) deserves special mention, since here, Laspeyres argues for a proportional tax on the 'enjoyments' or

benefits of the individual citizen (p. 106), including – in an almost Estonian way – that there should be no tax on reinvestments, but only on the income from these (pp. 109–10). Although these essays are perhaps the most doctrinaire and unrealist in his oeuvre, realism – as we saw in the case of Smithianism in the *History* – occasionally breaks its path.

Retirement Age

More directly illustrative of law and economics reasoning is Laspeyres' booklet on the question of professorial retirement age (1876), an excellent example of his statistical and analytical skills. Laspeyres analyses the effects of different retirement schemes according to which effect (on students and professors alike) they are supposed to have and which they actually do have, focusing on the question of how to retain good teachers and scholars, even when they are older, while being able to dismiss those who have become unproductive or, worse, outright senile. Although circumstances and possibilities have changed, his answers can still be read with much profit today, or perhaps especially today.

Conclusion

When looking at those of his colleagues included in this *Companion* who were the closest to Laspeyres in time and outlook, then, if compared with Schmoller's thinking about legal-political frameworks (Peukert, Chapter 53 in this volume), or Roscher's about the 'shaping of human behavior and of legal arrangements by production conditions and social institutions' (Streissler, Chapter 51 in this volume), Laspeyres seems to stand up rather well as a classic of law and economics. Sombart's thinking on 'the "spirit" of the system' and his elaboration 'in detail on the evolution of legal forms of enterprise and of business transactions' (Chaloupek, Chapter 55 in this volume), or Lorenz von Stein's on 'a kind of efficiency-oriented law' and 'the comparative analysis of legal structures designed to find out optimal solutions to given economic problems' (Grossekettler, Chapter 56 in this volume), appear to be, on the other hand, much more significant for, and closer to, law and economics.

Still, Laspeyres' insistence on realism in statistics, history, and policy, his use of different sources, the combination of theory and history, and especially his theory-relativizing approach and his interest in the actual effects of a certain regulation, make him, at least for those who hold a European, more inclusive definition of law and economics, one of the approach's (re-) founding protagonists in the nineteenth century.

There are surprisingly few publications about Laspeyres or his work, with the strong exception of the Laspeyres index of his work. His bibliography is in Rinne (1981, pp. 210–14). The *History* (1863) is available in two different reprints, the younger more than twice as expensive as the older (1961 and

1999). Rinne (1983) is by far the best printed source on Laspeyres' life and career; see also Maul (1930). An essay by the author (Drechsler 2000) deals with Laspeyres and law and economics, especially in the *History*; the current entry is based on it, although the former comes to more cautious conclusions. Archival material concerning Laspeyres at the Justus-Liebig-Universität Gießen (University of Gießen), Gießen, Germany, in the Universitätsarchiv (University Archives), PrA Phil 16 (Laspeyres, Etienne) and especially Phil K 20 (Laspeyres, Etienne), as well as at the Handschriftenabteilung der Universitätsbibliothek (University Library, Manuscript Division), e.g., Hs NF 80-1, Hs NF 109-3a, and Hs 139/100-12c.

References

De Bosch Kemper, J. (1863), 'Review of Laspeyres', *Tijdschrift voor staathuishoudkunde en statistiek*, **12**, 271–87.

Drechsler, W. (2000), 'Etienne Laspeyres' *History of the Economic Thought of the Netherlanders*: A law and economics classic?', *European Journal of Law and Economics*, **10**, 235–42.

Laspeyres, E. (1863), *Geschichte der Volkswirthschaftlichen Anschauungen der Niederländer und ihrer Litteratur zur Zeit der Republik*, 'Preisschriften gekrönt und herausgegeben von der Fürstlich Jablonowski'schen Gesellschaft zu Leipzig', 11, Leipzig: Härtel. Reprinted (1961) Nieuwkoop: De Graaf; and (1999) Dillenburg: Wissenschaftliches Antiquariat Lothar Gruber.

Laspeyres, E. (1865a), 'Adam Smith', in *Deutsches Staats-Wörterbuch*, **9**, J.C. Bluntschli and K. Brater (eds), Stuttgart and Leipzig: Expedition des Staats-Wörterbuchs.

Laspeyres, E (1865b), 'Staatsmonopole', in *Deutsches Staats-Wörterbuch*, **9**, J.C. Bluntschli and K. Brater (eds), Stuttgart and Leipzig: Expedition des Staats-Wörterbuchs.

Laspeyres, E. (1867), 'Staatswirthschaft', in *Deutsches Staats-Wörterbuch*, **10**, J.C. Bluntschli and K. Brater (eds), Stuttgart and Leipzig: Expedition des Staats-Wörterbuchs.

Laspeyres, E. (1871), 'Die Berechnung einer mittleren Waarenpreissteigerung', *Jahrbücher für Nationalökonomie und Statistik*, **16**, 296-314, reprinted (1981) **196** (216-)218–37.

Laspeyres, E. (1875), 'Die Kathedersocialisten und die statistischen Congresse. Gedanken zur Begründung einer nationalökonomischen Statistik und einer statistischen Nationalökonomie', *Deutsche Zeit- und Streitfragen*, **52**, Berlin: Lüderitz/Habel.

Laspeyres, E. (1876), 'Das Alter der deutschen Professoren. Ein Beitrag zur Universitätsstatistik und Universitätspolitik', *Deutsche Zeit- und Streit-Fragen*, **5** (74), Berlin: Habel/Lüderitz.

Meyers Großes Conversations-Lexikon 12 (1905), 6th edn, Leipzig and Wien: Bibliographisches Institut.

Lexis, W. (1910), *Die Entwicklung der deutschen Volkswirtschaftslehre im neunzehnten Jahrhundert*, Leipzig: Duncker and Humblot.

Maul, R. (1930), *Laspeyres als statistisch-ökonomischer Forscher*, PhD thesis, University of Frankfurt/Main, Rockenhausen: Buchdruckerei Blaufus.

Rinne, H. (1984), 'Zur 150. Wiederkehr des Geburtstages von Ernst Louis Etienne Laspeyres', *Gießener Universitätsblätter*, **17** (2), 50–61.

Rinne, H. (1983), 'Étienne [Laspeyres]', in *Neue Deutsche Biographie* 13. Berlin: Duncker and Humblot, pp. 660–1.

Rinne, H. (1981), 'Ernst Louis Etienne Laspeyres, 1834–1913', *Jahrbücher für Nationalökonomie und Statistik*, **196**, 194–215.

Stieda, W. (1921), 'Zur Erinnerung an Gustav Schmoller und seine Straßburger Zeit', *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft*, **45**, 219–57 (1115–93).

Vissering, S. (1863), Review of Laspeyres, 1863, *Jahrbücher für Nationalökonomie und Statistik 1*, 732–41.

46 Friedrich List (1789–1846)

Arno Mong Daastöl

Introduction

Friedrich List was one of the earliest and most severe critics of the Classical school of economics, the tradition of the Physiocrats and Adam Smith. His theoretical system is an empirically oriented system, in the sense that he claimed it to be based on historical experience. It is none the less logical and therefore coherent. List is generally known as a proponent of a protectionist, nationalist economic policy and of railroad construction in the early nineteenth century. This is only correct from a superficial point of view, as his fundamental ideas were far wider reaching, dealing with questions such as the ultimate and immaterial basis of economics and of civilization, within a dynamic long-term, global perspective.

List agreed with Smith on the desirability of global free trade. He claimed, however, that instant and radical free trade would lead to a monopoly under the strongest nation, technologically and economically. Other nations therefore had to be lifted up to the level of the leading nation in order to promote the potential wealth of the individual developing nations as well as the global common good. This had to be done gradually through legal and regulatory arrangements nationally and internationally. This would involve, among other instruments, limited and differentiated protection at home and international legal agreements.

List may even be a greater free trader than his main adversary, Adam Smith, in the sense that List's strategy would promote long-term competition to a larger degree than would Smith's strategy, and thereby promote global wealth creation more efficiently. This is a matter of perspective, of time, and of economic complexity, regarding for instance the interrelationship between markets. List would claim that Smith might be said to be a free trader only from a superficial, static short-term and relatively narrow-minded perspective concerning the interests of Britain only, and that this was the deliberate choice of Smith. List, like Smith, had a global perspective and a historical perspective. List, however, claimed to be and seems to have been more aware of deeply rooted international power relations.

List's basic argument against Smith was that his materialist, static and superficial generalizations hid the crucial differences that made different policies in different circumstances necessary (concerning goods, capital, markets, institutions, private versus public interest and historical stage of

development). In particular, this concerned the differences between private versus public interest, between commodities and refined goods and between the levels of development of a nation in all respects. Thereby Smith could overlook the necessity of installing an active government that would create and maintain a policy that differentiates between the above categories. This government would defend the macro point of interest, by contributing to the establishment of regulations and legal arrangements, nationally and internationally. List also argued that Smith and his followers confused causes and effects in their arguments by using non-historical static arguments (List, 1841, pp. 126, 135). This added to the tendency to disregard the need for legal and regulatory intervention in the economic sphere. As regards underdeveloped nations in particular, List alleged that this generally promoted merchant interests contrary to national interests.

List claimed that his economic strategy would also promote the basic and crucial non-monetary factors for economic development that Smith overlooked. His perspective not only paid attention to material factors, as he claimed Smith did, more or less. List, in contrast, saw the immaterial factors as the most important for the development of economics, as well as civilization in general. List was, perhaps, too much of a free trader, often showing too much faith in the withering away of necessary public regulation, being at heart a liberalist, emotionally and politically. Nevertheless, on other occasions he argued rationally for the continued role of government policies. The basic core of List's contribution to economics, or rather political economy, may be said to be that of, 'a prophet of the ambitions of all underdeveloped nations' (Laue, 1963, p. 57).

Although development of human civilization at large was indeed List's main preoccupation, his heart and activity was also devoted to the promotion of more concrete and intermediate matters such as promoting larger markets through economic and political integration, thereby allowing economies of scale and technological efficiency. This was to be realized by political and economic integration and by innovations and investments in activities related to communication, first of all, but also investments in industry and agriculture. More efficient transport systems would foster the power of the individual and of democracy, he believed, and thereby further boost the creativity of the individual. This would again boost scientific, moral and economic progress, and so on. Therefore he emphasized the crucial roles of two (three) phenomena, all intimately related to legal arrangements, arranged here according to importance:

1. political and religious freedom, security and morality. Although a primary goal by itself this also served the next point;
2. arrangements to invoke incentives for and investments in education,

science, research and communication/transport as well as in production in manufacturing and agriculture. This would also serve the former point. The next point was one instrument to achieve the present point;

3. regulation of national trade. Differentiated protection outwards and liberalization inwards, as well as voluntary customs to reap economics of scale benefits.

List's relation to law and economics

The prime instrument in List's strategy was the legal system. Changed legal regulations were to promote social progress, however much they had to be fought through in political and bureaucratic battles. Educated in law and having practised within the administrative and parliamentary system, this was logical to him. List's ideas, in spite of their close affinity to the law and economics approach, have nevertheless been barely noticed among students of law and economics. I shall show that his insights qualify List as an important forefather of the law and economics profession. List argues against a French colleague of Smith, J.B. Say:

> The prosperity of a nation is not, as Say believes, greater in the proportion in which it has amassed more wealth (i.e. values of exchange), but in the proportion in which it has more developed its powers of production. Although laws and public institutions do not produce immediate values, they nevertheless produce productive powers, and Say is mistaken if he maintains that nations have been enabled to become wealthy under all forms of government, and that by means of laws no wealth can be created. The foreign trade of a nation must not be estimated in the way in which individual merchants judge it, solely and only according to the theory of values (i.e. by regarding merely the gain at any particular moment of some material advantage); the nation is bound to keep steadily in view all these conditions on which its present and future existence, prosperity, and power depend.
>
> The nation must sacrifice and give up a measure of material property in order to gain culture, skill and powers of united production; it must sacrifice some present advantages in order to insure to itself future ones. (List, 1841, p. 144)

The relation between law and economics is the *practical core* of his world of ideas and economic efficiency in the widest sense is the goal for his efforts to reform the legal system. He says, 'Every law, every public regulation has a strengthening or weakening effect on production or on consumption or on the productive forces' (ibid., p. 307). Nevertheless, the ultimate goal was of an immaterial and moral nature, closely related to law as well. The prime instrument and crucial tool and lever of his plans for a more humane and efficient economic system was, in fact, law and regulation. He preferred, in many cases, to give law precedence over regulation because of the restrictive effect law has on potential excesses of the bureaucracy. For the same reason, he worked in favour of the jury system, 'one jewel out of the treasure house of

freedom' (ibid., p. 50) in order to correct the arbitrariness and inefficiency of the bureaucracy (as was Gottfried Leibniz's leading idea in his work on legal issues: Anners, 1983, p. 211).

List was educated and practised primarily within the legal system (as a junior and later senior tax clerk, as well as a student and, later, professor) and within law making (as a member of parliament). He was throughout his life devoted to the issue of arranging, first of all, the national and international legal systems so as to serve economic efficiency, in the interest of the common good. He writes:

> Thus the question as to whether, and how, the various nations can be brought into one united federation, and how the decisions of law can be invoked in the place of military force to determine the differences which arise between independent nations, has to be solved concurrently with the question how universal free trade can be established in the place of separate national commercial systems. (List, 1841, p. 114)

National and international legal arrangements were also the main preoccupation of forerunners in Germany, like Nicholas Cusa, Leibniz and Wolff.[1] Legal reform was the essence of List's reform plans as a local civil servant in Württemberg and the essence of his reform plans as a journalist, consultant and politician for a German, European and global legal and economic system. For List there was not a great conflict between justice and efficiency since justice would serve efficiency, and vice versa. He argued that injustice was a major reason for existing economic problems, for instance regarding slavery in the southern states of the United States. In fact, looking at much of List's agenda in opposition to the bureaucracy – freedom of expression, accountability, trial by jury – he might be mistaken for a neoliberal. His opinions on the spiritual origin and character of wealth and prosperity and on the corresponding need for moderate and differentiated regulation make most of the difference. Precisely because he insisted on the important role of regulation, he had all the more reason to be critical of the behaviour of the bureaucracy – like the main thrust of the Ethical-Historical School in economics, such as Max Weber. His personal experiences underline this point.

Surprisingly, it is little realized today, in the so-called 'information age', that List was also the prophet of the economics of communication. His strategy of furthering communication through the establishment of infrastructure in all ideal and material aspects was based on an idealistic image of Man, regarding the human spirit as the ultimate source of wealth and of power. Concerning economics of growth as well as law and economics, this gives List a lead, as he thoroughly discussed the incentive structure in many aspects, and regarded it as crucial for entrepreneurship.

List's background and career

Friedrich was the son of a respected tanner in the Southwest German town of Reutlingen. After a short and uninspiring period in his father's business, he decided to enter the civil service as a clerk and, in 1811, he took up a position in the neighbouring ancient university town of Tübingen. He studied law, part time, and gave up his position in 1813 in order to concentrate on his studies. He never sat for the final lawyer's examination, but instead passed the actuary examination in 1814. He re-entered the civil service as an accountant and was promoted to chief examiner of accounts (ministerial under-secretary) in 1816, under his mentor, Minister von Wangenheim. One product of his studies in law was the treatise on Roman law (List, 1962). He condemned the influence of this law (List, 1841, p. 80).

Apart from numerous reports, List later wrote at length on the importance of agricultural reform for democratization and industrialization. A pioneer paper on land reform is 'Die Ackerverfassung, die Zwergwirtschaft und die Auswanderung' (Agricultural constitution, small business and emigration) (List, 1927–36, vol. 5, pp. 418–547). This is a blueprint of legal reforms in agriculture necessary for higher efficiency in agriculture, higher revenue and therefore surplus for investments in industrialization. But the political aspect of the agricultural constitution was as important. The Japanese and Korean reforms are examples of industrialization starting with land reform. In both countries, the works of List were well known.

In this comparative country study, List elaborates on the ideas of the Cameralist, Justus Möser. List spells out the necessary reforms in order to create the political preconditions for a modern representative parliamentary system. The civil liberties of this system were supposed to be conducive to the creation of an urbanized and industrial society. List advised a golden middle way, creating a class of independent middle-class farmers instead of the British capitalist large-scale type of agriculture. This long-ignored work is perhaps the first systematic work in the historical tradition (with its empirical methodology as opposed to the rationalist introspective methodology of the classical British school) and thereby set a standard for a later method within the historical tradition of economic thought. Marx copied large portions from both this work and from 'The national system' (Lenz, 1930, p. 15).

List was an ardent supporter of the political liberal movement in the German idealistic tradition of Cusa, Leibniz and Wolff. Like them, he acted against the arbitrariness and inefficiency of the bureaucracy and also against the autocratic force of Austria in particular. He put forward numerous practical reports and suggestions for reform of the local inefficient administration and its legal system, some of them actually manifestos for reform. This came to the attention of his superiors and he became the protégé of von Wangenheim, who had similar ideas and intentions, but also made him enemies, especially

within the bureaucracy, to whom he became especially vulnerable after the resignation of von Wangenheim. The time was not yet ripe for these liberal reforms – but when would it ever be without the martyrdom of strong-minded and visionary forerunners? This outcome of List's reform efforts was to haunt him politically and financially for the rest of his life.

He supported the establishment of a chair in economics at the local university of Tübingen. Although he was not appointed to this chair, he was appointed to the full-time chair of political administration (*Staatspraxis*) in 1817, although he had no formal credentials for such a position. In fact, the chair was created for him personally. However, List was too energetic to keep away from life outside academia, and he became involved in the establishment of, and as consular secretary for, a new society for trade and manufacturing (the German Association of Trade and Commerce) in Stuttgart. This aimed at the abolition of internal impediments to trade in Germany, acting against internal customs barriers and for the erection of barriers to foreign manufactured goods, that is, for international customs barriers – a typical (state-) mercantilist programme. This provoked his opponents and gave them an excuse to demand his resignation, partly on the grounds that it was improper for a civil servant to hold a political position and partly because of his absenteeism. List withdrew from his chair in 1819, and thereafter his life was completely devoted to political and economic reform in the service of the common good. His financial position was to be, in consequence, correspondingly bad.

Political struggle: travel and public education

Descriptive of his life's work, as well as of the reasons for the persecution of which he was a victim, is List's claim that the content of his main book, *The National System* (List, 1841), could be summed up in two words: freedom and national unity. One is tempted to add industrialization, although he saw this as an instrument for the above. His main enemies throughout his life continued to be local public bureaucracy, the landed interest and parties of the international power game. These joined ranks to fight him, and did so successfully during his lifetime. Nevertheless, his ideas later bore fruit through his young disciples, who eventually entered into positions of power in Germany and elsewhere, such as in Sweden, Russia, Japan, the United States and China.

In 1819, the citizens of his home town, Reutlingen, elected him as their deputy to the state representative assembly of Würtemberg. He then took part in the fight for a constitution for Würtemberg. These reform activities brought him before the court, and he was tried and convicted for sedition in 1822. He escaped into exile in the same year. He came back to Stuttgart in 1824, when he was imprisoned. In 1825, he was exiled and he emigrated with his family

to the United States following the advice of the Marquis de Lafayette, the French nobleman turned US military general. In the United States he travelled extensively, meeting several leading political figures.

List wrote a series of letters that were published as *Outlines of American Political Economy*[2] (List, 1827), exposing its features and contrasting it with the extreme free trade *Elements of Political Economy* by the southern secessionist, Thomas Cooper. His work was hailed nationally as a textbook of the American system (of economics), as opposed to the British system (non-regulated free trade). His efforts were in part responsible for the introduction of the protectionist US tariff laws of 1828. The conflict between these two strategies was later to be the central reason behind the US civil war where the industrial North was protectionist and allied with Russia and Japan, and the agricultural South was free trade and allied with Britain.

List may therefore be described as an American economist as much as he may be described as a German economist. In 1830, List visited Paris. He returned to Germany in 1832. Back in Europe, he met people in politics, science and the arts. In Paris, in 1837, he wrote two treatises for the French Academy of Sciences, *The Natural System of Political Economy*[3] (List, 1837a) and *The World Moves* (List, 1837b).[4] *The National System of Political Economy* was published in 1841. Until his death in 1846, he wrote extensively for several journals. He travelled extensively on behalf of the Trade Association and met many leading politicians and heads of state. Throughout his life he propagated the same ideas which had followed him from his youth, outlined already in 1816, those of political freedom, political unification and economic development, with railroad construction and protection of manufacture as some instruments among many, all in the tradition of the potential 'harmony of interests'.

List's contribution to economics in general

List's contribution to economic theory may be summarized briefly: his ultimate goal was the moral elevation of global civilization. This could only succeed through industrialization and a corresponding refinement of the individual citizen. He saw the legal system as a lever for this development. He warned that a lack at industry had destabilizing effects (List, 1837a, p. 56). On the other hand, he argued that industrialization would elevate civilization by demanding a highly developed infrastructure and therefore educated and skilled workers with a high moral standard ensuring high-quality conditions of work and trade. High morality and skill would require general welfare. In order to industrialize, any country's government would need to consciously develop the country's infrastructure in all of its ideal and material aspects: its educative, communicative and administrative system, including the legal system, which was indeed to have the pivotal role.

According to List's stage theory of productive powers, which he says he developed after his American experience (List, 1827, p. 161), an industrializing country would have to go through a period of free trade and export of commodities and gradual introduction of industry, followed by a period of protective trade policy, in conjunction with the establishment of protective navigation laws and naval policy. Finally there would be a return to free trade when all economic sectors had been developed.

Urbanization and industrialization for a more humane civilization, individual freedom, democracy and rule of law

In List's mind, economic progress was inseparable from progress in civilization, which in his opinion meant a liberal model according to the British experience (List, 1841, pp. 48–52, 56, 130, 139). His insistent and repeated criticism of cynical British power policies towards its emerging competitor states should not make us forget that Britain was his model country, both in its civilized liberal political regime at home and in its imperial strength. It was List's firm belief that religious and political freedom could be attained only through industrialization and vice versa (ibid., p. 142). This had to be enacted through the legal system, establishing a rule of law, of just and egalitarian law. This spirit runs throughout his writings:

> It has been the experience of all ages and of all countries that freedom and industrial progress are like siamese twins. (List, 1837a, p. 153)

> In the manufacturing State the industry of the masses is enlightened by science, and the sciences and arts are supported by the industry of the masses. (List, 1841, ch. 17, p. 200)

> The greater the advance in scientific knowledge, the more numerous will be the new inventions which save labour and raw materials and lead to new products and processes. (List, 1837a, pp. 66–7; see also pp. 64, 67–9, 79)

List pointed out how manufacturing, as opposed to agriculture, creates a higher potential for diversification of social activity and enhanced possibilities for utilization of individual abilities, especially mental abilities, thereby enhancing and harmonizing equal rights to develop one's abilities and happiness with social welfare and prosperity (List, 1841, ch. 17, p. 200). In the typical German idealist and rationalist Renaissance tradition, as opposed to the materialist Enlightenment tradition and to the irrational Romantic tradition of the nineteenth and early twentieth centuries, List argues for the humanistic and liberating benefits for the individual of the industrially based urban lifestyle (List, 1837a, p. 69). This urban-oriented tradition is very strong throughout German history. We find the same ideas with, for example,

Cusa in the fifteenth century, with Leibniz 150 years later, and with Karl Bücher (1893) and Georg Simmel (1902).

Similarities and differences of approach: List versus law and economics

Some characteristics of List's approach resemble modern law and economics analysis and some elements distance him from this type of analysis. If analysed with modern economic language, there is a lot to be learned from List's insights for students of law and economics, as well as for students of history, development, communication and innovation. Distancing him from the pricing inclination of the law and economics tradition would be his emphasis on the immaterial production factors. Another distancing factor would be his corresponding contempt for the exchange school of Adam Smith working mainly with the monetary aspect of economics, that is to say with the trade aspect, and far less with the productive and creative aspect. List criticized Smith explicitly and repeatedly on this point (for example, List, 1841, ch. 12 – 'The theory of the powers of production and the theory of value'. List regarded the factors that cannot be priced on a market, in particular the immaterial aspects, as the most important factors for the generation of both prosperity and the elevation of culture (List, 1827a, pp. 59, 63, 67; 1841, ch. 12 – 'The manufacturing power and the personal, social and political productive powers of the nation'). To him, market prices were only one practical instrument among many, as part of a larger plan concerning the ultimate goal, the elevation of human culture. Therefore, it is likely that List would have been critical of several characteristics of modern law and economics. This would most likely have been that of pricing a legal arrangement, the reason being that he belonged to the code of the Continental law tradition, where you do not bargain over (semi-religiously) given axioms. He would most probably also have been somewhat critical of the assumption of freedom of choice, since he was critically aware of the path dependencies created by historical power structures.

Nevertheless, the above characteristics of immaterialism and power also constitute factors which unite List with the law and economics tradition. Reminding us of the law and economics approach is List's emphasis on a policy which uses governmental regulation and law making concerning, for example, competition, privileges, taxation and subsidies to promote long-term efficiency in the legal system and also within the economic system.

List's criticism of short-sightedness and the beggar-thy-neighbour policy

List's criticism of the one-eyed exchange value approach ('monetarism') of Smith's followers (List, 1841, pp. 133ff.) did not prevent him from using relative prices as a part of his own analysis of economics. Nevertheless, it

might be argued that he saw this in a wider perspective. List's criticism was generally directed against the inclination to short-term evaluations and the related narrow-mindedness in economic affairs. His criticism had four specific targets: (1) landed interests (particularly England), (2) merchant interests (particularly Holland and Britain), (3) governmental regulation (in general, both the lack of it and its excesses) and (4) international politics (particularly England).[5]

In all these cases, List pointed out the international aspect of the problems. In addition, in all these cases, he insisted in a Socratic manner that the actors who were his targets did not have a sufficient understanding of their own interests and how these could benefit from contributing to the interests of other actors. They therefore acted contrary to their own interests. His suggestion for remedying this was in part through legal and regulatory arrangements as well as education and moral enlightenment.

Mental capital as a basis for entrepreneurship and innovation

'Mental capital' was in List's opinion the core of the productive powers (List, 1827, pp. 57–67; 1841, pp. 49, 140, 159 plus ch. 12). He claimed that 'Mental work is in the social economy what the soul is to the body' (List, 1927–36, vol. 5, p. 42). By basing their method on the erroneous labour theory of value once established by Aristotle, both Smith and Marx focused on matter instead of mind. List mocked those who do not distinguish between the potential productivity of a Kepler and that of a donkey (List, 1841, pp. 142ff.; cf. p. 159). He maintained that Smith overdid his focus on exchange (monetary) value. List claimed that he ignored the intellectual, moral and religious activity behind the only apparent productive forces and thence behind exchange value:

> It follows that certain conditions must be fulfilled before men's productive powers, and their intellectual and physical labours, can be successfully applied to the production of material goods that have an exchange value. There must be good laws, effectively enforced. Persons and property must enjoy the maximum security. … high moral and religious standards … a good system of education. Science and the arts … adequate protection. … Moreover the labours of those who promote the expansion of productive powers are just as productive as those who actually make goods that have an exchange value. (List, 1837a, pp. 184–6)

List was deeply engaged in what is today normally termed 'liberal' or even 'neoliberal' causes such as freedom of the press and criticism of bureaucratic excess and arbitrariness. He did not see any contradiction between justice and economic efficiency and, quite the contrary, argued that only a free and just legal system could mobilize the mental powers of the individual citizen, in particular as entrepreneur, crucial to economic development. The most obvi-

ous example might be List's repeated attacks on the institution of slavery (List, 1827, Letter VI, pp. 86–7; 1837a, p. 184; 1841, pp. 200, 416). List's stress on the universality of law (trial by jury) and freedom of expression (for the press and so on) can be seen as an attempt to correct imperfections of the market for ideas and entrepreneurship, through vested interests and power structures. List intended to establish an efficient market for ideas, for innovation and for entrepreneurial activity, through implementing his liberal ideas. His work for security of property and for protection of investments can likewise be seen as intended to establish a market for innovation and for entrepreneurial activity, whatever the field, and all intended to secure an efficient working economy, to the benefit of general welfare.

> In order duly to estimate the influence which liberty of thought and conscience has on the productive forces of nations, we need only read the history of England and then that of Spain.
>
> The publicity of the administration of justice, trial by jury, parliamentary legislation, public control of State administration, self-administration of the commonalties and municipalities, liberty of the press, liberty of association for useful purposes, impart to the citizens of constitutional states, as also to their public functionaries, a degree of energy and power which can hardly be produced by other means. We can scarcely conceive of any law or any public legal decision which would not exercise a greater or smaller influence on the increase or decrease of the productive power of the nation. [Footnote on J.B. Say] If we consider merely bodily labour as the cause of wealth, how can we then explain why modern nations are incomparably richer, more populous, more powerful, and prosperous than the nations of ancient times? (List, 1841, p. 139)

List paid much attention to the role of incentives in economics and how these could be promoted by regulative and legal arrangements. He devoted a whole chapter to this in his *National System* (List, 1841, ch. 25, 'The manufacturing power and the incentives to production and consumption', pp. 303ff.). Patent laws were one legal measure for promoting the mental powers of production (ibid., pp. 56, 307). For other measures, see the section on tariffs, below. As noted above List regarded the mental powers (creativity and morality) as the bedrock of development, emphasizing the mental and therefore the entrepreneurial aspect to which the legal system is crucial. This makes the legal system even more crucial for development in his scheme than it could ever be in the so-called 'classical' 'orthodox' and 'neoclassical' economic tradition. (The same relation has been noted above for the emphasis and role of regulations and the public bureaucracy).

Law and economics deals specifically with the possibility of changing the incentives structure, but mainly within the standard economic image of Man as purely self-interested, that is, egotistic and in essence asocial. List's idealistic approach deals more than anything with the incentives structure, but

from a wider image or conception of Man, where Man is considered to be fundamentally but only potentially rational and moral, spiritual and social – in combination and in addition to being self-interested. This implies that standard economics has a narrower and more unrealistic conception of Man and therefore of reality, to which normal law and economics is also prone. List's conception is therefore more realistic but less easy to formalize.

Inventions, change and power: the obligation of law and economics

'Every new invention has some inconvenience for a number of individuals, and is nevertheless a public blessing' (List, 1827, Letter VI, pp. 86–7). As hinted above, List was a keen observer of power and of social structures, and therefore also of the implications of entrepreneurship for power. Inventions are a critical threat to many parts of an establishment, since power is ultimately based on control of some resource, whether material or mental/ideal or whether on a local as well as on a global scale. Technological and economic growth imply change, and therefore a restructuring of the power base, sometimes even on a vast and global scale.

Although possibly beneficial for the majority of a society, such change may be detrimental to some parts of the establishment which, in consequence, will try to use its dominant position to block change before it is too late, thereby cementing the structure of society and the economy, resulting in stagnation, eventual decline, sudden revulsion and upheavals. Hegel and Marx have described in a similar way the dynamics of social life with their theory of dialectics. It is thus of crucial importance for the survival of a community and eventually of a civilization that such static digressive behaviour is prevented and that a dynamic flow of change is permitted to take place. This line of reasoning applies on the local national level as well as to the international community. This problem, in particular, makes law and economics a crucial field of study and its students here have a crucial obligation.

Externalities like vested interests and power structures create transaction costs that render markets inefficient. This is a central theme of law and economics, which gives us yet another reason to claim that List should indeed be regarded as one of the important forerunners of law and economics. For List, power was at the core of economics and economic policy, both as a result and as a prerequisite. This was the source of one of his major criticisms of Adam Smith, who conveniently avoided this aspect in most of his writings, well aware as he was that the current power structure favoured Britain. Added to this point is the fact that Smith sometimes, contradicting other statements of his, was an ardent supporter of government intervention (Smith, 1759, pt. VI, ch. I). He was therefore also a supporter of military activity, being the latter's firm admirer: 'The art of war is certainly the noblest of all arts' (Smith, 1776, bk. V, ch. ii, pt. i). This insight was also

Smith's reason for applauding the very protectionist British Act of Navigation, 1651, since it injured commerce (in the short term) but strengthened the navy (Smith, 1776, bk. IV, ch. ii).

Smith's generalizations obscure the conflict between private versus community interest

A key to understanding List's theoretical criticism of Smith and his followers is to observe how he criticized the strategy of Smith as being too generalizing, disregarding the empirical and historical particulars of each practical phenomenon (List, 1841, pp. 171, 224ff., 316). For a further discussion of this, refer to the section in the introduction above. List often criticized the radical market school for their extreme and anti-social individualism, which he claimed would be destructive to communities – in opposition to his own socially oriented individualism (ibid., pp. 169–71). List's criticism of Smith's confusion of private versus national economic characteristics and interests is almost endless (for example, List, 1827, Letter 5, p. 75; 1841, p. 172). Hildegard Schwab-Felisch has edited a collection of List's writings devoted to this issue (see List, 1920).

This difference between private and public interests is the basis for differentiating between private and public goods. Public goods have in common the feature of concentrated costs and dispersed benefits; there tends to be structural underinvestment in these areas, since government regulation always is inadequate or inefficient, in lack of better remedies 'per definition'. This might not be a major problem if it had not been for the fact that these areas function as a carpet and a productivity-enhancing locomotive for all other economic activity in, practically speaking, any society. For List, these activities were related to innovation and communication, and to the former belonged education, science, arts and also the machine tool industry (List, 1841, p. 314). Public goods activities are therefore a prime target of governmental regulation and law making. List never used the phrase 'public goods' nor did he explain their basic characteristics – concentrated costs and dispersed benefits (as opposed to those of rent seeking: concentrated benefits and dispersed costs). Nevertheless, his criticism of Smith does in practice take as its crucial and given point of departure the difference between private and public interests, and thereby the difference between private and public goods, as a crucial and given point of departure (List, 1827a, Letter 5, p. 75; 1841, ch. 14, 'Private and national economy').

List claims that because of denying the sometimes-existant fact of a conflict of interest between private and community interests, Smith makes the logical mistake of playing down the necessity for organized action, through the institution of the nation (List, 1841, ch. 14, p. 163). He also claims that Smith's and Cooper's conscious confusion of private and public interests is a

reason for their playing down of the role of public regulation, lawmaking, and therefore the role of the nation (List, 1827, Letter VI, p. 87; cf. 1841, p. 166). List was no admirer of regulation for its own sake, but saw clear advantages of regulation for justice and prosperity in opposition to the principle of *laissez faire et laissez passer* (List, 1827, Letter VI, pp. 86–7).

Tariffs shall promote skill and liberty

List elaborated on the system of differentiating and temporary protective tariffs regarding different questions: incentives and security for the investor (List, 1837a, p. 89; 1841, pp. 167–8); need for uninterrupted production and stability (List, 1841, p. 298); protection and importance of the home market (ibid., pp. 24, 186–7, 191); trade war and dumping (ibid., pp. 95, 299); differentiated tariffs according to skill, experience, machinery and capital involved (List, 1837a, p. 145; 1841, pp. 178–9); according to necessity of life (List, 1841, p. 311); according to time, that is a bell shape of tariffs along the time axis (List, 1837a, p. 145; 1841, p. 314); special key branches such as the machine tool industry (List, 1841, p. 314); historical setting (List, 1837a, p. 145; 1841, pp. 115, 130, 314, 329); fiscal side secondary in importance (List, 1837a, p. 36); necessity of averting inefficient monopolies (ibid., p. 81; 1841, pp. 81, 169–71); necessity of state credit and interest-free loans as a kind of subsidy or negative tariff (List, 1841, pp. 296, 300, 315); state investments in infant industry, preferential interests rates to the investors; and temporary subsidies to promote infant industry (ibid., p. 315).

The German term for infant industry tariff was *Erziehungszoll* or education tariff, as opposed to *Schützzoll* – or protection tariff – of 'grandfather industry'. The fiscal side to his proposals were, however, totally secondary in importance (List, 1837a, p. 36). The economic activities to be protected more than any other were knowledge-intensive activities, since these activities had most to give at a later stage through raising the productive potential of the economy. Knowledge-intensive activities were also the most vulnerable since they were more difficult to foster and maintain. Accordingly, they had to be cultivated and fiercely protected and cared for with the utmost attention. List explains the similar British practice repeatedly (List, 1841, p. 39; cf. p. 111).

List criticized excessive protection that did not conform to the promoting principles of awaking and sharpening the productive powers of the nation but instead stupefied and blunted them (ibid., pp. 309–11).

The period in which List was active (1811–1846), had a poorly developed system of regulation in comparison to modern systems over 150 years later. The easiest or sometimes even the only way to regulate was mainly through the means of prohibition and tariffs. This is why the same strategy today would suit the developing countries best, with their, relatively speaking, poorly developed regulatory structure and infrastructure. More industrially

developed countries, on the other hand, potentially have many more financial regulatory instruments available. Many of these instruments are not being used today because of free-trade treaties and customs unions.

One instrument of very great importance in the industrialization process after List has been and still is public procurement (purchase) for various purposes related to a standard liberal conception of public duties, such as the upkeep of infrastructure in the widest sense: health, education, transport, law, security, defence etc. List seems to have missed this option, and the reason is possibly the small scale of public procurement of his time. His pupils, however, saw this clearly and extended his general principles into this field, utilizing public procurement to promote domestic know-how, production, jobs and a wider tax base.

Customs unions avert monopolies and inefficiency

List repeatedly claimed that protection might be damaging in a small nation, as it was likely to establish an inefficient monopoly. Small nations would have to cooperate through customs unions, arranged by means of international legal agreements. There would be no internal barriers to trade, so that trade was made efficient by internal competition – a classic mercantilist strategy. In order to facilitate efficiency, competition ought to be furthered through the following factors: education, innovation, division of labour, economics of scale, transfer of technology and other cultural achievements, technical standardization and the infrastructural grid, including of course the legal framework. In this regard he criticized Smith for his theory of (exchange) value as an expression of narrow-minded merchant interests (List, 1837a, p. 102). For List, protection should offer privileges to any individual who is willing to risk his capital for the public good, limited to the period during which it serves public interest. There are therefore both good and bad monopolies, and bad ones raise domestic prices permanently whereas good ones lower them in the longer term (ibid., ch. 15, p. 81).

Interrelationships of markets and protection for long-term efficiency

A variant of transaction costs is interrelationships between markets. This occurs when actors in one market are dependent upon other markets but have limited influence on them, or sufficient incentives to engage in them, even though this malaise might affect all the actors. In neoclassical terms this lack of joint action might cause market imperfections. List was a keen observer of the interrelationship between different markets, how the way markets functioned was crucially shaping, connected to and reliant upon how other markets functioned (List, 1841, pp. 39, 387). The character of some markets approaches the characteristics of public goods more than others (concentrated costs and dispersed benefits) and therefore this deserves special attention.

Concerning education, administration and communication, List knew the basic and crucial function these constituted as a carpet and locomotive for other activities. He was therefore on the alert to make these basic markets fulfil their functions by shaping the constitutional, legislative and regulative system with this in mind.

Furthermore, he was a keen observer of the fact that markets have different characteristics and came to the conclusion that therefore, in order to work properly, they need to be treated in different ways through public legislature, and that this in particular pertains to public goods markets. In a larger perspective, regulation and protection limited by time (and trade) may be seen as contributing more to competition than immediately introducing free trade. Protection may be seen as a remedy to correct market imperfections where some actors have the upper hand. Indeed, this was List's opinion concerning the strong position of English producers in his day, and this way of interpreting List is not new. As J.S. Nicholson points out, in his 'Introductory essay' to the 1904 reprint of List's *National System*, Henry Sidgwick argued, 'that ultimately the world at large might gain by the temporary protection of the constituent nations'.

Temporary protection will eventually lead to fiercer competition and a more efficient economy at a later stage. List answered his critics with precisely this argument and claimed that foreign domination of domestic markets consisted in a monopoly (List, 1841, pp. 169–71, 176–7). He saw no advantage in a foreign (British) monopoly over that of, not domestic monopoly, but internal competition (ibid., 1841, pp. 184, 189–93). He therefore set out to establish domestic production and competition. Establishing infrastructure was one important element of making this come about, as no competition can exist without communication.

Thus List's fight for the employment of legal arrangements for protection perfectly matches the agenda of law and economics, seen from a larger perspective. Indeed, List's concept of economics was in some profound ways wider reaching than today's law and economics.

Notes

1. These German authors all belong to the tradition of natural (rational) law, code of law as opposed to common law, but to the idealist section that adheres to Thomas Aquinas' verdict that the ultimate duty of Man is moral perfection. The opposed materialist section of this tradition of natural law, dominated by people such as Hugo Grotius, Thomas Hobbes and Samuel Pufendorf, saw biological survival as Man's ultimate duty. See, for instance, Christian Wolff (1749, ch. 1: 'Duties of nations to themselves and the rights arising therefrom'), for instance paragraph 35: 'Of a nation's duty to perfect itself and its form of government' and paragraph 51: 'How far this applies to the ruler of the state'.
2. The title first intended seems to have been *The American Economist*.
3. The subtitle of List's treatise was *Et la patrie, et l'humanité*. List's work was rediscovered in 1925 and published two years later in French and German. The question posed by the French Academy of Moral and Economic Sciences was: 'If a country proposes to introduce

free trade or to modify its tariff, what factors should be taken into account so as to reconcile in the fairest manner the interests of producers with those of consumers?'. No contester was awarded, but List's work was among the three mentioned as *ouvrages remarquables*.

4. The subtitle of this treatise was *On the Effects of Steam Power and the New Means of Transportation*. Erroneously this treatise was thought lost, for instance by Henderson (1983).
5. Since Britain was the most powerful and influential nation at that time. He mentions the Hanse, Venice and Holland as historical examples and the United States as a likely future example.
6. The British Act of Navigation 'accidentally' promoted British manufacture and power in the long run and led to the perhaps decisive Anglo-Dutch war. The English victory in 1651 cemented English world domination, later perpetuated by the United States.

References

Bücher, Karl (1893), *Die Entstehung die Volkswirtschaft*, 2 vols, Tübingen: Verlag der H.Laup'schen Buchhandlung.

Lasok, D. (1990), *The Customs Law of the European Economic Community*, 2nd edn, Deventer and Boston: Kluwer.

Laue, Theodore von (1963), *Sergei Witte and the Industrialisation of Russia*, reprinted New York: Columbia University Press (1969).

Lenz, Friedrich (1930), *Friedrich List, 'die Vulgarokonomie' und Karl Marx*, Jena: G. Fisher.

List, Friedrich (1827), *Outlines of American Political Economy*; newly edited by Michael Liebig and published with a parallel translation in German, Wiesbaden: Dr.Böttiger Verlag, 1997.

List, Friedrich (1837a), *The Natural System of Political Economy*; reprinted Totowa, NJ: Frank Cass, 1983, ed. W.H. Henderson.

List, Friedrich (1837b), *Die Welt bewegt sich. Uber die Auswirkungen der Dampkraft und der neuen Transportmittel*; reprinted Göttingen: Vandenhoek & Ruprecht, 1985, ed. Eugen Wendler.

List, Friedrich (1841), *Das Nationale System der politischen Oekonomie*, Vierte Auslage, Jena: Verlag von Gustav Fisher, 1922; translated into English as *The National System of Political Economy*, Fairfield, NJ: Kelley, 1991. (Reprint of the first English translation and edition; London, 1985, Longman's Green & Co., reprinted 1904.)

List, Friedrich (1920), *Staatsinteresse und Privatwirtschaft, eine Auswahl aus den Schriften von Friedrich List*, ed. Hildegard Schwab-Felisch, Berlin: Deutsche Bibliothek.

List, Friedrich (1927–36), *Werke. Schriften, Reden, Briefe*, 10 vols, Berlin: Reimar Robbing.

List, Friedrich (1962), *Grundriss des Römischen Rechts; nebst Übersicht über das Eherecht nach dem Codex juris canonici*, Baden-Baden: Verlag für Angewandte Wissenschaften.

Schumpeter, Joseph A. (1954), *History of Economic Analysis*, reprinted New York: Oxford University Press, 1976.

Simmel, Georg (1902), *Gemeinschaft und die Grossstädte*, reprinted Stuttgart: Brücke & Tür, 1957.

Smith, Adam (1759), *The Theory of Moral Sentiments*, Edinburgh: A. Kincaid & J. Bell; reprinted Indianapolis, IN: Liberty Fund, 1984.

Smith, Adam (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*; reprinted Indianapolis, IN: Liberty Fund, 1981, 2 vols (reprint of the Clarendon Press edition, Oxford 1976, with Edward Cannan's original index from 1922).

Sombart, Werner (1928), *Der Moderne Kapitalismus*, vol. 2, *Das Europäische Wirtschaftsleben im Zeitalter des Frühkapitalismus*, Munich and Leipzig: Duncker & Humblot.

Wolff, Christian (1749), *Jus Gentium Methodo Scientifica Pertractatum*, Halle; translated as vol. II, and reprinted as vol. I, Oxford: Clarendon Press, 1934.

47 Achille Loria (1857–1943)

Giampaolo Frezza and Francesco Parisi

Introduction

Achille Loria was born in Mantua (Italy) in 1857, and died on 6 November 1943, in Luserna, Turin. In a posthumous note by the Italian economist Augusto Graziani, the *Commemorazione del Socio Achille Loria* (Commemoration of the Fellow Achille Loria), published in 1949 by the National Academy of the Lincei, Loria is described as a 'Sovereign of Science', whose entire life was dedicated to the effective and tireless investigation of truth, a search which continued into old age. Every problem he encountered was thoroughly examined. Loria's wide knowledge of fact and theory is comparable to John Stuart Mill's, especially for the power, ingenuity, and receptive spirit and acceptance of new truths, including those from other disciplines, and from the confutation of mistake and renewed sophism in the social sciences.

Loria started producing his scientific works at the age of 22 with a monograph that captured the attention of Italian and foreign academics for its highly critical references, originality of thought and depth of examination.[1] His prolific scientific output is concentrated in the following research areas:

1. *Economic studies* Loria's economic studies are remarkable in the *Analisi della proprietà capitalista* (Analysis of capitalist property), Turin, 1899; *Le leggi di popolazione e il sistema sociale* (The laws of population and the social system), Siena, 1882; *Carlo Darwin e l'economia politica* (Charles Darwin and political economy), Milan, 1884, *La terra ed il sistema sociale* (The earth and the social system), Padua, 1892; *Nota sui debiti pubblici e le imposte* (Note on public debt and taxes), in Sienese studies, Verona Padua, 1897; *La Costituzione economica odierna* (The modern economic constitution), Turin, 1897; *Il capitalismo e la scienza* (Capitalism and science), Turin, 1901; *Sintesi economica. Studio sulle leggi del reddito* (Economic synthesis. Study of the laws of income), Turin, 1908; *Economia politica* (Political economy), Turin, 1934; *I fondamenti scientifici della riforma ecomonica* (The scientific foundation of economic reform), Turin, 1922; *Dinamica economica: studio sulle leggi della variazione* (Economic dynamics: study of the laws of variations), Turin: UTET, 1935.

2. *Studies of Marxism* These include a variety of studies, all integrally reproduced in the volume entitled: *Marx e la sua dottrina* (Marx and his doctrine), Milan, 1902.
3. *Studies of the value of money* These include preservations of his writings: *Le peripezie monetarie della guerra* (Dangerous monetary circumstances of war), Milan, 1919; *Le peripezie monetarie del dopo guerra* (Dangerous monetary circumstances after war), Milan, 1924; *Aspetti sociali ed economici del dopo guerra* (Social and economic aspects after war), Milan, 1921.
4. *Social and Sociological Studies* *Il movimento operaio* (The workers' movement), Padua, 1903; *Il salariato* (The salaried), Milan, 1915; *La socialogia, le sue scuole* (Sociology and its schools), Padua, 1900; *La morphologie sociale* (Social morphology), Paris, 1906; preface to the *Ricchezza della Nazioni* (Wealth of nations), by Adam Smith, Turin: UTET, 1927; *Verso la giustizia sociale* (Towards social justice), Milan, 1920.

Studies of the economic foundation of the law are relevant to the present discussion. They are forcefully argued in Loria's work, particularly in 'Le basi economiche del diritto' (The economic basis of the law), previously published in the *Giornale del economisti* (Economists' Journal), Second Series, 1893, 347, et seq., and subsequently published entirely in the second part of the monograph entitled *Le basi economiche della costituzione sociale* (The economic basis of the social constitution) (Milan, Turin, Rome: Fratelli Bocca Editori, 1913, 135, et seq.).[2]

In his work, Loria articulates an innovative 'economic theory of the law' through an argumentative investigation, developed from two perspectives: (i) the relation between the general theory of the law and economics; and more specifically, (ii) the confirmation of this theory in the most significant legal institutions, from their genesis to their progressive development, which from Loria's point of view are only a necessary reproduction of economic relations. From this point of view, he analyses in turn the following legal institutions:

1. family law;
2. property law;
3. succession law;
4. relations between proprietors;
5. relations between owners and non-owners;
6. criminal law; and
7. international law.

These institutions reflect the logical and dispositive order proposed in Loria's works.

The following discussion will explicitly examine Loria's economic theory of law. Subsequently, the general meaning of this theory in relation to the above legal institutions will be demonstrated.

Loria's 'economic theory of the law'

Loria believes that the economic theory of the law means the study, from one point of view, of the economic structure of legal sanctions, and from the other, of the economic nature of legal change. In this context, Loria devotes particular attention to the confutation of the opposing theories, from which the law, a product of the rational will of men, always precedes, and does not follow, the economic order.

The point of departure in his investigation is represented by the definition of the law as a legal sanction in accordance with the economic relationships of society. In this sense, Loria's theory appears to adopt a concept of sanction, similar to that of modern law and economics, whereby legal rules are understood as a system of imposed prices on the conduct of various subjects of the law. In particular,

> Whoever examines the law in the limited form of the economy, or in the association of free labour, would see that it would take into account a whole range of norms, ensuring to single producers the peaceful enjoyment of the results of their labour and of their accumulation of wealth, and that these prescriptions would be spontaneously respected by everyone else because they essentially conform to the real self-interests of each one.

Loria later contrasts the 'free association of labour' with the 'compulsive association of labour'. In the first case, in a society where producers, who have equal bargaining power, spontaneously associate, whoever dares to take opportunistic advantage of the other's right will provoke the other individual to break the association, declining at once the productivity and benefits of the association with the opportunist. Therefore, the antithesis of this self-interest (mutual benefit) is spontaneously abandoned. Under such conditions, the law has no other task than to identify the individual interests of each associate and to impose them jointly on all the associates. As such, the self-interest of each individual is spontaneously observed by everyone, and a legal sanction is not necessary.

This state of affairs is contrasted to the case of compulsive relationships. Here, there is an evident and necessary distinction between moral behaviour and the law. Ironically, in a command legal system detached from the economic order, the violation of the law may increase and improve the wealth and destiny of the subjects of the law. In the general case, however, the

offender of the law improves his/her well-being at the expense of another's: 'Opportunism pushes men to reciprocal breach, or antisocial actions'. Consequently, 'the law can no longer limit itself to a theoretical affirmation of individual economic interests, but must be armed with a severe sanction for opportunistic behaviour'. Considering the previously explained egoistical interests of men, it is evident that the difference between the law and moral behaviour is that the latter is unable to guarantee a social order of things; therefore, it is helped by the workings of the law. This touches the heart of the matter through the identification of an *effective sanction* against the egoistical behaviour. Selfish behaviour differs from moral behaviour, for the latter discredits the former, making its omission look like egotism. In summary, the right of coercive institutions is a necessary product of the conflicting interest of self-interested individuals. This is why people connected by their reciprocal self-interest will coordinate their actions in the pursuit of a common utility. Further, it is the completion and integration of the coactive moral behaviour where it is insufficiently manifested to its scope. This insufficient scope compared to the coactive moral behaviour represents a more complex and articulated organism.

It is often stated that the law represents a society-approved sanction against individual action. Differences in creeds and social beliefs, however, may ultimately lead to a coherent synthesis of collective wants. Consider the historical–legal comparison of the Germanic or Roman populations, which differed profoundly in terms of race, customs and climate. Despite these differences, however, the populations developed economic relationships. Loria confirms:

> [The] entire history of the law shows us how law in those societies is far from being the product of abstract reason, or the result of a national conscience, or the emanation of race. Economic relations are a constant, migrating from nation to nation, shifting from one century to another, from one phase of history to another. Yet if it is in doubt that the law transforms itself with the mutation of the economic relations, it is also true that this transformation does not always happen promptly.

At this point a painful contrast appears that is sometimes fixed between the new economic relationships and the antiquated legal institutions. If today we find the law crystallized or almost paralysed, this, in Loria's view, is because even the law is not attentive enough to the underlying economic reality; nor is it sufficiently tractible. In fact, while new conditions in rural industry require more elastic agricultural contracts and more benefits to the producer, the law is still shrouded by the dark funeral cape of Roman formalism.

According to Loria, further specification is implied by the relationship between economics and law, and the transformation of the latter to the former.

In particular, the derivation of the law from economic reality does not preempt the idea that once the economy is developed, it cannot proceed on its own. Thus, the law cannot be autonomous, but it becomes part of the political economy. It follows that during its own autonomous development the law becomes a powerful factor in the entire social constitution; thus, it is efficient, albeit a modification of the economic relationship. Therefore,

> [T]he economic theory of society never negates, nor does it soften, the incontestable and powerful reactions of the legal factor to the economic factor. In noting this important fact, economic theory intends that the legal factor should never forget that it is not spontaneous but at the same time it is the necessary product of prior economic factors.

Based on a similar consideration, Loria radically concludes that, 'each person examining social relationships will at some point convince himself of the influence of the legal fact that modifies the economic fact. It is not merely the last thing, but it is an influence of the economic fact that modifies itself'.

Thus, the 'solemn' affirmations of Rudolf Stammler's doctrine (contained in the work entitled *Wirtschaft und Recht*, Leipzig, 1896), seem, in the eyes of Loria, 'arbitrary and indeterminable' and, therefore, fundamentally refutable. Stammler, in particular, thought that the origins of economic relations arise from legal derivation, that they cannot be conceived except generally through a view of a specific legal environment under the direction of the law. A similar irrefutable proposition is invalid in its inverted meaning: according to Stammler, the economy cannot ever influence the law. Thus, the law is not analysed as the study of facts or concrete phenomena, but as the *theory* of the phenomena. Accordingly, it is an extended dogmatism by its own relative and positivistic dimension. The general theory of the law would be validated in every place and time independently of the economic order. Loria considers this prospect in relation to its abstract dogmatism, considering the study of the fact in its historical context, valid for every type of scientific approach. In this sense, the dispositions of private property, the protection of property, and the termination of the legal condition of an heir, have no reason to exist without reference to a specific case.

Stammler is careful even when he specifies that the relation between law and economics is not due to cause and effect but to form and substance. The law would be the form of the economy, where the economy represents the substance. This perspective in Loria's view contains a contradiction. Loria refutes the presumed derivation of the economy from the law, since 'if the law is the form and economy is the substance, how can Stammler pretend that the former precedes the latter? If between the form and the substance there is an anterior element from the other, the law can be nothing but the substance'. Then, the conclusion is inverted, 'if the form proceeds the substance, thus

creating the latter, it gives itself a form that is more convenient and beneficial'. All of this confirms that 'legal institutions are the necessary and natural products of economic relationships'.

What is not convincing about Stammler's work is highlighted in Loria's works by his consideration of the law attached to the will of human beings. After all, will is intended not as a purely arbitrary, nor as a subjective egoism, but rather as 'reasonable, rational, and partially circumscribed in the social environment. Individuals will point toward the high idealization of a community of free men who are freely conscious, and they tend to exist without ever completely fulfilling their higher form'. This expression represents a 'pure indeterminism' that reduces the indeterminate will of men to a supreme factor of the legal institutions: that is, the whole of social life. Loria found reassurance concerning the validity of the origin of the law from the economic system, in view of the lack of reliability of the opposing theories. This economic theory of the law had already been sensed by Plato (in the *Republic*, 1, 9) and by Cesare Beccaria in, *Dei delitti e delle pene* (On crimes and punishments) (1764), affirming that 'most of the laws are nothing but privileges, a tribute by all for the use of the few'.

The 'economic theory of the law' and the study of legal institutions

Proceeding from the history of the law to an analysis of individual institutions, Loria's study and theory seem right. The institution of the family 'is the result of the necessity of production'. The shift from the promiscuous state, 'vacant and wild', to a primitive family based upon a maternal figure, appears 'as necessary to make working associations and the coordination of producers more efficient, a system imposed to arrange individuals into family units'. That the husband occupies himself with war while the wife is concerned with family production justifies the organization of the family upon a maternal base. The so-called matriarchy represents 'the first method adopted to unify the work of many individuals in a defined territorial space; it is also an imposed limit on the wild dispersion of workers, the first attempt to coordinate the productive forces'. In other words, the family institution is created by economic necessity. Moreover, the shift from the matriarchy to the patriarchy is also justified by economic reasons. The man, being the stronger individual, assumes importance in the production process through purchasing. The man always determines the economic power, and as a consequence, the legal power. Related to this concept is the transformation of the property from the collective concept to the individual and private dimension, for this implies that the limits of relationships between relatives assume not only a philosophical meaning, but also an economic one. As such, this transformation creates hereditary succession. Economic factors determine the rise of this important legal institution.

According to Loria, economic factors are always at the core of the institution of the private property, which determines the rise of the institution of patriarchy. Consequently, 'patriarchy is an organizational arrangement for the natural production of private property that allows the man to join with the woman and together produce children. The domestic sovereignty is then subjected by other transformations, or mutations, along with the mutation of the structure of property'.

Along similar lines, Loria suggests that the institution of divorce is created for essentially economic reasons: 'In fact, the expansion of job opportunities for women creates the economic independence of the woman'. On the other hand, Loria suggests that the multiplication of the ownership of real estate always creates the predomination of the capitalist class. This is the class for whom the institution of divorce is profitable, a system ready to break the economy when it cannot freely change itself.

In sum, Loria views the influence of economic relationships on the legal disposition of property as self-evident. The essential economic foundation of property rules is already to be found in Roman law, which did not demand subjecting property to harsh restrictions as soon as these were required by the demands of production. At that time, legal servitude was instituted, 'inspired by the opportunity to promote the rural production; it also allowed the pursuit of private funds, because production does not hurt agriculture, still developing today; indeed, production helps agriculture with the killing of animals that are dangerous to it'.

Above all, this last statement is extremely interesting not only for its practical content (these days we cannot talk about extensive agriculture and hunting, which makes it even more effective, but, instead, it is the exact opposite problem), but also because it refers to the method utilized by Loria, who was committed to enunciating a methodology that is similar to the economic approach to law followed by modern academia.

Continuing the analysis of the economic foundation of property rules, Loria believes:

> [T]he disposition of the Roman law was based on the right to put a price on whatever was a potential object of ownership rather than compelling individuals to undertake natural barter. As noted by Jhering, this concept has a clear economic significance. To promote the cultivation of the land it is stated that a person who works on abandoned land becomes the proprietor after a period of two years. Economically, allowing the acquisition of the land (through possession extended over time) rewards the entrepreneurial spirit, and punishes the inertia of the landowners.

Loria continues his analysis of the origins of legal rules in the Roman tradition, observing:

> [I]n the primitive Roman law, the good faith possessor can be expelled by the proprietor and even deprived of the property and its fruits. This inflicts a strict sanction and obstacle on the progress of agriculture, and necessitates lenient legislation. It thus becomes possible to let the good faith possessor retain the fruits of his work.

Loria concludes that if the law is inspired solely by the interests of the property owner at any time, then it compromises other incentives for the efficient use of scarce resources. As an example, Loria cites France, where in the eighteenth century it was customary to entitle agrarian property; on the other hand, in the following century it was customary to privilege industrial property.

The evolution of hereditary law is explained ontologically by economics. Loria considers that the philosophers of the law, Gans and Lasalle, were not able to reach a satisfactory conclusion regarding the legal nature of hereditary institutions. That is, wills are far from being the product of 'man's freedom', as stated by Gans and Lasalle. This concept forms the basis of Loria's idea of the economic causation of legal institutions.

According to Loria, the rise 'of the economy based on slavery', was associated with the transformation of property from the collective to the individual, and the custom of making a will developed as a legal consequence. In certain circumstances,

> [There are] serious reasons [for] allowing the proprietor the right to utilize his belongings even in a different historical time. Slavery inflicts upon production and wealth accumulation harsh limits that can paralyse, or soften, mercy with equal power. Now this stimulates the making of wills which reaches its pinnacles, sharpening the desire to accumulate and transforming the accumulated into an idea that never dies.

This situation starts to weaken when, 'under the action of the imminent antagonism of slavery, the previous accumulation creates a universal impoverishment. Therefore, the ancient necessity to stimulate wealth accumulation creates the necessity to contain it'. Likewise,

> [T]he more intensive the production, the more necessary it becomes for a coordinated direction of the family firm that is absolutely incompatible with the collective intestate inheritance. Therefore, it is a difficult moment when the necessity to further perfect the institution of primogeniture is realized, understanding that only the first son can benefit from the inheritance rather than all of the children. From this we see the birth of the law of primogeniture.

Even the disappearance of the servant economy contributes to the abandonment of registered inheritance, and it imposes the necessity of instituting the freedom of testator. In an economy based on wage accumulation, this free-

dom is crucial to maintain the incentives of the worker in the latter years of his life: 'It is easy to understand why an economy based on salaries creates the right to make wills. It becomes the normal form of transmission of property. Thanks to its own force, the accumulation proceeds more rapidly, and the accumulation of the belongings is even more rapid'.

Even in an economy based on salaries, as compared with one based on slavery, accumulation eventually encounters severe limits. Beyond these limits, the accumulation can go no further without ruining the entire system: 'At this point, even the capitalist class's interest demands that capitalization and accumulation of wealth must be limited'. Therefore, hereditary rights show how economic rules govern, following the alternating and recurring stages of its own evolution.

Moving from the hereditary laws to the analysis of the law that regulates the relationship between proprietors, the logic of Loria's reasoning does not change, but rather is confirmed, for even this law represents the product of mutation of the economic order. For instance,

> [T]he rigid rules of ancient Roman law allow that bad faith in contracting parties is the basis for much of the reluctance of honest people to engage in contracting. Therefore, with the progression of society and the necessity of frequent economic relations, the standard of good faith must be imposed upon those contracting, and the rules of the primitive *strictum jus* must be abandoned.

In addition, lease contracts deserve particular attention. Even a superficial study of this type of contract shows:

> [There is a] substantial diversity between the conditions of the Roman tenant who has only *jus ad rem* (even if later extended through the judicial decisions) and the modern tenant who possesses *jus in re*, with the enjoyment of full real remedies. Now, the reason for this division is found in the different conditions of the economy, initially based on slavery, subsequently based on wage labour. In the economy based on slavery, the reduction of cultivation of most fertile lands almost completely excludes income from property. This made it impossible for progressive increments to urge the proprietor to expel a renter off the land in order to earn a higher income. Instead, in the economy based on salaries, the difference between cultivated lands creates income from property. The growth of property income induces the proprietor to break the contract with a renter as soon as the income is more than the designated contract. At this point, it becomes obvious that the renter's conditions, which were previously ignored, have a direct impact on the agricultural industry. It also creates the necessity to guarantee the renter a safer position, defended from the arbitrary behaviour of the proprietors.

The logic of Loria's argument closely resembles the modern law and economic explanation of contractual safeguards in the presence of incomplete contracts with asset-specific investments. In this way the legal property con-

tract is substantially mutated by the change in the underlying economic relationships: 'These relationships transform the incompatibility of rent due to the persistence, and the normality, of production. The position of the tenant who at first was exposed to the arbiter of the proprietor is now more and more invading and threatening to the law'.

In Loria's view, the area of the law deserving particular attention is that which disciplines the relationships between the owners and the non-owners: this area of the law reveals the dependency of legal choices on economic ones. Related to this issue is Loria's study on the employment relationship, which is an agreement between two parties, employer and employee, who generally have unequal bargaining power. When Loria was writing, legislators refrained from imposing strict restrictions on employment contracts, leaving most of the contractual terms to the freedom of the market. In this setting, Loria's assertion that the discipline of the civil code does not contemplate the problems embedded in wage-based employment relations and the consequent overbearing behaviour of the dominant class can be explained. Here, Loria is overcritical of the neutrality of the code with respect to distributional issues related to employment contracts and usury.

Loria is also critical of the greater vulnerability of the worker which 'allows the master to be the workers' judge, and to inflict upon the workers his own wicked desires without any sort of control'. He is wary of the idea that the market provides any substantive constraint on the employer's discretion, given the imperfect information which characterizes the labour market.

Loria then moves from the study of private law to the study of criminal law, and the interrelation between the legal and the economic relationships in the criminal arena. The economic nature of criminal sanctions appears clear in Loria's mind: 'Economic considerations powerfully impact crime rates, and the choice of the appropriate penalty'. Indeed, Loria thinks that crime can be considered as the product of two fundamental factors: a social factor essentially economic, and an individual factor directly anthropological and indirectly economic: 'crime is the byproduct of both factors'. Loria suggests, paradoxically, that crimes against property would not be a reason to exist in an environment characterized by perfect economic equality. The thief could not benefit from the additional share of wealth that he illegally appropriated. In the real world, perfect equality is not achievable, rendering Loria's observation empty of practical content.

Finally, Loria considers the social costs and benefits of criminal sanctions. He contemplates the evolution of penalties, observing that 'the penalty of mutilation was abrogated to avoid the burden upon society of the maintenance of an invalid'. Associated with this, it is necessary to consider Loria's belief that, 'the harshness of the penalties inflicted by the state for crime, is proportionate to the intensity of the associated compulsion of the criminal'.

The optimal social penalty is that which maximizes the deterrence associated with its imposition and minimizes the social cost of its infliction. Obviously, Loria believes that the criminal system is not immune from political bias. He considers that each state punishes more severely the crimes that damage the interest of the class for which political power is created.

Loria's excursus of the economic theory of law is certainly not limited to these areas of law. Of all areas of law considered by him, international law appears most directly influenced by economic relationships. Its genesis and development are indeed connected exclusively to economics. International law arises because of the need to facilitate commerce between people of different nations. International commerce greatly benefits the trading partners and wars may annihilate such exchange opportunities. The evolution of international law then raised commerce to a new and superior role, to the great advantage of people who live together and produce in a global economy.

Notes

1. *La Rendita fondiaria e la sua elisione naturale* (Income from property and its natural annulment), Milan: Hoelpi, 1880.
2. The following quotations are derived from this monograph.

48 Karl Marx (1818–83) and Friedrich Engels (1820–95)

Heath Pearson

Introduction

Karl Marx was born in Trier, a Rhenish town then belonging to Prussia. At the age of seventeen he entered the Faculty of Law at the University of Bonn; the year following, he transferred to the Law Faculty at the University of Berlin and, while he continued his training in jurisprudence there, he also pursued his interests in philosophy and history. In the end those collateral interests prevailed, and his doctoral degree was granted in philosophy. Thwarted in his career ambitions in the academy, the young Marx wrote for and edited several opposition newspapers. In 1842 he met Friedrich Engels, the scion of an industrial family, who was to be his lifelong collaborator and supporter. The two would spend most of the rest of their lives abroad, in England, France and Belgium, under conditions of explicit or self-imposed exile from their native Germany. Marx, always the towering figure of the pair (even after his death), read and wrote prodigiously, occasionally as a newspaper correspondent but more commonly as a freelance critic. In addition, he and Engels both tried their hand at political organization and communist agitation.

Marx and Engels were not the first scholars to attempt an economic analysis of law; nor, arguably, were they the best. Neither one held an advanced degree in economics or in law and, as self-avowed revolutionaries, they were studiously ignored by most professionals in both fields. But their exclusion from nearly all retrospective treatments of the law and economics movement is not therefore justifiable, or wise. Their writings crystallized in compelling prose the nineteenth century's regnant views of the nature of man and society. More importantly, their views on the economic foundations of law reached and informed a far wider audience that any commentator before or since can reasonably lay claim to. This entry will set the record straight by showing (i) how Marx and Engels came by an economic conception of law, (ii) how they applied the economic way of thinking to concrete problems in jurisprudence, and (iii) the systematic differences remaining between their approach and that of the modern mainstream law and economics movement.

The emergence of a materialist conception of law

The young Marx's intellectual development recapitulated in just a few years the slow evolution of Western jurisprudence over many centuries. Nineteenth-century jurisprudence stood out among the human sciences for its long and hearty resistance to the historicizing and sociological tendencies of Enlightenment thought. Accordingly, as a young law student in Bonn, Marx was at first swept away by the timeless categories of Roman public and private law. But as he explained in 1837, in a letter to his father, gradually he grew dissatisfied with the project of elaborating such a 'system of metaphysical principles'. Immersing himself in the study of history and modern languages, he suddenly concluded that jurisprudence as a self-sufficient discipline was fundamentally inadequate: 'A curtain had fallen, my holy of holies was rent asunder, and new gods had to be installed' (Marx, 1837, pp. 15–18).

In its stead Marx turned to G.W.F. Hegel, who had long insisted that law be viewed in a wider frame of reference. But Hegelian jurisprudence, while it did subject law to systematic scrutiny, did so from a wholly idealist perspective. By its lights, law had to be understood ultimately as emanating from the metaphysical idea of state. Even after Marx left Hegelianism and, following Ludwig Feuerbach's humanism, subjected Hegel to a 'transformational criticism' which made man, not incorporeal concepts, the protagonist of history, the result was still far from social science as we conceive it today. For the ultimate criterion in assessing law remained a normative one: whether and how it served the higher purpose of human self-realization. In effect, Marx and the others who had left Hegelianism (most famously P.-J. Proudhon, in his *What is Property?* 1840) continued to posit the ancient distinction between natural law, which deserved its sublime status as 'Law', and mere man's law, which did not. Thus he wrote in 1842 of Gustav Hugo, a member of the German Historical school of jurisprudence which was then experimenting with a sociological understanding of law, that he '*profanes* all that the just, moral, political man regards as holy' (Marx, 1842a, p. 204; see also Kelley, 1990, p. 257). True laws, he continued elsewhere, had rather to be understood as 'the positive, clear, universal norms in which freedom has acquired an impersonal, theoretical existence independent of the arbitrariness of the individual' (Marx, 1842b, p. 162). By contrast, private interest – the stuff of political economy and of civil society itself – was 'by its very nature blind, immoderate, one-sided; in short, it is lawless natural instinct, and can lawlessness lay down laws? Private interest is no more made capable of legislating by being installed on the throne of the legislator than a mute is made capable of speech by being given an enormously long speaking-trumpet' (Marx, 1842c, p. 261).

Gradually, Marx evolved towards a more materialist and determinist understanding of law, in which economic concepts held centre stage. A kernel

of materialism was present already in Feuerbach's attack on metaphysics; that kernel was nourished, and its husk of humanist essentialism corroded, by Marx's growing familiarity with the Scottish Enlightenment, with the French Saint-Simonians, and with the harder-headed Engels. Already in 1843, Marx was referring disparagingly to 'the speculative philosophy of law, that abstract extravagant *thinking* on the modern state' (Marx, 1843–44, p. 181). But it is with *The German Ideology* (1845–46), which for many commentators marks the watershed in the Marx–Engels work between youthful humanism and mature scientism, that the materialist conception of law achieved primacy. There they asserted, contra the insular pretence of jurists, that 'law has just as little an independent history as religion'; it needed rather to be rooted in mankind's secular evolution through the stages of material culture (ibid., pp. 32–4, 90–92, 329, 365). 'In each historical epoch,' Marx stipulated in *The Poverty of Philosophy* (1847), 'property has developed differently and under a set of entirely different social relations. Thus to define bourgeois property is nothing else than to give an exposition of all the social relations of bourgeois production. To try to give a definition of property as of an independent relation, a category apart, an abstract and eternal idea, can be nothing but an illusion of metaphysics or jurisprudence' (ibid., p. 197). In sum, 'legislation, whether political or civil, never does more than proclaim, express in words, the will of economic relations' (ibid., p. 147).

In the ensuing decades, Marx immersed himself in the study of economics, which he credited with laying bare 'the anatomy of civil society'. Law assumed a pivotal role in the so-called 'base–superstructure' model of social evolution which emerged from his titanic efforts. The classic expression of this model is to be found in *A Contribution to the Critique of Political Economy* (1859), and is worth quoting at length:

> In the social production of their life, men enter into definite relations that are indispensable and independent of their will, relations of production which correspond to a definite stage of development of their material productive forces. The sum total of these relations of production constitutes the economic structure of society, the real foundation, on which rises a legal and political superstructure and to which correspond definite forms of social consciousness. … At a certain stage of their development, the material productive forces of society come in conflict with the existing relations of production, or – what is but a legal expression for the same thing – with the property relations within which they have been at work hitherto. From forms of development of the productive forces these relations turn into their fetters. Then begins an epoch of social revolution. With the change of the economic foundation the entire immense superstructure is more or less rapidly transformed. (p. 263)

In other words, in the long run law is determined by resource endowments and the stock of technical knowledge. In the short run, however, law evinces

some inertial properties and thus can come into tension with the slowly evolving logic of production. It is this tension that gives rise to revolutionary upheaval. Engels took up this leitmotif in his own works, indeed stressing more baldly than Marx the social determinacy of law, which he termed nothing but 'the ideologised, glorified expression of the existing economic relations' (Engels, 1872–73, p. 381).

Concrete examples of the economic analysis of law

In their historical narratives and their commentaries on current events, Marx and Engels occasionally drew attention to these general considerations, in a way that bears more than a passing resemblance to present-day law and economics. Early social institutions were of special interest for Engels. 'At a certain, very primitive stage of the development of society', he wrote in 1873, 'the need arises to bring under a common rule the daily recurring acts of production, distribution and exchange of products, to see to it that the individual subordinates himself to the common conditions of production and exchange. This rule, which at first is custom, soon becomes *law*' (Engels, 1872–73, p. 380). As he later detailed in *The Origin of the Family, Private Property and the State* (1884), communal property right was one such adaptation to relative scarcity, as was the practice of 'group marriage', whereby each member of a community was guaranteed conjugal rights to every other member of the opposite sex. Engels saw this mating rule as an innovation unique to man among the higher mammals, which typically practised monogamy or polygamy. It stemmed, he reasoned, from the human animal's relative defencelessness in the isolated state, and the consequent incentive to exploit economies of scale in organized militancy. But this latter could only be sustained if male jealousy were held firmly in check (as it conspicuously was not in animal hordes during the mating season); by formally countermanding the male's instinctive urge for exclusive sexual possession, group marriage served that goal admirably (ibid., p. 145).

As early man evolved out of the hunter–gatherer economy with the domestication of animals and plants, new legal practices were drawn in train. Private property began to displace common ownership, as resources grew scarcer and were invested with more human labour. Private property in human beings – slavery – emerged at this juncture as well. In the hunter–gatherer economy, human labour had yielded no surplus over the costs of its own maintenance: thus 'outsiders' were either assimilated as colleagues or else simply killed. But 'with the introduction of cattle breeding, of metalworking, of weaving and, finally, of field cultivation, this changed. … The family did not multiply as rapidly as the cattle. More people were required to mind them; the captives taken in war were useful for just this purpose, and, furthermore, they could be bred like cattle themselves'. Still another innovation was

patriarchy. Since men had traditionally been the family hunters, domestic livestock naturally fell to their purview; moreover, as their musculature lent itself relatively well to the heavier agricultural tasks, they began to displace women in the production of vegetable foodstuffs as well. The upshot was that most of the new sources of wealth accrued to men, and their new-found economic leverage was readily converted into legal authority over womenfolk. The law of monogamy, for example, appeared at this juncture, primarily as a means to compel female fidelity and thus facilitate inheritance of property through the male line. Monogamy was 'based on the supremacy of the man; its express aim is the procreation of children of undisputed paternity, this paternity being required in order that these children may in due time inherit their father's wealth as his natural heirs. … It was the first form of the family based not on natural but on economic conditions, namely, on the victory of private property over original, naturally developed, common ownership'. Last but not least, these economic transformations issued also in the first appearance of the matrix of rules known as the state. On the one hand, the increasing division of labour gave rise to an incipient structure of economic classes, which undermined traditional government by consensus. On the other hand, the growing stock of wealth posed an incentive for intercommunal predation. Thus emerged the state, as an organ of internal administration, of collective defence and of organized aggression (ibid., pp. 162–73, 263–72; see also Marx, 1972). Henceforth the logic of public choice – where economic interests vie in the political arena for favourable legislation – would structure law's historical development.

The long transition to capitalism threw up further evidence of the economic logic of law. The revival and elaboration of civil law in the high Middle Ages stemmed, like its Roman exemplar, out of just such a process as we have been describing, developing 'simultaneously with private property out of the disintegration of the natural community' (Marx and Engels, 1845–46, pp. 90–92, 343). The revival of intra-European trade and the expansion of contact with the wider world impelled the European elite to accumulate lucre rather than social status and court retainers, thus producing a new economic class, the bourgeoisie, which pressed for a distinctive legislative agenda. The English Statute of Labourers of 1349, which imposed ceilings on wage rates, Marx linked to the Black Death which, by curtailing the supply of labour, threatened the dominant classes with a redistribution of income towards the nascent proletariat (Marx, 1867, ch. 10, sec. 5). Subsequent laws prohibiting vagrancy were born of the same redistributive intent, as were all parliamentary acts promoting the enclosure and clearance of common land. Royal legislative prerogatives posed some inertial resistance to capitalist enclosure, but to little effect in the long run: ultimately, 'the law itself becomes now the instrument of the theft of the people's land' (ibid., ch. 26).

Contemporary legislative debates too were interpreted in the light of economic interest and public choice. The passage of the 1815 Corn Laws demonstrated, according to Marx, how 'landlords everywhere exert considerable, and in England even overwhelming, influence on legislation, [and] are able to exploit this situation for the purpose of victimising the entire class of tenants' (Marx, 1894, ch. 37). But the victory of a declining class could only be pyrrhic: the industrial and commercial middle class of Britain understood that the Corn Laws raised their wage bill and made British exports less competitive, and so made its repeal a central legislative goal of the 1840s. Repeal was achieved precociously via an alliance with the working class, the bourgeoisie offering provisional support for the Factory Acts in exchange for proletarian support (Engels, 1845c, pp. 657–61). Those Factory Acts were themselves 'not at all the products of Parliamentary fancy. They developed gradually out of circumstances as natural laws of the modern mode of production. Their formulation, official recognition, and proclamation by the State, were the result of a long struggle of classes' (Marx, 1867, ch. 10, sec. 6). The economic logic of factory legislation was twofold. On the one hand, by hampering the sweated-labour practices of traditional workshops, they served the interest of the rising industrial bourgeoisie by speeding the consolidation of production in modern, capital-intensive factories (ibid., ch. 15). On the other hand, some regulations were pushed through parliament by the factory system's enemies. The Ten Hours' Bill of 1847, for example, was championed by a coalition including not only workers, but also aristocrats and those sectors of the bourgeoisie hostile to the factory owners. Even so, the Ten Hours' Bill could become law only with the acquiescence of the dominant social element, at a peculiar economic conjuncture: it was passed

> at a moment of neither prosperity nor crisis, in one of those in-between periods in which industry is still labouring sufficiently under the consequences of over-production to be able to set only a part of its resources in motion, in which the manufacturers themselves therefore do not allow full-time working. At such a juncture, when the Ten Hours' Bill limited competition among the manufacturers themselves, and only at such a juncture, could it be tolerated. (Engels, 1850, p. 296)

In a different vein, Engels used the principle of rational self-interest to argue that proposed regulations of interest rates would be largely self-defeating: faced with legally capped interest rates, financiers would (i) continue to lend at usurious rates, but only to persons too downtrodden to bring them to book for it, and/or (ii) divert a large part of their capital to equity investments. Either way, the benefits of regulation would not accrue to the parties intended (Engels, 1872–73, pp. 317–37).

One last point of affinity between Marx and Engels and contemporary law and economics is in the treatment of criminal behaviour. 'Like law,' they

wrote, 'so crime, i.e. the struggle of the isolated individual against the predominant relations, is not the result of pure arbitrariness. On the contrary, it depends on the same conditions as that domination' (1845–46, p. 330). Specifically, Engels argued that the calculus of crime was sensitive especially to the life-chances attendant upon a life of obedience. 'If the influences demoralising to the working-man act more powerfully, more concentratedly than usual, he becomes an offender as certainly as water abandons the fluid for the vaporous state at 80 degrees' (Engels, 1845a, p. 425). As a result, 'crimes of passion are becoming fewer and fewer in comparison with calculated crimes, crimes of interest' (Engels, 1845b, pp. 248–9). The statistical regularity of crime

> proves that crime, too, is governed by competition; that society creates a *demand* for crime which is met by a corresponding *supply*; that the gap created by the arrest, transportation or execution of a certain number is at once filled by others, just as every gap in population is at once filled by new arrivals; in other words, that crime presses on the means of punishment just as the people press on the means of employment. (Engels, 1843, p. 442)

Marx and Engels versus modern law and economics

So far, so closely parallel to the economic analysis of law as we know it today. Why, then, have Marx and Engels not been better recognized as progenitors? In part this may be due to their radical political affiliation, which agrees poorly with the 'Chicago' orientation of most present-day practitioners; in part, too, it must owe something to their lack of formal credentials in law or economics, and especially to their dismissive attitude towards jurists (whom they regularly termed 'ideologues', purveyors of 'mystification' and 'false consciousness'), who today bulk so large in the field. But there are more intellectually creditable reasons as well.

First, their considerations of law, substantial though they may appear in the narrow compass of this entry, were not in fact of paramount importance to them: considered in the context of the many thousands of pages of their collected works, their comments were strewn thinly indeed on the ground. Second, in explaining law, Marx and Engels privileged the logic of conflict over that of cooperation, of distributional struggle over allocative efficiency; this much is clear from their discussion of the enclosure of common lands, and it permeated their approach at large. Third, they occasionally softened their position of strict economic determinism, accepting that law could be shaped also by forces that had nothing to do with the forces of production, including such plainly 'superstructural' factors as the legalistic mind-set itself:

> As soon as the new division of labour which creates professional lawyers becomes necessary, another new and independent sphere is opened up which, for all its

> general dependence on production and trade, has also a specific capacity for reacting upon these spheres. In a modern state, law must not only correspond to the general economic condition and be its expression, but must also be an *internally coherent* expression which does not, owing to internal conflicts, contradict itself. And in order to achieve this, the faithful reflection of economic conditions suffers increasingly. (Engels, 1890, p. 481)

Jurisprudence which denies the economic basis of law 'forms what we call *ideological outlook*, influences in its turn the economic basis and may, within certain limits, modify it'. This rings true, but it has not had much of a hearing in the recent literature.

Fourth, we find reflected in the Marx and Engels approach to law the millenarian impulse that suffused their view of social evolution as a whole. After the victory of the proletariat, crime would be all but eliminated and criminal law correspondingly obviated; the coercive monogamy of male domination would be replaced by the consensual monogamy of love; in sum, with the end of human self-alienation, law would lose its *raison d'être*: 'state interference in social relations becomes, in one domain after another, superfluous, and then dies out of itself; the government of persons is replaced by the administration of things, and by the conduct of processes of production' (Engels, 1885, p. 268). Mankind, they believed, was not conclusively fallen.

Finally, underpinning Marxist chiliasm was a more pervasive contrast with today's law and economics, namely that Marx and Engels tended to analyse legal phenomena by reference to the will or needs of ambiguous entities like 'class', 'state', 'society', the 'mode of production' or even 'history' itself. This approach is excluded by the methodological individualism prevalent in economic analysis today; and rightly so, in so far as it prevents us from appreciating the fact of intra-class conflict, the fact that the state might not act as an 'executive committee' for its own interests, let alone anyone else's, and the fact that history is not going anywhere in particular.

Bibliography

Works by Marx and Engels

Engels, Friedrich (1843), *Outlines of a Critique of Political Economy*; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 3, pp. 418–43, London: Lawrence & Wishart, 1975.

Engels, Friedrich (1845a), 'The Condition of the Working Class in England'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 4, pp. 295–596, London: Lawrence & Wishart, 1975.

Engels, Friedrich (1845b), 'Speech in Elberfeld'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 4, pp. 243–55, London: Lawrence & Wishart, 1975.

Engels, Friedrich (1845c), 'History of the English Corn Laws'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 4, pp. 656–61, London: Lawrence & Wishart, 1975.

Engels, Friedrich (1850), 'The English Ten Hours' Bill'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 10, pp. 288–300, London: Lawrence & Wishart, 1978.

Engels, Friedrich (1872–73), *The Housing Question*; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 23, pp. 317–91, London: Lawrence & Wishart, 1978.
Engels, Friedrich (1884), *The Origin of the Family, Private Property and the State*; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 26, pp. 129–276, London: Lawrence & Wishart, 1990.
Engels, Friedrich (1885), *Anti-Dühring*; reprinted in: *Karl Marx, Frederick Engels: Collected Works*, Vol. 25, pp. 5–309, London: Lawrence & Wishart, 1987.
Engels, Friedrich (1890), 'Letter to Conrad Schmidt'; reprinted in *Karl Marx, Frederick Engels: Correspondence, 1846–1895*, pp. 477–84, New York: International Publishers, 1934.
Marx, Karl (1837), 'Letter to his Father in Trier'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 1, pp. 10–21, London: Lawrence & Wishart, 1975.
Marx, Karl (1842a), 'Philosophical Manifesto of the Historical School of Law'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 1, pp. 203–10, London: Lawrence & Wishart, 1975.
Marx, Karl (1842b), 'Debates on Freedom of the Press and Publication of the Proceedings of the Assembly of the Estates'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 1, pp. 132–81, London: Lawrence & Wishart, 1975.
Marx, Karl (1842c), 'Debates on the Law on Thefts of Wood'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 1, pp. 224–63, London: Lawrence & Wishart, 1975.
Marx, Karl (1843–44), 'Contribution to a Critique of Hegel's Philosophy of Law. Introduction'; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 3, pp. 175–87, London: Lawrence & Wishart, 1975.
Marx, Karl (1847), *The Poverty of Philosophy*; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 6, pp. 105–212, London: Lawrence & Wishart, 1976.
Marx, Karl (1859), *A Contribution to the Critique of Political Economy. Part One*; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 29, pp. 257–417, London: Lawrence & Wishart, 1987.
Marx, Karl (1867), *Capital*, Vol. 1; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 35, London: Lawrence & Wishart, 1996.
Marx, Karl (1894), *Capital*, Vol. 3; reprinted New York: International Publishers, 1967.
Marx, Karl (1972), *The Ethnological Notebooks of Karl Marx (Studies of Morgan, Phear, Maine, Lubbock)*, Assen: Van Gorcum.
Marx, Karl and Friedrich Engels (1845–46), *The German Ideology*; reprinted in *Karl Marx, Frederick Engels: Collected Works*, Vol. 5, London: Lawrence & Wishart, 1975.

Other works

Cain, Maureen and Alan Hunt (1979), *Marx and Engels on Law*, London: Academic Press.
Kelley, Donald R. (1990), *The Human Measure: Social Thought in the Western Legal Tradition*, Cambridge, MA: Harvard University Press.
Landau, Peter (1973), 'Marx und die Rechtsgeschichte' ['Marx and legal history'], *Tijdschrift voor Geschiedenis*, **41**, 361–71.
Pejovich, Steve (1982), 'Karl Marx, property rights school and the process of social change', *Kyklos*, **35**, 383–97.

49 Carl Menger (1840–1921)

Richard E. Wagner

Carl Menger would surely occupy a secure place in anyone's pantheon of influential economists. Along with William Stanley Jevons and Léon Walras, his work in the early 1870s initiated the neoclassical or marginal utility approach to economics. This approach continues to dominate economics even today. Menger was also one of the three founders of what came to be known as the Austrian school of economics, the other two being Eugen von Böhm-Bawerk and Friedrich Freiherr von Wieser. Recognition of a distinctive Austrian school emerged in the 1880s, flourished until the 1930s, then entered a period of quiescence, though it has shown some signs of revival, primarily in the United States, over the past two or three decades.

Life and work

The basic facts of Menger's life are simple and few. He was born in Galicia, which now lies mainly across Poland and the Ukraine, on what normally would be the last day of February, save that his birth year, 1840, was a leap year. His life as a university student commenced in 1859 with a year in Vienna, followed by three years in Prague, and then Krakow where he wrote his dissertation. In 1873, he was appointed to the University of Vienna where he remained until his retirement in 1903, save for the 1876–78 period which he spent tutoring and travelling with Crown Prince Rudolph. Menger retired from his teaching duties at the early age of 63 so that he could devote more time to his scholarship. While his post-retirement life lasted 18 years, and while from all reports he kept busily engaged with reading and note-taking, he wrote little after his retirement. Menger's reputation is almost wholly based on work he published between 1871 and 1892. Had Menger published nothing after 1892, his intellectual standing would have been little affected, if, indeed, it would have been affected at all.

The vast preponderance of Menger's works were collected in four volumes by F.A. Hayek and published by the London School of Economics between 1934 and 1936, as part of a 'Series of Reprints of Scarce Tracts in Economics and Political Science'. Most of the excluded writings can be ascertained by consulting the bibliography appended to Volume IV, though there are some writings, particularly anonymous journalistic offerings, that have not been included. A second edition of this collection was issued by J.C.B. Mohr between 1968 and 1970, unchanged from the first edition save that Hayek's

original English introduction was translated into German. (There was also a brief, amusing note in which Hayek retracts the statement in his original introduction that Menger was a tall figure, and notes instead that Menger was barely of average size.)

Volume I contains the first of Menger's two books (1871), which appeared in English as *Principles of Economics* (1950). Among other things, this book shifts the explanation of prices from cost of production to consumer marginal valuation, a shift that is generally regarded as the key distinction between classical and neoclassical economics. Volume II contains the second of Menger's books (1883). This was published in English as *Problems of Economics and Sociology* (1963). This book is best known for its initiation of what came to be called the '*methodenstreit*' between Austrian and German economists, a clash that in turn helped to solidify the impression of a distinctive Austrian school of economics. Beyond this, an important accomplishment of *Problems* is its elaboration of the theme that there are many beneficial social institutions that arise not through plan or intention but as unintentional byproducts of people pursuing their individual interests and plans. While Menger mentioned a number of examples briefly, most of his effort in this regard focused on money. In this respect, Volume IV of his collected works contains Menger's monetary writings, and opens with the third edition of his famous (1892) monograph 'Geld' which appeared in the *Handwörterbuch der Staatswissenschaften*, a few glimpses of which were also published in English as 'On the origins of money' (1892). Volume III of Menger's collected works mostly contains further writings of methodological disputation, comprising his contribution to the *methodenstreit*. It also contains five biographical essays, two of them in commemoration of hundredth anniversaries of births (Friedrich List and John Stuart Mill), and three of them obituaries (Lorenz von Stein, Wilhelm Roscher and Eugen von Böhm-Bawerk).

Marginal utility, economic order and economic process

Prior to the 1870s, economists typically attributed differences in product prices to differences in the costs of producing those products. Products command higher prices because they are more costly to produce, according to this line of explanation. While each of the neoclassical triumvirate had his own mode of presentation and expression, they were united in replacing this cost-of-production theory of pricing and resource allocation with a marginal utility theory. Each member of the neoclassical triumvirate explained prices in terms of the utility that consumers derived from a marginal addition to their stock of an item. It was not a high cost of production that elicited high prices from consumers, rather it was the high consumer valuation of additional output that justified large expenditures by producers to bring such products to market.

John Stuart Mill's *Principles of Political Economy*, which was originally published in 1848, was probably still the most widely referenced statement of the economic principles of pricing and resource allocation when it appeared in a revised edition in 1871. Shortly after he remarked that 'there is nothing in the laws of value which remain for the present or any future writer to clear up; the theory of the subject is complete', Mill waded into a presentation of the diamond–water paradox, and in much the same fashion that Adam Smith had done nearly a century before. Diamonds are expensive, water is not. Yet people can live without diamonds, but not without water. This formulation appeared to present a paradox, which was resolved by invoking a distinction between use value and exchange value and by limiting the economic explanation to exchange value. Yet the paradox dissolves instantly under the neoclassical formulation. Market prices depend on the values that consumers place on one more diamond or one more bottle of water. Until diamonds begin to sprout like dandelions in the spring or until people come to think that beautiful is ugly, diamonds will fetch much higher prices than water. Yet if one monster monopolist could control the entire stock of water and another the entire stock of diamonds, the water monopoly would be far more valuable. Relative market prices depend on relative marginal utilities, but this says nothing about relative total utilities.

The marginal utility apparatus represents a clear advance over the cost of production approach for explaining prices and resource allocation. It is clear that Menger shared with his neoclassical brethren this focus on explaining variations in prices by variations in consumer valuation of the marginal unit. Yet it is also clear that he used his analytical apparatus differently than did the other neoclassicals. For those economists, economics was centrally concerned with the explanation of prices and resource allocation. To this day, this centrality is illustrated by a vocabulary that makes repeated references to economic theory as being price theory or price and allocation theory. For Menger, however, pricing and allocation were of only secondary interest, and for this reason Menger stands apart from his neoclassical brethren. For Menger, price and allocation theory was only a small part of economic theory.

In the Walrasian orientation that came to dominate neoclassicism, the central analytical problem is to explain the structure of relative prices and the allocation of resources among competing uses in a setting where knowledge is frozen and time is suspended. In this analytical schema, changes in knowledge are sources of exogenous shocks when time is allowed to elapse. In sharp contrast, Menger sought to develop an analytical framework where the passing of time and the development of knowledge, institutions and organizations that accompanied the passing of time occupied the analytical foreground, and were not injected as exogenous shocks. Consumer valuations and prices were always present, to be sure, but so was the acquisition of knowledge, the

development of new products, and the emergence and creation of new institutions and organizations.

This point is illustrated nicely by Menger's treatment of pricing, particularly his emphasis on price formation as opposed to price determination. Where Walras sought to explain some equilibrium pattern of prices in terms of a logic of marginal utility, Menger refused even to impose the requirement that the same product sell everywhere for the same price. Such price uniformity might arise in some historical settings, Menger would surely acknowledge, but the central analytical task would be to explain the formation of such uniformity, and not to impose it by assumption. Menger's central concern lay with explaining the formation of actual prices, and not with expounding a logic of assumed equilibrium prices. This interest in a process of economic evolution through time led to a treatment where competition evolves from monopoly, and where one of the central features of the economic process is the generation of new knowledge and products, not as exogenous shocks to some logic of equilibrium, but as a natural feature of a scarcity-induced competition through time.

Spontaneous order, institutional evolution and legislation

An important part of Menger's view of social and economic processes as centrally involving movement through time, as against being denoted by a structure of relationships at some point outside of time, is his treatment of spontaneous order and the evolution of institutions. There is an important distinction between what are direct objects of individual choice and what are unintentional consequences of the interactions among such individuals and their choices. Patterns of economic activity are generally orderly and apprehensible, but that orderliness is no-one's intention, responsibility or creation; rather, it is simply a byproduct of social interaction.

A primary task of economics thus involves the explanation of the evolution and emergence of institutional arrangements that are socially beneficial and yet which result from no-one's plan or intention. Menger gave a number of examples, among them money, language and law. By far his greatest efforts were given to the explanation of the emergence of money, with other examples more mentioned or listed than developed or elaborated.

Menger explained the emergence of money out of barter as his paradigmatic example of institutional evolution. Money was not created through legislation, after which the economic process began. Rather, money emerged as an unintentional byproduct of people's efforts to promote their economic interests. Without money, opportunities for trade were confined by the double coincidence of wants that barter required. Gradually, people came to recognize that some objects were more readily saleable than others. People would accept these items in trade, rather than refuse to trade, even if they had no

immediate need for them, because those highly saleable items could be traded relatively easily at a later time for something that was desired. While history has seen many objects occupy such positions of easy saleability, such precious metals as gold and silver are the most prominent, along with certificates that represent title to such metals.

To attribute an invisible hand explanation to the emergence of institutions is not, however, to say that the results of evolution are incapable of improvement through legislation, Menger also argued. In particular, he argued in Appendix 8 of *Problems* that the spontaneous emergence of legal rules sometimes generated rules that were detrimental to the common welfare, and the correction of which was a task for legislation. He would not have argued for any efficiency of common law, to pick one instance where a large literature now exists. He would have denied that people need consciously to adopt some legal framework before economic life can proceed, and would have claimed instead that such a framework will emerge spontaneously and conterminously with economic activity. At the same time, however, he would have held out the possibility that in some instances rules that emerged spontaneously might be susceptible to improvement through legislation.

As a conceptual matter this is certainly an entertainable proposition. Yet Menger did not elaborate with concrete illustrations nor did he develop any schemata that would make it possible to judge when legislation might truly be corrective and when it might accomplish something other than correction, say as illustrated by the literature on rent seeking that Gordon Tullock (1967) initiated. While Menger showed clearly how money emerges spontaneously, it is also the case that the free banking that accompanies the spontaneous or market generation of money has not been competitively robust against central banking created through legislation. Is the historical dominance of central over free banking an example of what Menger had in mind by the legislative correction of what he would have regarded as deformities in the outcomes of the evolutionary process? It would have been good to have seen Menger's mature thoughts on the relation between spontaneous evolution and deliberate legislation in a variety of concrete instances. Menger retired from the University of Vienna to devote his remaining years to reading and writing. The reports from his various unpublished notes and fragmentary manuscripts, however, suggest that no significant progress was made, though such efforts are now at the forefront of various research programmes in law and economics, new institutional economics, evolutionary economics and public choice.

The *methodenstreit* as destructive spontaneous evolution?

Perhaps the *methodenstreit* itself can serve as an illustration of the way social deformity can arise through spontaneous evolution. The course of the *methodenstreit* was clearly a product of spontaneous evolution, and was not

the result of some deliberate design. One significant consequence of the *methodenstreit* was the generation of the presumption that there is something that can meaningfully be called Austrian economics. Without the *methodenstreit*, there almost surely would never have arisen the sociological phenomenon of an Austrian school of economics, or of a German Historical school, for that matter – and particularly the notion that the one is the antithesis of the other.

The *methodenstreit*, set in motion by Menger's 1883 book, became a nasty and passionate battle that spanned the better part of two decades until its participants tired of the controversy. To be sure, there were significant differences between the thinkers and the approaches they represented, yet these were minor as compared with the differences between the commonalities shared by those participants and the growing neoclassical emphasis on economic theory as price and allocation theory outside of time. Carl Menger's orientation towards economics and social science similarly characterized the Germanic tradition of economic scholarship in the nineteenth century, whose principal exemplar was Gustav von Schmoller, who was also nearly Menger's exact contemporary. While Schmoller lacked Menger's theoretical acuity and decisiveness, it is clear that Schmoller's *Grundriss der allgemeinen Volkswirtschaftslehre* reflects an orientation similar to Menger's in its animating spirit and its central concern with the evolution of institutional arrangements through time.

The *methodenstreit* would have to be classified as a family feud, though it should also be noted that most murders are committed among family or friends. In this respect, Joseph Schumpeter (1954, p. 815) argues that the *methodenstreit* was primarily a clash of temperaments, and this surely has considerable plausibility. In terms of Isaiah Berlin's distinction between the hedgehog and the fox in his famous essay by that name on Tolstoy, Menger was more the hedgehog who was seeking one unifying principle, a universal solvent of the social world, so to speak. In contrast, Schmoller was more the fox, who emphasized variability and singularity, so much so that general organizing principles conducive to conceptual clarity tended to be slighted. Such differences in temperament may in turn have been magnified against the background of Prussian and Austrian hostility that included a short war in 1866, and which helped a passionate family controversy to turn bloody.

Family feuds can have enduring consequences, and the *methodenstreit* would seem to be no exception. While there ceased to be any identifiable Austrian-oriented approach to economics by the late-1930s, a resurgent interest in Austrian economics has emerged over the past two or three decades, particularly in the United States, an interest which has been chronicled crisply by Karen Vaughn (1994). While the main intellectual figures in this resurgence were Ludwig von Mises and Friedrich A. Hayek, Vaughn quite correctly

locates Carl Menger as the central inspirational source behind these two figures.

Also published in 1994 was Nicolai Foss's *The Austrian School and Modern Economics*. As with Vaughn, Mises and Hayek are the two central figures in Foss's narrative. Only Foss does not write of a resurgence of Austrian economics; rather, he writes of it as a degenerating research programme, primarily because its practitioners offer more by way of critical commentary than by constructive example. Foss is writing for people who are interested in such currently robust research programmes as law and economics and the new institutional economics. He argues that Mises and Hayek still have many valuable insights to offer such scholars, and he counsels such scholars not to let the Austrian identification that is commonly attached to Mises and Hayek interfere with a recognition of the value of their insights.

Foss and Vaughn share an admiration for Mises and Hayek, as well as for Menger, only they differ sharply in their projection of the future of Austrian economics. That such an issue would even be discussed today is a product of the *methodenstreit* and the social formations that emerged in its aftermath. The *methodenstreit* surely aided the Walrasian ascendency in economic theory, and has also generated an inadequate appreciation of the intellectual heritage of that which has been displaced by that ascendency. Had the family controversy between Menger and Schmoller not erupted in the manner it did, it is arguable that we would find ourselves speaking today neither of an Austrian nor of a German historical school of economics, yet at the same time we would find before us a robust alternative to the Walrasian type of neoclassicism. That alternative would have been centrally concerned with the development of organizations and institutions through time rather than with a logic of prices and allocations outside of time, and it would have incorporated the progressive components of what now are fragmented under such rubrics as the new institutional economics, evolutionary economics, law and economics and Austrian economics. But the *methodenstreit* did happen, and as a result economics took on a different history in the twentieth century than it might have taken had cooler heads prevailed in Vienna and Berlin in the closing decades of the nineteenth century – and, moreover, a history that has made more difficult the articulation and implementation of an alternative to Walrasian neoclassicism, even though that alternative was already in place prior to the *methodenstreit*.

References

Foss, Nicolai Juul (1994), *The Austrian School and Modern Economics*, Copenhagen: Munksgaard International Publishers.

Menger, Carl (1871), *Grunndsätze der Volkswirtschaftslehre*, Vol. 1 of *Gesammelte Werke*.

Menger, Carl (1883), *Untersuchungen über die Methode der Sozialwissenschaftern und der Politischen Ökonomie insbesondere*, Vol. 2 of *Gesammelte Werke*.

Menger, Carl (1950), *Principles of Economics*, Glencoe, IL: Free Press.
Menger, Carl (1963), *Problems of Economics and Sociology*, Urbana, IL: University of Illinois Press.
Menger, Carl (1968–70), *Gesammelte Werke*, 4 vols, 2nd edn, Tübingen: J.C.B. Mohr.
Schumpeter, Joseph A. (1954), *History of Economic Analysis*, Oxford: Oxford University Press.
Tullock, Gordon (1967), 'The welfare costs of tariffs, monopolies and theft', *Economic Inquiry*, **5**, 224–32.
Vaughn, Karen I. (1994), *Austrian Economics in America: The Migration of a Tradition*, Cambridge: Cambridge University Press.

50 Plato (c. 427–349 BC)

Wolfgang Drechsler

Introduction

To let law and economics begin, not in the late 1950s in Chicago, but some 2300 years earlier in Athens (Athens, Greece, that is), and with Plato in particular, runs the risk of attracting a certain amount of derision. First, it seems to cater to the cliché of letting everything begin with the ancient Greeks ('Already Aristotle has said …'), as an educated-bourgeois equivalent to the late Soviet praxis of having everything begin with Marx or Lenin. Second, Plato of all thinkers seems to be an odd choice as the founding father of the decidedly realist approach of law and economics, because it is his student Aristotle who is universally regarded as the first economist (after all, he first coined the term in his *Politika*) as well as the original realist. Plato, in contrast, has the image of an aristocratic abstract theorist, dwelling in a world of ideas.

However, the mental framework within which Western (and, to a certain extent, global) civilization in the late twentieth century operates is indeed based on and still substantially determined by the ancient Greek one. Plato's thought (as well as Platonism, although that is another matter) provides us with an original structure with which (rather than about which) we invariably still think. Second, the 'real world' as such, as well as its legal systems and its economies, are arguably actually Aristotelian, and there is good reason for recognizing in Plato the master of theory and ideas. Yet, Plato's focus on *praxis*, especially in matters concerning humans living together, is also obvious to those who take the time to study him with an open mind, especially in his genuine guidebook, the *Politikos* ('Statesman'), and in the *Nomoi* ('Laws').

Biography

Plato, the son of Ariston, was born around 428–7 BC in Athens and died there in 349–7 BC. From an aristocratic background, he had political ambitions in his youth, and probably started to exercise them as well. However, he lived in uncertain and violent times, when the *polis* of Athens was already in severe decline, and his shocked reaction to the forced death by poison of his teacher Socrates is frequently seen as the beginning of the antagonism between philosophy and politics (see Arendt, 1990). After a lengthy visit to Sicily, he founded his own philosophical school in Athens, the Academy, which survived for 900 years. Later in his life, Plato travelled twice more to Syracuse

to educate the young tyrant, Dionysios II; the extent to which his political writings were texts for or results of this, or neither, is heavily debated, as are most facets and data of his life. Plato has attracted violent attacks and criticism for more than two millennia, but very few people concerned with philosophy doubt that he was one of the greatest philosophers of all time, and many would say the greatest of all.

Law and economics: the *Nomoi*

The *Nomoi* (it is best to retain the Greek name because our meaning of 'Laws' is quite different) constitute Plato's last and longest work. In dialogical form, the founding of a new state or a more or less independent colony is discussed, and an ideal or model constitution and laws, including detailed codification-like catalogues, is outlined. The *Nomoi* are the logical place to look for law and economics principles, because they are immensely practical, as anyone who has ever been actively involved in law-making will observe. They offer motives and foundations as well as a systematical, finalized legal catalogue regulating phenomena that usually have to be regulated today as well, and in almost all important areas of modern legal order. They are also Plato's most realistic work – not less so because it is a utopian book, but more so, as a utopia with realistic goals that are only communicated in specific ways.

Nevertheless, according to Ryle (1967) no 'modern statesman or political theorist would suggest remodelling Canada, say, or Russia or Ghana according to the prescriptions of the Laws – or even retrospectively, Macedon or Rome'. After all, the *Nomoi* are a 'would-be practicable plan' (Ryle, 1967: 332). The *Nomoi*, too, are, as Hans-Georg Gadamer says, 'an immense utopia' and like the *Politeia*, 'an educational state' (Gadamer, 1991a: 288–9; see *passim*). The *Nomoi*, in short, are Plato's classical utopia, which has only been obscured by the long-lasting miscasting of the *Politeia* in that role. The key to understanding the *Politeia*, however, lies in Gadamer's interpretation of the dialogue as a heuristic utopia, that is, as a utopia in the sense of the presentation of a non-existent ideal state *which is never supposed to be realized*; indeed, which should not even be approximated (Gadamer, 1934, 1991a; see also Drechsler, 1998, 2002a).[1]

Theory

'This is what the law-maker must often ask himself: What is my purpose? Do I indeed achieve this or rather miss my goal?' (744a), says Plato in Book V, in the context of a section dealing with wealth and worldly goods as a final goal. The two questions might serve well as a motto for law and economics generally. One of the key features of law and economics, the concern with the effect as opposed to the purpose of a law, is thus indeed first formulated by

Plato in the *Nomoi*. This feature is found throughout the book as a principle on which the entire legal order of the new state, or much of it, seems to be built. It even includes procedure, which means that legal provisions depend on how justice will realistically be applied. This concern was to remain neglected for more than two millennia after Plato first expressed it here.

However, the economic system of the *Nomoi*, from today's perspective, seems grossly anti-business, anti-trade and anti-finance.[2] It is not for nothing that Pöhlmann speaks of its 'anti-capitalist trade and business policy' (1925: 189). But that is just the ideology. Economics are prominent in the *Nomoi* – even the provisions of penal law can be argued to have at least also an economic impetus, in the sense that they aim at maintaining the economic system of the *Nomoi* (Bisinger, 1925: 102–5). More importantly, regardless of the economic system of the *Nomoi*, which is indeed 'anti-capitalist' – the way Plato gets there, and ensures its maintenance, is very much based on law and economics principles, and that is what matters. After all, as George Stigler says in a key law and economics essay, efficiency, often mentioned in this context as the basis of the free market approach, 'is to be judged only with respect to the goals one seeks' (1992: 459; see also Coase, 1960: 44, last paragraph; Campbell, 1985: 196. In any case, economic efficiency can also be seen as a collectivist principle).

Praxis

In *Nomoi* 915d–e, for instance, it is decreed that all sales must take place in the official market, by the vendor in his allotted space. Sales occur under the special protection of the state, so that the buyer is protected, for example, from purchasing goods with hidden damage. What is important here is not the consumer protection aspect, but rather that sales outside of the market are by no means prohibited: if the buyer chooses to trust the vendor, he may buy wherever he wants; however, in that case there is no legal protection. In other words, the free market (in the original sense of the word) is not prohibited, but the kind of control which is necessary for protection, and which is probably desired by most buyers, is something the buyer may waive. Thus freedom of contract is guaranteed. However, the state is only responsible for quality control if the purchase is made in a specific place where control can be easily exercised. (But see also Pöhlmann, 1925: 190; compare Lacour-Gayet, 1945: 35.)

According to *Nomoi* 916a–d, if a sold slave turns out to be physically or mentally sick, the sale may be revoked – unless the buyer has a specialist knowledge of humans, such as a physician or an athletics coach. It is thus made clear that not all parties to a contract are the same – this is the interesting feature in our context, not the slave trade. The nineteenth-century insistence on the fictional equality of all parties to a contract is thus superseded by a

realistic differentiation based on professional expertise, and the responsibility is allotted to the party which most easily commands that expertise. In contemporary German private law (for example, §§ 459ff. BGB), no such differentiation is made. In that sense, the *Nomoi* are – from a law and economics perspective – more advanced than the 2300-year younger codification, since the *telos* of the provision, which ultimately is the smooth flow of commerce, is much better ensured by Plato.

Nomoi 937d–938c deals with litigious lawyers who try to persuade people to sue, even if there is no good case. The mildest punishment for such behaviour is to be forbidden to sue for a certain period. Lawyers who are citizens and who have been found by a court to have frivolously sued twice for sheer fun, or once out of greed, even risk capital punishment. This seems outrageously harsh today, although theoretically, that is, if it does not concern one's own lawyer, it might be possible to find many who would agree with this measure. But regarding the severity of the punishment, one should remember the utopian quality of the *Nomoi*. This provision is a necessary realist corrective because the *Nomoi* have a system of 'punitive damages': double, or, more rarely, triple the worth of items in question, payable to the victorious party (see, for example, *Nomoi* 916b–d, 921c, 928b) Thus, the incentive for frivolous suits is very high, and without the corrective measures advocated in this section, the system might not work very well. Noticeable here is Plato's interest in procedure and the influence that procedure has on substance; this is a key law and economics feature.

Nomoi 849e–850a, mostly known for advocating the prohibition of (domestic) trade gain, and often referred to as 'anti-capitalist', also contains what can be interpreted contrariwise, namely, as an almost Coasean concern with the efficient regulation of the market over government intervention, even if the goal is 'anti-capitalist'. Plato wants to exclude from the *polis* any form of credit sale – 'Nobody shall give to another anything in advance' (844e; on Plato and credit, see Klingenberg, 1982). However, similar to 915d–e (above), this is not achieved by state regulation or intervention, but by the opposite, that is, the exclusion of state intervention in the case of non-payment of the debt: 'Or, if he necessarily has the trust in the other, he has to be satisfied whether he receives that what is his or not, because he may not draw legal claims from such an action' (849e–850a). It remains, thus, the vendor's choice. So, in the end there is no intrusive state (Pöhlmann, 1925: 185) at all; on the contrary.[3]

In *Nomoi* 844a–d, the passage on water supply – probably based on historical examples, as it explicitly claims (844a; Haliste, 1950, 143–4) – it is decreed that, if someone with land elevated higher than that of his neighbour has too much water on his land because of rain, he must agree with the latter on how to let the water flow away so that it creates the least damage *for either*

party (844c; Haliste, 1950: 147–9). An official is asked to intervene only if an agreement cannot be reached, (844c–d) – and that is certainly not historical, but Platonic.

One of the central achievements of Coase's 'The problem of social cost', arguably the foundational work of modern law and economics, is to demonstrate that the Pigouvian tax, which internalizes externalities by, for example, charging a polluting industrial plant for the costs of pollution, is not necessarily economically efficient, because 'taxing the generator of the externality in a measure corresponding to the difference between the private cost and the social cost of his own activity [fails] to consider the effects of potential victims' behaviour' (Parisi, Chapter 1 in this volume; see Coase, 1960: 2). The motive of *Nomoi* 844c is thus clearly Coasean, that is, the owner of the less elevated land in Plato's case is not, or not only, a victim, because without him, there would be no problem at all. 'The problem is to avoid the more serious loss', as Coase says (1960, 2), and that must be Plato's motive as well.

Conclusion

The *Nomoi* are historically the primary document of the school of thought that law and economics is today – or, better, of one way of pursuing this kind of thought and 'method' – and also the first formulated and detailed legal system that generally, and not incidentally, operates on law and economics principles. This becomes particularly obvious once one is aware of the realistic character of the *Nomoi* as a classic utopia, with all its attendant features. Plato's image and even goal should prevent neither the law and economics scholar nor the philosopher from seeing that the state of the *Nomoi* has a realistic impetus, and that realism implies facing the facts of reality and human performance. It is unimportant whether Plato uses law and economics principles in order to achieve 'anti-capitalist' goals.

The *Nomoi* give not only theoretical but also very practical attention to key law and economics tenets. Plato's main emphasis is certainly on something else, but in the *Nomoi* he does create a legal system in which things are not taken for granted. We see, for instance, that in spite of general equality before the law, not all people have equal expertise; that not everyone might want, and thus should not be forced into, consumer-protecting interchange, and that a complex legal system, granting high damages, will not suffer from spurious claims. Merely to address a problem *prima facie* does not lead to its solution – rather, one must always remember what the goal of the law is and ascertain whether this is really fulfilled. We even find an early, probably the earliest, application of the Coase theorem. Therefore, this entry is not only justified, but necessary and inevitable.

Bibliography

Plato's works are universally cited according to the pagination of a sixteenth-century edition by Henricus Stephanus, without this usually being mentioned; which makes it possible to refer to the exact citation without listing an edition in the bibliography. This citation consists of a (page) number, small letter (indicating the paragraph), and second (line) number, although most editions do not give the last. Unfortunately this system is not very useful, especially for non-classicists, because it pays little attention to units within the text. Benjamin Jowett's translations into English are most frequently used; they also have the advantage of being available, under various addresses, on the Internet. The most relevant of Plato's works for law and economics are, as mentioned, the political dialogues, *Nomoi*, *Politikos* and *Politeia*.

Regarding the nature of Plato's political writings and their implications, next to Gadamer (1985, 1991a, 1991b, as well as several other essays on the topic), Salin (1921) is particularly interesting. Good summaries of the economic system of the *Nomoi*, or of Plato's economics generally, regardless of the normative judgement some of them make, include Espinas (1914), Bisinger (1925: 56–91), Pöhlmann (1925: 171–91, 238–45), Lacour-Gayet (1945) Rameil (1973) and Schefold (1989: 25–33). Among the best investigations of the legal aspects of the *Nomoi*, especially regarding criminal law, which take seriously the encompassing regulations as a system, are Shuchman (1963), Stalley (1983), and Saunders (1991), even if they do not start, or only incidentally so, from the classical-utopian character of the *Nomoi* as argued in this entry. Plato and law and economics are discussed by the author in Drechsler (1998, 2002b, 2003), on which the revised version of this entry is based.

Notes

1. The view of the *Politeia* as a utopia in which the individual counts for nothing at all, as an early form of totalitarianism, strongly sustained by Karl Popper (1966), overlooks that one of the most important motives of Plato's political thought, if not the central one, is the defence against threatening tyranny (especially Koyré, 1962: 97, 152). For a general critique of Popper on Plato, see Gadamer (1991a).
2. See, for example, the famous *Nomoi* 849e–850a, about which more below, or 921d (prohibition of interest); the classic statement in Pöhlmann (1925: 185–7, 238–41). However, see also Klingenberg (1982) and particularly Rameil (1973, esp. 76–82). Rameil locates the primary goal of Plato's economics in stability – economic as well as societal – but that only underlines these concerns: no system is as volatile as the free market, or capitalist, system.
3. This segment might be a display of what Gadamer calls Plato's 'Attic humor' (1991b: 230), since *Nomoi* 849e–850a could easily be an ironic way of showing that the economics of the market depends on state structures, institutions and rules to function at all.

References

Arendt, H. (1990), 'Philosophy and politics', *Social Research*, **57** (1), 73–103.
Bisinger, J. (1925), *Der Agrarstaat in Platons Gesetzen*, *Klio*, Beih, 27 (NF 4), repr. 1963, Aalen: Scientia.

Campbell, W.F. (1985), 'The free market for goods and the free market for ideas in the Platonic dialogues', *History of Political Economy*, **7** (2), 187–97.
Coase, R.H. (1960), 'The problem of social cost', *Journal of Law and Economics*, **3**, 1–44.
Drechsler, W. (1998), 'Platons *Nomoi* als Objekt der Rechtsvergleichung', in Werner, Häberle, Kitagawa and Saenger (eds), *Brücken für die Rechtsvergleichung. Festschrift für Hans G. Leser zum 70. Geburtstag*, Tübingen: Mohr Siebeck, pp. 44–61.
Drechsler, W. (2002a), 'Hans-Georg Gadamers *Plato und die Dichter*', in E.J. Pickerodt-Uthleb (ed.), *Zukunft braucht Erfahrung. 475 Jahre Gymnasium Philippinum*, Marburg: Gymnasium Philippinum.
Drechsler, W. (2002b), 'Les *Lois* de Platon, fondement de l'Économie du droit', *Revue Française d'Histoire des Idées Politiques*, **16**: *Les Lois de Platon*, 399–410.
Drechsler, W. (2003), 'Plato's *Nomoi* as the basis of law and economics', in S. Scolnicov and L. Brisson (eds), *Plato's Laws: From Theory Into Practice*, St. Augustin: Academia.
Espinas, A. (1914). 'L'art économique dans Platon', *Revue des Études Grecques*, **27** (122), 105–29, and **27** (123–4), 236–65.
Gadamer, H.-G. (1985 [1934]), 'Plato und die Dichter', in *Griechische Philosophie* 1 = *Gesammelte Werke* 5, Tübingen: Mohr Siebeck; English translation, 'Plato and the Poets', in *Dialogue and Dialectic: Eight Hermeneutical Studies on Plato*, New Haven, CT and London: Yale University Press, 1980.
Gadamer, H.-G. (1991a [1983]), 'Platos Denken in Utopien. Ein Vortrag vor Philologen', in *Griechische Philosophie* 3: *Plato im Dialog = Gesammelte Werke* 7, Tübingen: Mohr Siebeck.
Gadamer, H.-G. (1991b [1988]), 'Plato als Porträtist', *Griechische Philosophie* 3: *Plato im Dialog = Gesammelte Werke* 7, Tübingen: Mohr Siebeck, pp. 228–57.
Haliste, P. (1950), 'Das Servitut der Wasserleitung in Platons "Gesetzen"', *Eranos*, **48**, 142–9.
Klingenberg, E. (1982), 'Das Zinsrecht in Platons "Nomoi"', in H.J. Wolf (ed.), *Symposion 1977 – Akten der Gesellschaft für griechische und hellenistische Rechtsgeschichte* 3, Cologne: Böhlau, pp. 99–112.
Koyré, A. (1962 [1945]), 'Introduction à la lecture de Platon', in *Introduction à la lecture de Platon suivi de Entretiens sur Descartes*, repr. Paris: Gallimard, 1991.
Lacour-Gayet, J. (1945), *Platon et l'économie dirigée*, Paris: Union.
Pöhlmann, R. v. (1925), *Geschichte der sozialen Frage und des Sozialismus* 2, 3rd edn, Munich: Beck.
Popper, K.R. (1966 [1945]), *The Open Society and Its Enemies* 1: *The Spell of Plato*, 5th edn, 7th repr., London: Routledge & Kegan Paul, 1991.
Rameil, A. (1973), 'Die Wirtschaftsstabilität und ihre Problematik in Platos Gesetzesstaat', PhD thesis, Ludwig-Maximilians-University, Munich.
Ryle, G. (1967), 'Plato', in *The Encyclopedia of Philosophy* 6, New York: Macmillan and Free Press; London: Collier Macmillan.
Salin, E. (1921), *Platon und die griechische Utopie*, Munich and Leipzig: Duncker & Humblot.
Saunders, T.J. (1991), *Plato's Penal Code. Tradition, Controversy, and Reform in Greek Penology*, Oxford: Clarendon.
Schefold, B. (1989), 'Platon und Aristoteles', in J. Starbatty (ed.), *Klassiker des ökonomischen Denkens* 1: *Von Platon bis John Stuart Mill*, Munich: Beck, pp. 19–55.
Shuchman, P. (1963), 'Comments on the Criminal Code of Plato's *Laws*', *Journal of the History of Ideas*, **24** (1), 25–40.
Stalley, R.F. (1983), *An Introduction to Plato's Laws*, Oxford: Basil Blackwell.
Stigler, G.J. (1992), 'Law or economics?', *Journal of Law and Economics*, **35**, 455–68.

51 Wilhelm Roscher (1817–94)

Erich Streissler

Life and work

Wilhelm Roscher was born in 1817 in Hannover (in a German kingdom then still in personal monarchical union with Great Britain), the son of a high-ranking civil servant in the Ministry of Justice. After studying and early teaching at the University of Göttingen (doctorate 1838, university lecturer 1840), in 1848 he became professor of history and social sciences in Leipzig, a university distinguished by its law faculty. He served in this position up to his death in 1894. He is the main German-language textbook author of economics in the second half of the nineteenth century, the term 'textbook', however, by far understating the scientific importance of his comprehensive treatises. In its tripartite structure (theory–policy–public finance) roughly following the innovative framework of Karl Heinrich Rau, his predecessor as the dominant German economist a quarter of a century earlier, Roscher's *System of Economics* deals with theory in Volume I, *Foundations of National Economics* (1854), a book which ran to 26 editions up to 1922; followed by Volume II, on agricultural economics (that is, economic policy, part I), first published in 1859, a book of 14 editions; Volume III, published only in 1881, on commerce and industry (that is, economic policy, part II), running to eight editions; Volume IV, on public finance, published in 1886 (five editions); and, finally, Volume V, *System of the Poor Laws and the Policy towards the Poor*, published in 1894 (three editions), a third instalment of the treatment of economic policy which had become important by the end of the century. Of Roscher's numerous other books mention should be made of his monumental *History of National Economics in Germany* (1874), as Roscher was certainly the most knowledgeable historian of economic thought of the nineteenth century (altogether he surveys at least a thousand authors).

Framework of thought

Roscher presents himself to the reader as the economist who knows simply all aspects of economics and – in his huge number of footnotes – is conversant with what everybody else has written on the subject. This treatment of 'vast masses of material' has clouded the fact that he was of very considerable innovative power himself. Above all, Roscher had great intuition and an astonishing degree of sound judgement, which stands up well to scrutiny one and a half centuries later. But as he states no easily discernible general

principles explaining his innovations and does not stress their innovative character, they appear to the reader as just a part of his magisterial treatment of everything knowable. The only German economist not discussed at length in his history of thought is, unfortunately, Roscher himself. Actually, his theoretical innovations lie in what we would now call macroeconomics, in production and distribution theory, in business cycle theory and in regional economics; and he is innovative in examining thoroughly a vast number of institutional aspects, showing, in particular, how they are shaped by economic costs and benefits.

Of course, Roscher very explicitly gives a unifying principle to the innovative part of his thought: the 'historical perspective' or the 'historico-physiological method'. But this is more of a red herring than really helpful, for it actually means a number of different things. First, it means that Roscher illustrates his theoretical statements with empirical examples, taken in particular from ancient history, his broad knowledge of medieval and early modern legal arrangements from all over Europe and his copious knowledge of (more or less contemporary) statistics. He demands that theory be 'checked' against such facts, which, in modern terms, only means that he derives from the data rough estimates of parameter values of economic relationships, which themselves are basically 'given' to him a priori. (He would have denied the latter fact, which is obvious to the reader all the same.) Second, the historical perspective means that he is interested in long-term economic developments, which he usually analyses as applied economic theory. And third, Roscher shows how institutions shape factor supplies and how socioeconomic institutions and economic behaviour of the individual are shaped by production relationships, on the one hand, and by incentive structures, on the other. Such a 'historical perspective' is very useful to Roscher with his vast knowledge and his superb economic intuition, but is of little use to a mere student who should try to emulate him; it is not at all easily learnable.

On a formal level, Roscher follows what Streissler (1994) has called the German protoneoclassical framework of Rau and Friedrich von Hermann: all prices are determined by demand and supply, the former being at least as important as the latter. The determinants of demand and supply have to be examined in detail. All factor prices are just prices determined in the same way as goods prices. Price theory is thus general and symmetrical. Following Johann Heinrich von Thünen and other German authors, Roscher states explicitly the marginal productivity derivation of the remuneration of all three classical factors of production, demonstrating that the demand for factors is just a derived demand for goods. Cost functions are not necessarily, not even typically, characterized by constant unit costs. Following Hermann, his cost functions usually show increasing unit cost due to determinants external to the firm (increasing production draws firms with ever worse production func-

tions into the market). Roscher himself, and this is one of his important innovations, adds that, owing to determinants internal to the firm, particularly because of fixed costs, cost functions in industry frequently have decreasing unit cost.

To this formal framework he adds his own perspective of how the world 'works', a perspective evidently closely linked to the thought of Adam Smith: production relationships (which themselves may be shaped by some basic social or economic institution) shape historical development and the human mind. On the supply side, economic relationships dominate history. As far as legal arrangements are concerned, Roscher has a purely materialistic interpretation of history: economic costs and benefits shape laws, but never the other way round. On the demand side, on the other hand, individual preferences and reactions are strongly influenced by social institutions. It should be evident that this 'world view' is close to that of Karl Marx. But it is, of course, the economically liberal and morally conservative Roscher, writing one and a half to two decades earlier, who, both as to his 'world view' and as to his notion of increasing returns to scale, must be the original source, and not vice versa. (Alone among German academic economists, Marx quotes Roscher ten times in *Capital*, Volume 1, though negatively, of course, for ideological reasons.)

Because of the importance in Roscher of demand aspects, which are shown to be 'subjective' in spite of their frequent social codetermination, one could also say that Roscher is close to Carl Menger and thus to the Austrian school – Menger, of course, dedicating his *Principles* to Roscher. Incidentally, at least in Menger's policy teaching to Crown Prince Rudolph, Roscherian incentive aspects figure prominently.

Two examples: slavery and the small-scale firm

Two examples may illustrate Roscher's 'world view' of historical development. According to Roscher, economic development takes place in three stages. In the first stage, the factor nature predominates in production, in the second, labour and in the third, capital. Only in the third stage does a society become 'fully developed'. The second and third stages depend on the increase in supply of certain factors of production. (Note how Roscher turns the three classical factors into a theory of economic development, the third stage being the result of a marked accumulation of capital, a notion of Adam Smith's.)

Roscher then asks why the economy of classical antiquity never reached the third stage and thus stopped in a stage of 'immaturity'. For him, as for David Hume, the distinguishing characteristic of antiquity is slavery. (Note the similarity to Marxist periodization: slavery–feudalism–capitalism.) The further analysis is then purely theoretical economics, using, without saying,

the concept of a substitutive macroeconomic production function: Roscher says that slave labour in antiquity is cheap labour; therefore there is no incentive to substitute capital for labour (numerous examples are given); consequently, capital is not accumulated, credit is not developed and the rate of interest remains high; therefore the 'third stage', which depends on a large accumulation of capital, is never reached. Slavery is thus explicitly less the consequence than the cause of a low stage of economic development. If, on the other hand, capital accumulation takes place, together with technical improvement in instruments and machines, labour relations are transformed first to medieval serfdom and subsequently to modern wage labour.

The article in question was first published in 1849 and republished in an easily available essay collection in 1861. It certainly predates Marx. It also predates Thünen's publication of marginal productivity theory in 1850, which Roscher implicitly uses. A little later he says of himself that he has developed Thünen's 'natural laws' in the direction of their 'social and legislative' consequences. One might also say that Roscher extends Thünen's purely microeconomic assertions into macroeconomic conclusions.

Roscher adds that in antiquity, slavery also inhibited the development of a mass market for consumption goods and thus prevented their mass production and the development of the division of labour. On the other hand, he is, with Adam Smith, of the opinion that slave labour is always low-quality labour, inefficient because of a lack of incentives. With the high price of slaves under modern production conditions, slave labour is, in contrast to the case in antiquity, by now uneconomical and 'the freeing of slaves becomes imperative merely from properly calculating self-interest' on the part of their lords. Slavery is unprofitable or will soon be so in the United States (this was written in 1849 and repeated in 1861, before the American Civil War began), so that it will soon break down: a pretty, though unfortunately wrong forecast of the 'invisible hand' type and a mark of Roscher's basic belief in the beneficial nature of economic 'progress'.

Also in 1849, Roscher published his articles on 'Industry on a large and a small scale' and 'The economic importance of machine using industry', in which he develops his idea of cost functions in industrial production demonstrating declining unit cost. Large plants (implicitly in their production optimum) are characterized by relatively low fixed cost. For this statement Roscher gives the 'engineering' argument: a furnace that is ten times larger than another one will certainly not need ten times as much fuel and building it will not need ten times as many bricks, and so on. He presents a large number of statistical cost calculations, gathered from the empirical literature. They imply (in modern terms) scale elasticities between 1.065 and 1.175. Roscher draws the interesting conclusion that the greater competitiveness of English industrial products is to no small degree simply due to economies of

scale. Quoting numerous statistics from around 1850, he shows that English enterprises are on average larger than those on the continent. Of course, economies of scale lead to a displacement of handicraft firms by factories. But Roscher shows his superb judgement by pointing out that, in general, handicraft or artisan firms will not be fully replaced. They will continue to be active in sectors with a highly individualized, localized or temporally variable demand and, in particular, in service and repair firms. (Roscher shows that, at his time, in the industrially most advanced parts of the developed countries, the number of employees in small handicraft plants still exceeded by far that of factory workers.) Fascinating is his forecast (in 1849!) that, as a result of greater specialization in large firms, university-trained professionals 'without capital' will find new types of careers in the management positions of large joint-stock companies (while formerly careers had been available to them only in the civil service, the army and the church), so that these firms evidently open up 'a new and important moment of national liberty'.

Thus Roscher's whole discussion of economies of scale in industry runs only in terms of economic change and technical conditions in production causing socioeconomic developments. The legal arrangements making these developments possible are hardly mentioned. In fact, as Roscher frequently states, more rapid economic change and greater competition create the need for more economic freedom and for liberal institutions, not the other way round. (This argument is later repeated by Carl Menger.) Also, as Rau had already said, industry is too various and complex, so that government lacks sufficient information for effective regulation. Thus liberalism is 'produced' by economic development: Marx would have spoken of the substructure producing the superstructure.

Roscher on socialism

With Roscher, in contrast to Marx, it is liberalism and not socialism that is 'produced' by 'capitalist' development (or, more exactly, it is a slightly mixed economy). For the process of industrial concentration remains partial and, owing to the development of professional industrial managers, the middle class is increased and not diminished by 'capitalist' development. Furthermore, socialism as such is devastatingly criticized by Roscher (from 1854 onwards) in a large section of his 'theory' text in one of the most breathtakingly prescient forecasts ever written in economic literature.

The only comparable (in length and time of publication) treatment of socialism is the somewhat earlier one of John Stuart Mill in his *Principles* (1848). But while both Mill and, later, Marx consider property relationships in the narrow legal sense as essential to the understanding of socialism, Roscher thinks of property rights in the much wider modern sense, including the whole incentive structure as well as determinate responsibilities and the

remunerations corresponding to them, as the keys to understanding socialism. While both Mill and Marx are naively convinced of the selflessness of socialist man, Roscher, in spite of his frequent statements that one ought to consider the communitarian spirit of man just as much as his egoism, uses only the notion of the self-centred average man so typical of Adam Smith. Finally, while Mill treats socialism as merely a form of distribution, Roscher analyses it in the production part of his book: property arrangements are to be considered akin to a factor of production in a wider sense.

Roscher starts out with an argument in political theory: as communists are only interested in economic organization, they are basically indifferent between kinds of political organization as long as they serve their aim of economic revolution. In general, however, they will opt, if possible, for relentless despotism, because that will serve their aims best. He then asks whether community of property, once it is established (and without considering the 'terrible revolution destroying all culture' of its establishment), will be viable. Yes, he says, among either 'animals or angels', but also among human beings bound by true love. After all, every 'exemplary' family is a kind of community of property. In larger societies, however, such love is to be found only in the case of the 'highest religious enthusiasm' (he refers to the Acts of the Apostles), which enthusiasm seldom lasts long. Roscher explicitly states the free-rider argument: in general, every member of a community of property will try to work as little and to enjoy as much as possible; among 100 000 members, his own inactivity only concerns him to the amount of one in 100 000, which is 'as much as not at all!'. Perceptively, he adds that all self-interest will therefore be channelled into fights over distribution. Thus, under socialism, everyone will always harm the common good, while under the present economic system he does so 'only under exceptional circumstances'.

Roscher explains that because of this most theoreticians of a community of property add the idea of 'central planning of production and consumption either by an existing public authority or one newly to be established'. But this would imply 'a despotism which as yet has scarce ever had its equal in the world: a Caesaropapism which would also usurp the role of the private householder'. (Roscher, the concerned Protestant, identifies socialism with a kind of religion which demands the whole man and therefore leaves no space for religious freedom.) But even such an economic dictatorship will not assure 'active labour and thrift' because there are no incentives to motivate supervisors. 'At best, supervision will be lax', the term 'at best' implying that, most likely, the supervisors will be corrupt and bribable.

He then argues that political freedom is historically closely linked to free modes of production. And what would be gained by socialism? The economic equality, once established, will not last long: 'in spite of all laws', human

diversity both in abilities and preferences will soon re-establish inequality of wealth. Therefore, in the course of economic development, all forms of communal property have been progressively shed. Basically, then, socialism is an attempt to return to economic primitivism. The proximity of Roscher's argument to Friedrich von Hayek's *Road to Serfdom* is evident. Roscher, who always sees countervailing forces as well, ends on another note, however. There is also another 'no less important tendency': in the course of economic development the demand for public goods also increases, and private property owners cause an increasing amount of external effects on others. Therefore the 'realm of public responsibilities' is also extended. Roscher sees no essential conflict between increasing property rights and at the same time an increasing provision of public goods and increasing controls in the interest of other private individuals. Typically of German nineteenth-century tradition and of Rau's subtle reinterpretation of Adam Smith, the government of a liberal state to Roscher is also a strong government.

Roscher's analysis of socialism can be seen as an essay on the futility of legal establishments not supported by economic development. He repeats this theme again and again.

The shaping of human behaviour and of legal arrangements by production conditions and social institutions

While in his arguments on socialism he takes human motivation and behaviour as given and immutable relative to a mere change of property rights – and rightly so in his historical framework of thought, which deals with a relatively short-run analysis of a given stage of economic development – in the historically very long run, in contrast, Roscher considers preferences, work habits, human knowledge and the average state of economic information as shaped by conditions of production and the levels of per capita income. Quite explicitly, he pronounces the basic political constitution of a society to be the mere result of its modes of production (and never the other way round).

The reader of the first of Roscher's economic policy volumes to be published, that on agriculture, may be surprised to be first treated to a history of the development of the powers of government within the last five thousand years. Three important 'natural laws' explain everything: (i) the development of personal freedom and private property up to a state of general free competition within society, (ii) the progressive centralization of the state, and (iii) the development of all branches of industry (in particular of agriculture) from extensive to ever more intensive production.

The centralization of the state or of government is due to the tendency of ever-increasing responsibilities or purposes of the state with the rise in 'culture' (that is, national per capita income). This is due to the increasing

intensity of all production which causes more frequent and intensive economic interactions of individuals, which in their turn imply increasing possibilities for external effects. For Roscher, however, increasing state responsibilities and centralization are not in conflict with increasing liberalism. He says, for example, that ideally in a highly developed society public care for the poor is to be replaced by legally regulated, but private, insurance and state regulation of industrial employment is to be replaced by free contract. Thus increasing state concerns do not necessarily increase state activities. In fact, in the rude state of society there is neither need nor capability for much government activity. Government activity (particularly in industry and trade, as we are told in later books) then rises on both these counts, while the final stage will be free, but legally regulated, trade: 'Freedom of trade proceeds parallel to the establishment of a constitutional government'. Roscher is also clear as to the limits to all government: if opposed by strong economic forces, 'the law giver frequently gains not only relatively, but absolutely more if his demands are moderate'. Furthermore, Roscher is never one-sided. Centralization of government, for example, is to be held in check by the principle of subsidiarity, which also has much to be said for it – England in this respect showing the ideal example of a 'healthy equilibrium'.

As to the basic political constitution of a people, we are told that the agriculturist is trained 'by his strict dependence on nature to a blind obedience in human affairs as well'. Thus the prevalence of agriculture makes for a conservative and aristocratic government and – in the case of artificial irrigation – even for absolute monarchy. But the number of agriculturists declines in the higher stages of economic development; and industry and trade, and in particular small-scale artisan firms, make for democracy, and also for enlightenment and rationality, as traders and manufacturers constantly have to examine natural causes. (Roscher takes the materialist interpretation so far as to liken the art of Corinth and Rhodes in classical antiquity to that of Venice and the Dutch Republic, because all those were commercial people.)

As to agriculture, the incentive effects and the possibilities of control by the landlord in various systems of land holding are discussed extensively on very modern lines. The importance of legal security is stressed, 'which is to the advantage of not only those who gain at law, but to all'. Inheritance is of great importance, as it makes for long-term planning. Roscher explicitly explains, for example, the economic prosperity of the high Middle Ages as 'in particular due to fiefs having become hereditary from the middle of the eleventh century onwards'. Finally, competition is the panacea for many social ills.

Becoming more and more devout as he aged, Roscher in the book on public finance adds 'the sinfulness and the tendency to err of human beings' as an argument against absolute government and also concludes that for the

same reason no tax will be possible that causes no 'other disadvantages' besides the financial sacrifice. Tax collectors should get a share of the amount of taxes illegally evaded which they recover.

While Roscher's last book on the policy towards the poor, published when he was 77, contains little of novelty, his book thirteen years earlier, on trade and industry, contributes a wealth of interesting socioeconomic or legal-economic insights. Take the following three: (i) it is one of the legal advantages of joint-stock companies that, because of their greater permanency, they can hire more highly qualified personnel, especially better managers; (ii) if the obligation of the central bank to exchange its notes against coin be suspended by the government in a crisis, the bank is likely to fall victim to the government, which can then force it to finance any amount of government expenditure; (iii) improvements in transport and communication, particularly the introduction of the railways, will, by vastly cheapening transport and by levelling regional and social differences, increase the possibility of centralized control by the state, make for a common big city culture and create a national consciousness.

The book written in 1881 shows Roscher, the institutionalist, to particular advantage. In his materialist interpretation of all legal institutions he not only has to explain why legal arrangements of former times have become dysfunctional and how laws are first disregarded because of pre-eminent economic advantages and then swept away; he also has to show why these now dysfunctional arrangements had good economic reasons for them at the time when they arose. Of particular interest are the arguments for the (former) protection of trade in contrast to the (now to be recommended) freedom of trade. For, explicitly against Adam Smith, protection cannot have arisen 'purely out of error or even because of deceit'.

Sometimes the arguments may not be fully convincing, as when Roscher contends that the medieval rights of staple of certain towns are to be explained by the fact that, in times 'full of rapine and feuds', commercial capital was safe only behind the walls of towns, and that with the then very thin markets there was an advantage in concentrating trade at certain places (reduction of transaction and information costs for buyers).

More interesting is the argument for international protection. Its good economic sense, though as something 'conditional and transitory', is argued on infant industry lines (quoting Friedrich List) as a measure of education, of developing 'slumbering productive forces' and of stimulating industrial diligence. Also the initial advantage of foreign production might give the foreigner monopoly power. (Roscher repeatedly – and, of course, implicitly – refers to the Pareto inefficiency of monopoly due to the fact that it does not use resources socially optimally.) Thus time has to be exactly 'ripe' for a successful introduction of free trade: Turgot, for example, had introduced free trade

in France at too early a moment. Roscher adds a characteristic political explanation: protection had been introduced by absolute monarchs in order to help their bourgeoisie develop more rapidly and thus to weaken the power of their aristocracy. However, 'full' protection, by giving too much security, will ruin the habits of the manufacturers. Thus, in all matters of economic policy, according to Roscher, it is crucial to find exactly the right dose and exactly the right time.

References

Roscher, W.G.F., *System der Volkswirthschaft: Ein Hand- und Lesebuch für Geschäftsmänner und Studierende*, Stuttgart: J.G. Cotta:
- Vol. 1 (1854), *Die Grundlagen der Nationalökonomie*; 26th edn (with R. Pöhlmann) (trans. English, French, Russian, Polish, Serbian) 1922;
- Vol. 2 (1859), *Nationalökonomik des Ackerbaues und der verwandten Urproductionen*; 14th edn (with H. Dade) (trans. French, Italian, Russian, Polish, Hungarian, Swedish) 1912;
- Vol. 3 (1881), *Nationalökonomik des Handels und Gewerbefleißes*; 8th edn (with W. Stieda) (trans. French, Polish, Hungarian) 1913–17;
- Vol. 4 (1886), *System der Finanzwissenschaft*, 5th edn (with O. Gerlach) 1901;
- Vol. 5 (1894), *System der Armenpflege und Armenpolitik*, 3rd edn (with C.J. Klumker) 1906.

Roscher, W.G.F. (1861), *Ansichten der Volkswirthschaft aus dem geschichtlichen Standpunkte*, Leipzig and Heidelberg: C.F. Winter; 2nd edn 1878; reprinted, Düsseldorf: Wirtschaft und Finanzen, 1994.

Roscher, W.G.F. (1874), *Geschichte der National-Oekonomik in Deutschland*, Munich: Oldenbourg; reprinted, Düsseldorf: Wirtschaft und Finanzen, 1992.

Streissler, E.W. (1994), 'Wilhelm Roscher als führender Wirtschaftstheoretiker', *Vademecum zu einem Klassiker der historischen Schule*, Düsseldorf: Wirtschaft und Finanzen.

52 Emil Sax (1845–1927)

Manfred Prisching

Emil Sax, professor of economics at the German University of Prague from 1879 to 1893 and member of the Austrian parliament, tried to apply Austrian economic theory to politics, state and public finance, and further to all kinds of human communities and social associations. While the Austrian school emphasized methodological individualism and provided the basis for an analysis of the market economy, thus preparing the foundation for the libertarian views of later generations, Sax tried to develop a theory of public economics based on Austrian ideas. Thereby he gained insights into the relation of law, state and economy that deserve a closer look (Neck, 1989).

The analysis of individualistic and collectivistic phenomena

The mainstream of present economic theory confines its analysis to individual actions, even when collectivistic phenomena are examined. In Sax's view, however, social circumstances must be derived from two forces, individualism and collectivism. This is an astonishing point of departure for an Austrian economist who claims to cling to the individualistic approach, but nevertheless acknowledges collectivism as a feature of human nature as well as individualism.

Individualism is the inborn quality of man who considers himself the centre of the social circle. It is his attitude to relate everything, means and men, to himself. While the individualistic idea of personality is a product of civilization, the germ is present in every person (Sax, 1884, pp. 50f.). However, the efforts of individuals collide with those of other persons as far as individual spheres of activity touch each other, and evasion is impossible. People have a predisposition for common living, and their community is shaped by heritage, kinship, language, religion, historical reminiscence and nationality. Arguing against historical materialism as well as against exaggerated conservative ideas, Sax underlines that associations of individuals cannot be fully explained by economic processes.

Collectivism is the combination of individuals in a wholeness which commands individuals in order to fulfil collective purposes. There are communities such as state, church and nation: 'They are not a product of individualism, though a product of individuals' (ibid., p. 57). Sax is convinced that it would be an error to reduce institutions completely to individual actions.

Collectivism excludes individualism, as far as it pertains. On the one hand, collectivism constrains the free decisions of individuals when they interfere with other people; on the other hand, it combines and regulates individual activities that contribute to common functions wherever the free play of individual actions does not suffice, and individual egoism is constrained. Free activities of individuals are made possible without endangering social life through permanent conflicts and fights (ibid., pp. 63f.). At this point the *legal structure* enters the scene: it determines the relation between individual and community.

Both aspects, individualism and collectivism, are the basis for the classification of economic spheres: there is the private economy ('*Privatwirtschaft*') and the public economy ('*Gemeinwirtschaft*'). Public economics ('*theoretische Staatswirtschaft*') analyses the economic aspects of collective communities. We have further to distinguish economic actions and socioeconomic processes. Economic actions are set by individuals dealing with nature. By analysing models of isolated economic actions and using the inductive method for deriving general conclusions, one can find general economic categories. Socioeconomic processes ('*ökonomische Socialvorgänge*') are beyond individual control, and by analysing them one has to deal with basic social relations ('*sociale Grundverhältnisse*').

Sax draws attention to the mistake of identifying individualism with egoistic behaviour and collectivism with altruistic (or at least norm-directed) behaviour. Individual behaviour – as well as the behaviour of collectivities – may be egoistic (as it usually is in market relations), mutualistic (combining solidaristic elements with self-interest, as in collective insurance or class-based behaviour), and altruistic behaviour (as in families). By emphasizing both poles of human behaviour, Sax acknowledges the use of these concepts by writers as diverse as Adam Smith, Francis Hutcheson, Adam Ferguson, Karl Knies and Gustav von Schmoller, Émile Durkheim and Albert Schäffle.

Individualistic and collectivistic law systems

Historians tend to assume a process of increasing individualism through history, while economists presuppose the autonomous individual throughout time. Sax draws a somewhat different picture. History deploys a general process of the increase of individualism as well as collectivism, but more important is a sequence of cycles in which once individualistic and then collectivistic structures prevail. This picture of long-term cycles resembles Pitirim A. Sorokin's model (1937, 1953). Law is part of the framework that shapes the periods in question.

Individuals start living in tribes. The hardship of their fight against nature and against other groups demands that they join closely to secure food and defence. With increasing civilization, the need for collective strictness de-

creases. People gain increasing knowledge about how to use nature, states are consolidated and fighting becomes rarer. They have the chance of more individualism which is also promoted by property and the division of labour. Some peoples curb this process, for instance by developing caste systems; but there is a general tendency towards more individualism which eventually deploys egoism and degeneration. The sudden change comes with Christianity proclaiming general altruism, and youthful peoples contribute their more collectively oriented law systems. (Collectivistic) Germanic law replaces (individualistic) Roman law. Stronger ties lead into the Middle Ages, and people are strictly placed into collectivities again. But tendencies towards individualism start moving again, and they become mighty in the age of the Renaissance and Reformation. This is also the age of discoveries and innovations, of a rising monetary economy, and finally the age of traffic and technique. Again the individualistic Roman law is used, and the rule of law with its guarantees of liberties becomes anchored in constitutions. Thus the period culminating in the nineteenth century is an individualistic epoch. But Sax observes that, again, as time passes, collectivizing elements start gaining strength. Probably once more a more collectivist age is dawning.

There has been a long discussion about the nature and the genesis of the state. Sax criticizes individualistic theories which are trying to dissolve the peculiarity of the state, but he is equally critical of idealistic theories which shroud the state (as the realization of moral aims) in beautiful words or organistic theories which metaphorically equal the state to human bodies. Sax is also dissatisfied with realistic models, like the theory of Ludwig Gumplowics, which links the existence of states to occupation and the exploitation of groups (Sax, 1884, pp. 28ff.; 1887, pp. 106ff.). The economic theories of the state – exchange theory, consumption theory, productivity theory, capitalistic theory – are not convincing. Sax acknowledges merits of the new production theory, offered by Adolph Wagner, and the reproductivity theory, offered by Lorenz von Stein, but he again finds some flaws in them.

States are organizations for balancing individualistic and collectivistic impulses, attitudes and functions, and it is necessary to analyse the economic and legal background. There is one general historical tendency: the centralization of collective functions, 'a concentration of purposes from smaller to larger associations' (Sax, 1887, p. 424). The law of the expansion of collective functions which Sax borrows from Wagner shifts collective purposes from local to central authorities, from kinship, neighbourhood and community to region and, further, to the centralized state (ibid., p. 69). However, most economic considerations about collective institutions remain valid for all kinds of associations.

General economic categories and socioeconomic processes

Sax accepts the conceptual foundation laid for economic theory by Carl Menger, Friedrich von Wieser and Eugen von Böhm-Bawerk. Needs must be satisfied by goods, because man knows that he depends on them. Goods need preservation, efficient use and sustainability. They are produced by labour. The whole process is determined by subjective valuation. Goods may be used for consumption and investment. Capital serves the future satisfaction of needs. Economic decisions are guided by the comparison of costs and return (Sax, 1887, pp. 16, 28, 42ff.; see also Sax, 1916).

These processes are, in a certain sense, matters of technique and psychology, and they have to be distinguished from the second category, from socioeconomic processes. The description of these categories follows traditional paths. The category of *services*, however, is dealt with differently. Man needs fellow human beings in order to shape his relations with God, to realize his urge towards knowledge, to create emotional excitement through art, to arrange defence against enemies and struggle against diseases. People who provide services (which is essentially a non-economic activity) do not produce 'goods', but live from the redistribution of goods. Immaterial goods, like human capital or rights, are not goods (Sax, 1887, p. 33) and Sax notes that, in excluding rights and relations from the concept of goods, he is in agreement with Böhm-Bawerk who also argued for a conceptual difference in his *Rechte und Verhältnisse* (1881). The distinction seems to be influenced by the classical dichotomy of 'productive' and 'unproductive' labour. Sax notes that not all economic processes can be depicted as processes of production and that they are not necessarily productive (1887, p. 202). But there may also be some further theoretical potential in the distinction: services are really different from goods in several characteristics, such as storage, insecurity, quality, price negotiations and qualification of labour.

The economic structure gets its legal complement, because these categories and relations are secured and shaped by law. Possession, for example, becomes private property: the right of individuals to do with their property whatever they like is secured, as long as others are not afflicted to an extent that impedes general social development. According to Sax, private property is not a matter of course, it is not a necessary element of human personality and it has not been founded by a decisive act of legislation. The legal arrangement of property is an expression of the tensions between individualism and collectivism, and possession is the outcome of human egoism shaped by the endeavour to maintain human existence under the pressure of the environment (ibid., pp. 149f.). Permanent fighting is ruled out by individual and collective protection of rights, including the maintenance of minimal living standards for all individuals through means provided by the collectivity. Other economic categories are also shaped by the law.

Collective needs and public goods

Sax's treatment of the theoretical foundations is not particularly surprising (with the exemption of services) and more interesting is his thinking as regards the public sphere. Sax acknowledges *collective needs* that spring from the common living of men in relation to their environment, and he criticizes Wagner's proposition that purposes of power and law, on the one hand, and purposes of culture and welfare, on the other are essential functions of the state. 'It is not the abstractum state that can put up purposes, and that is able to feel and act, but only the concrete inspired individuals, its elements, that are able to do that' (Sax, 1887, pp. 191f.). But collective needs cannot be dissolved into individual needs. Individuals sense these needs only as members of a community and, incidentally, not all citizens at all times sense these needs. Therefore collective needs are an expression of individual feelings, but these feelings have their origins in collective ties and aim at collective purposes. Individuals may not be conscious of collective needs, and sometimes their implementation must be coerced, but if they are conscious of them, *they order them according to their intensity, alternatingly and together with individual needs*, and they get their satisfaction from them as in the private sphere. More complicated is the process with collective goods proper from which no tangible benefits for individuals can be derived: the valuation of these goods is done *collectively*, but Sax is not quite clear about how the process works. It is obvious that individuals usually have stronger feelings for their individual needs, but sometimes they are also prepared to sacrifice blood and property for national honour and dignity (ibid., p. 195). Mutualistic and altruistic elements are always present in economic assessments.

Collective or public goods are in the possession of the collectivity: Sax prefers the legal to the economic definition, while the latter was to dominate the discussion in modern theory (ibid., pp. 26 and 36). The value of goods is determined by the individuals' preferences, by their equating the marginal utilities of different goods. But a kind of average value emerges from the social context: the evaluation is not a completely individual decision. In the public sphere, *taxes are an expression of collective values*. Executive bodies of collective life carry out the assessment in their function as representatives of single economic units. Expenditures are justified when individuals appreciate the collective goods produced with their money or their private goods more than the goods they would have possessed or acquired if they had used their resources for private purposes. It must be the aim of politics to realize an equilibrium between individual and collective needs. Public revenues are a phenomenon of valuation: 'A Robinson and a 100–million Empire obey in their actions the same law: the law of value' (ibid., p. 308).

There is a *valuation of share* ('*Antheilswertung*') – one assesses the amount which individuals get from the collectivity, and the percentages of single

purposes – and a valuation of costs ('*Kostenwertung*'): one assesses the amount which is necessary for satisfying collective needs and which has to be collected for goods, services and altruistic actions. Revenues and expenditures have to be *mutually* adapted. By consenting to the whole amount of collected resources the individual decides on his own share, and by having a certain imagination about his contribution to the community he influences the whole amount collected. Sax is not specific about how the process can be imagined in detail, but he objects to the solutions proposed by Wicksell (1896) or Lindahl (1919). The valuation of collective goods proper is shaped by influential political groups and elites. Actually the decisions are reached by the political representatives, but they have also to adapt to the wishes of their voters. By this process, we can consider the valuation of costs a collective, societal assessment of value.

Sax rejects the formula, well known in his time, that in the private household income determines expenditures, while in the public sphere expenditures determine revenues: the process of *mutual* adaptation of revenues and expenditures describes a different procedure. And he modifies the rule of thumb that public investment, durable consumption goods and irregular expenditures have to be covered by the extraordinary budget, that is, by public credit. This depends, as in private decisions, on the valuation of present necessities and future possibilities, therefore irregular expenditures could well be covered by current revenues which can be increased *ad hoc* for this purpose. Sax is fascinated by the simple explanation that unveils the nature of the public financial world. It is a solemn phenomenon that we observe in the collectivist household 'a living together not only of millions of contemporaries that have little in common but also the communication of different generations that are often far from each other. Thereby, the equilibrium of numerous collective and individual needs in such large dimensions of space and time is maintained with the highest level of human perfectibility. And that happens by the same psychological forces that guide the smallest, inconspicuous economic unit' (Sax, 1887, p. 380).

Collective purposes and public institutions

The overall objective of an association of people is social welfare maximization. The structuring of collective bodies takes place in accordance with the preferences of the individuals. The law is the expression of their wishes and needs, and the administration is the executive instrument:

> Collectivism expresses its influence by integrating the individual in the whole feeling and the whole will in regard to common living. Because of this motivation individuals as members of the association realize concrete purposes in their relative importance for common life, with the clear consciousness or the dark feeling that everyone shares it, and they yield to the necessity of withdrawing essential

> goods from individual purposes … The consent of will is expressed by common decision [usually by organs of the community]. (Sax, 1924, p. 218)

We have to distinguish two categories of state activities. First, there is the *regulating administration* ('*regelnde Verwaltung*') setting incentives for individuals; it is justified by the necessity of coordination purposes as well as by decisions of the majority to enforce certain aims. Second, there is the *acting administration* ('*selbstthätige Verwaltung*') through which the state supplements or replaces activities by individuals; it is justified in cases when favourable individual actions are missed, impossible or detrimental, as in cases of market failure or merit goods (Sax, 1887, p. 65).

Public enterprises and institutes serve to fulfil the tasks of government. *Public enterprises* provide private goods, like railways. The 'prices' they demand ('*Taxpreise*') are similar to the prices that would be demanded by private firms, but deviations might appear because of political considerations and long-term calculations of the state. Public supply is justified by decreasing average costs. The state will not demand monopoly prices because, owing to its long-term perspective, it will rather aim at the advantages of mass production. If the state nevertheless demands higher than competitive prices, the situation is comparable to the case that an indirect tax is surcharged on private goods. Because of the lack of competition, there is also the possibility of differentiating prices according to individual assessments of the goods by consumers (ibid., pp. 73–5). But as there is no mechanism to reveal demand, it is essentially a political decision to balance private and public needs.

Public institutes ('*öffentliche Anstalten*') provide merit goods, for instance education. There is a stronger element of collective purposes in the goods supplied by public institutes, and they are financed by fees ('*Gebühren*'). The rates paid by consumers may be created differently: they may consider individual ability to pay and mirror political decisions about the burdening of social groups, but they have also to take into account considerations of efficiency, they must be sufficiently simple to avoid unnecessary transaction costs (ibid., p. 475). Sax acknowledges different degrees of 'publicness': the difference between collected fees and costs of the good provided is the measure of 'collective content'. However, the supply of goods free of charge should be an exceptional case: 'Even the "wrong communism" of the complete abolition of a fee may sometimes take place, but the collectivistic egoism of the community directed towards single individuals is generally strong enough to balance such endeavours, and often it drives the decision to the opposite direction' (ibid., p. 477). Acquainted with the development of modern interventionist states, we will have to add that this assumption was rather optimistic.

Systems of needs and principles of taxation

In the case of *general* or *pure public goods*, it is impossible to ascribe a certain amount of utilization to an individual (Sax, 1887, pp. 71 and 80). This is obvious in cases like politics, general administration, science, legal order and the avoidance of epidemics. These expenditures are covered by taxes. According to Sax, dominating tax theories provide no explanation. They speak of the moral obligation of individuals to provide the means for the state's existence. They resort to unexplained ideas of justice or an equalization of sacrifice ('*Opfertheorie*'). They use the idea of ability to pay ('*Leistungsfähigkeit*', '*Billigkeit*', '*Steuerfähigkeit*', '*Steuerkraft*'). But they do not offer an economic explanation of the tax system. But tax laws have to be explained by economic theory, not by recourse to morality (ibid., p. 524). And the economic explanation can only be based on the idea of a sequence of private and public needs that are ordered according to their utility.

If individuals had a true knowledge of the facts, they would assess their taxes according to their demand for collective goods, and government would only imitate the individual valuations. The consent of every individual is not necessary – later theorists will argue over the free-rider problem. In Sax's view, it is overcome by general consent: by the political decision that is passed in the framework of a legitimate constitutional system. Individual consent is an 'as if' argument, an intellectual fiction (Sax, 1924, pp. 207f.). In implementing the tax system, the state has to take into account unrealistic assessments of individuals, and it has to counter individual egoism. But the central point is that taxes are the *expression of collective valuations*.

Sax pleads for progressive taxation. His arguments are not based on the simple, but finally unconvincing, idea of the declining marginal utility of rising amounts of money which is usually offered following Carl Menger's theory (Menger, 1883; Sax, 1892, p. 54). We develop Sax's model in a number of steps. First, Sax resorts to the model of a *pyramid of needs* which is similar to Abraham Maslow's (1954) pyramid. At the bottom of the pyramid, there are existential needs which have an infinite intensity; then follow needs that are preconditions for a decent life; then demands connected with socially appropriate standards of life; beyond that level we meet higher cultural aspirations; and finally there are casual impulses satisfied by sheer luxury. In Sax's view, it is the *quality* of the needs that provides the justification for a progressive tax scale: goods at the higher levels of the echelon can be more easily forgone than those at the lower levels.

Second, the intensity of needs is so different that they cannot be depicted by proportional relations, as models of the theory of equal sacrifice have assumed: 'We know that the decline of the intensity of needs and therefore the decrease of value advances, until a certain point, faster than the corresponding amount of goods (income) increases … From this fact we can ex

ipso conclude that the progression of the tax has to take place' (Sax, 1892, p. 93). The intensity of needs would decline relatively fast, but Sax accepts Robert Meyer's (1887) idea that the decrease is, to a certain extent, compensated by a *greater variety* of needs and goods at higher levels of the pyramid. Sax expresses his hope that further statistical research would help to clarify this question, and he refers to the statistician Ernst Engel for recent findings.

Third, according to the complicated structure of the intensity and variety of needs, tax cannot have a regularly designed structure, and it is inadequate and a fantasy to look for a general mathematical formula. There are only broad classes of taxpayers in relation to the pyramid of needs: one could imagine a tax-free basic income, and the largest jump in the rate must necessarily occur at the step from the physical to the social minimum of living; but the further differences of tax rates in relation to income levels are shaped by the specific assessments of society.

Fourth, individual circumstances can be taken into account for the design of the tax structure, for instance whether there is a whole family that depends on a certain amount of money, or only a single individual who has the same amount of resources at his disposition. Obviously, it is necessary to reach a certain average treatment of groups of taxpayers. In reality, the tax structure will be a combination of both standards: individual property (income) and individual needs (Sax, 1887, p. 514). A just taxation leads to the reduction of opportunities for individual development that is the same for every member of the society.

Fifth, no theoretical prescription for a tax schedule is possible. The tax system is the expression of subjective valuations of the citizens, and the final outcome in legislation is determined by the political process in which group interests and strategies of political elites have their roles. It is similar to an experimental process, a trial-and-error procedure.

Sixth, Sax is against simple tax rates or single tax systems. Taxes can only be conceived as a complex of taxes that are composed as parts of a system. No individual tax can be assessed without the background of the whole tax system. Not every tax needs to be progressive; a tax may have the function of compensating the exaggerated progressivity of other taxes (Sax, 1892, pp. 43ff.). There are also special taxes ('*Zwecksteuern*', '*Spezialsteuern*'), taxes in the framework of a federal constitution, contributions ('*Beiträge*') or specific taxes ('*Umlagen*'). The 'art of taxation' lies in the appropriate combination.

The sequence and intensity of private and public needs is permanently changed and reassessed, and the political conflict continues. Sax wants to trace the rules applied to public and private economic activities back to common principles. In his model of taxation the same principles guide the absolute amount of the public budget and the distribution of the burden for

individuals, the revenues and the expenditure side of the budget, and the private and public decisions of individuals.

References

Böhm-Bawerk, Eugen von (1881), *Rechte und Verhältnisse vom Standpunkte der volkswirthschaftlichen Güterlehre. Kritische Studie* [Rights and Relations from the Perspective of the Economic Study of Goods. A Critical Analysis], Innsbruck: Wagner.

Lindahl, Erik (1919), *Die Gerechtigkeit der Besteuerung. Eine Analyse der Steuerprinzipien auf Grundlage der Grenznutzentheorie* [Justice of Taxation. An Analysis of Tax Principles based on Marginal Utility Theory], dissertation, Lund.

Maslow, Abraham (1954), *Motivation and Personality*, New York: Harper & Row.

Menger, Carl (1883), *Untersuchungen über die Methode der Sozialwissenschaften und der Politischen Ökonomie insbesondere* [Investigations into the Methods of the Social Sciences, with Special Reference to Economics], Leipzig: Duncker & Humblot.

Meyer, Robert (1887), *Das Wesen des Einkommens. Eine volkswirthschaftliche Untersuchung* [The Nature of Income. An Economic Analysis], Berlin: Hertz.

Neck, Reinhard (1989), 'Emil Sax's contribution to public economics', *Journal of Economic Studies*, **16** (2), 23–46.

Sax, Emil (1884), *Das Wesen und die Aufgaben der Nationalökonomie. Ein Beitrag zu den Grundproblemen dieser Wissenschaft* [The Nature and Functions of Economics. A Contribution to the Principles of Economic Science], Vienna: Hölder.

Sax, Emil (1887), *Grundlegung der theoretischen Staatswirthschaft* [Principles of the Theory of Public Economics], Vienna: Hölder.

Sax, Emil (1892), 'Die Progressivsteuer' ['The Progressive Tax'], *Zeitschrift für Volkswirtschaft und Sozialpolitik*, **1**, 43–101.

Sax, Emil (1916), *Der Kapitalzins. Kritische Studien* [Interest of Capital. Critical Studies], Berlin: Springer.

Sax, Emil (1924), 'Die Wertungstheorie der Steuer' ['The Valuation Theory of Tax'], *Zeitschrift für Volkswirtschaft und Sozialpolitik*, NS, **4**, 191–240.

Sorokin, Pitrim A. (1937), *Social and Cultural Dynamics*, New York: American Book Co.

Sorokin, Pitrim A. (1953), *Kulturkrise und Gesellschaftsphilosophie*, Stuttgart: Humboldt.

Wicksell, Knut (1896), *Finanztheoretische Untersuchungen nebst Darstellung und Kritik des Steuerwesens Schwedens* [Investigations in Fiscal Theory, Including a Description and Critique of the Swedish Tax System], Jena: Gustav Fischer.

53 Gustav von Schmoller (1838–1917)

Helge Peukert

The man and his time

Gustav von Schmoller, the main representative of the younger German Historical school, was born on 24 June 1838 in Heilbronn, where his father had been *Kameralverwalter* of the Württembergische fiscal interests since 1833, a family tradition deeply influencing young Schmoller and which goes back to 1651, when an ancestor became a public servant. After finishing Gymnasium in Stuttgart in 1856, he stayed for one more year in his father's office, learning a lot about financial and administrative law. In 1857, he began to study financial, state and administrative law at the University of Tübingen as the start of a career as a civil servant. Furthermore, he attended lectures in philosophy, history (under Max Duncker) and even the natural sciences (for his biographical background, see Balabkins, 1988, chs 1–4; Kaufhold, 1988).

The combination of economics (called 'national economics' at the time) and history is already manifest in his first major scientific undertaking 'Untersuchung der volkswirtschaftlichen Anschauungen zur Reformationszeit', written in 1860. It is a study of the economic conceptions of the Reformation period, with which he won first prize in a competition and for which he was awarded his PhD. In this study the main question of his intellectual life already played a major role: the tension between the necessity of an individual sphere and a strong, market-regulating state. The second part of his time as junior barrister he passed at the Württembergische statistical office headed by his brother-in-law, G. Rümelin, a friend of his father. He thoroughly analysed the census of industry of 1861 (1863) and in 1864 become professor at the University of Halle, where he published 'Die Arbeiterfrage' (the labour question) in which another tension of his economic approach stands out: on the one hand, he argues against the free marketeers and Manchester liberals like John Prince-Smith and their '*Kongress der Volkswirte*' and their strict and general principle of non-interventionism. On the other hand, he argues against the socialism of Ferdinand Lassalle and against Marxism (in so far as his polemical labelling as a 'socialist of the chair' by H.B. Oppenheim is at least slightly misleading). He developed a different solution to the social problem (we face it again today in the presence of the globalization of production and capital movements) and to the relationship between capitalists and workers which he did not define as a zero-sum power play but as a productivity-enhancing cooperative relationship (a careholder value of the

firm concept) where both sides can realize (material) gains if rules and compromise are instituted (foreshadowing the concept of a social market economy, the '*Soziale Marktwirtschaft*'). He argues against the left and the conservatives like H. von Treitschke (1875), who was convinced that accumulation and work discipline could only be upheld together with a very low subsistence level income of the working class (today this opinion might be called a 'natural level of misery').

In 1862, he published (first anonymously) a short tract about the French commercial treaty and its enemies. He argues in favour of the Prussian–French treaty and against a treaty with Austria, a solution which was favoured by most southern German states and which clearly reveals his conviction that the future of Germany and its economy stands and falls with Prussia, which should unify the many small German states and abolish the manifold tariffs and other 'mercantilist' trade restrictions, on the one hand, but at the same time develop the backward German economy by establishing a dense legal–institutional framework (the market as an instituted set of legal rules) including interventions in the distribution process to solve the social question (exploitation, long working hours, no insurance systems, child labour and so on). He believed in a benevolent, efficient and ethically motivated state and its bureaucracy in a monarchy which should include the best elements of parliamentarism and a liberal state, where the monarch should be neutral and above the class and interest groups, pursuing the common good.

With this perspective in mind, Schmoller studied the constitution, administration and economy of the Prussian state intensively (which was his great love and weak point as far as objectivity is concerned, but which does not necessarily invalidate his general policy conclusions). Although looking backwards in his historical research, as in his history of German small business in the nineteenth century (1870), his economic policy conclusions had a very topical perspective because in 1869 the principle of freedom of trade (*Gewerbefreiheit*) had been introduced in the commercial laws of some states. Schmoller was arguing that then and now there are some legitimate fields where the intervention of the state is for the common good, to solve the social problems and to support industrial development.

In 1872, Schmoller received a call to Strasburg where in 1879 he published a study about the Strasburg cloth and weaver guild. He stressed again that a certain mercantilist interventionism is not entirely bad and may be in the interests of social harmony and economic development. After going to Berlin in 1882, he edited the *Acta Borussica* (1882), a broad compilation of source material, a result of his empirical archive studies, following his methodological principle of inductive–empirical research methods (his reproach against members of the older Historical school like Wilhelm Roscher, Bruno Hildebrand and Karl Knies was that, for example, they constructed stage

theories without basing them on sound empirical evidence – what Schmoller called false abstractions). In 1900/1904, he published his major theoretical contribution, the 1400-page heritage of the Historical school, the *Grundriss der Allgemeinen Volkswirtschaftslehre*, in which he wanted to delineate his approach to the general questions of economic science. In it another tension in Schmoller comes to light: on the one hand, he was an ardent defender of empirical–inductive research, on the other, he had a strong inclination to generalize, that is, to apply a more deductive–'theoretical' method. The interesting methodological point in Schmoller is not that he was simply a theoretical 'historicist' but that he tried to bridge the great (and maybe unresolvable) antinomy between a bottom-up and a top-down research method.

Besides giving innumerable lectures, public talks and reports, he became a member of the Prussian state council and rector of Berlin University in 1897/98. From 1899, he represented the university in the Upper House (*Herrenhaus*) and became a member of the order *Pour le Mérite*; he received many honorary doctorates (including one in 1896 in Breslau from the Law Faculty) and corresponding memberships inside and outside the country. Although his direct influence on the social policy of the Bismarck state (which he criticized in a very direct manner, even going too far in some respects) was rather weak, his main achievement at this time has to be seen in the foundation of the 'Verein für Socialpolitik' (Lindenlaub, 1967) in 1872, in Eisenach. It was inaugurated to contribute to the evolutionary solution of the social question by scientific research and policy recommendations: a young generation of economists without dogmatism but with ethically minded enthusiasm should help to abolish social misery. Against other voices, Schmoller preferred an 'open society model', where no scientist should be excluded for his economic or political opinions. A major achievement of the 'Verein' was its annual meetings, where all opinions were voiced. The accompanying monographs, from the question of the development of seaports to the lofty problems of monetary and fiscal policy, are examples of sound theoretical and statistical research. Once again a main tension in Schmoller became apparent, one of the real dilemmas of the life of the mind, the dilemma of the ascetic and the worldly, of detachment and involvement, the dispassionate and (what Max Weber called) value-free research versus the formulation of tentative solutions to build up a good society, the desire to know the causes of things versus the desire to change them.

Schmoller died on 27 June 1917, in Harzberg, so he did not witness the final defeat of the Prussian state and the monarchy (as a result of the aggressive expansionism of the young German Reich in the First World War) on which all his hopes were placed; Schmoller as a moderate nationalist supported the war with some enthusiasm. Historically, he was both right and wrong at the same time. The Prussian monarchy was dead, but the failure of

the Weimar Republic was that it was a state and bureaucracy above everyday interests. A (republican) ideal was missing. The middle class had no common denominator, social policy was of no central concern and in the great depression the role of the state was defined as a 'hands-off' policy (invisible hand plus tight monetary and no fiscal policy beyond the balanced budget concept). The chaos and tyranny after 1933 was even worse than that which Schmoller tried to make us believe would be the result of a non-interventionist policy.

Schmoller on law and his economic theory

The mainstream opinion about Schmoller and the Historical school can be summarized as follows (see Fischer, 1968; but compare the more receptive accounts in *Journal of International and Theoretical Economics*, 1988, pp. 524ff.; Bock et al., 1989; Schiera and Tenbruck, 1989; Backhaus, 1993; Giouras, 1994): he proposed a historical–realist, psychological–ethical, inductive–descriptive programme which should include law, morals and culture. The research programme was more historical than economic and some blame him for having delayed scientific progress in economic theory (and not only in Germany). He has even been held responsible for the inflation of 1923 (Barkai, 1991) by neglecting simple quantity theoretical connections and so influencing the economists of his time and long after. According to this view, his theory was neither coherent and worth consideration nor an alternative to neoclassical mainstream thinking.

Schmoller's weakest point was his methodology (a subject he was not really interested in), which became apparent in the so-called 'first debate about method' ('*der erste Methodenstreit*') between him and the Austrian, Menger. Schmoller's 20-page review (1883) of Menger's *Investigations* (1883) and Dilthey's book about method (1883) was the start of a series of (self-) misunderstandings on both sides. Menger could not solve the problem of distinction between empirical and exact laws, between explaining and understanding, a representational versus a constructivist approach which showed up in Schmoller's work as the problem of induction versus deduction (compare Schmoller, 1949, 1900, ch. II. 4, and Menger, 1884). Both in principle asked the same questions and both failed (see Peukert, 1997, chs 3 and 4).

The point here is that Schmoller cannot be linked to one methodological camp. At the same time he held the succession model (first induction, then deduction), the division of labour model (induction and deduction as complementary activitites), the imperialist model (induction is better than deduction) and the pragmatist model (deduction is bad because it leads to Manchester atomism). For Schmoller, on the one hand, classification develops by itself and according to empirical findings; on the other hand, all thinking is abstraction and all mature science is deductive. Theoretical elements of David Hume,

Immanuel Kant, John Locke and others can be detected in his reflections. The antinomies between a natural science and a philosophical–sociological, a psychical and a physical–organical approach, the freedom of will versus necessity, and the ethical individual motivation versus the determination of action by the environment stand out and none of these tensions is ever resolved. It may be mentioned here that all his major findings were already apparent before he discovered the social problem and before he had undertaken any historical or empirical studies (see Schmoller 1860, 1864/65).

In one sentence only (1883, p. 247), Schmoller raises the fundamental substantive question: what holds (capitalist) economies together and directs their development – exchange, value, money and so on (Menger's quasi-Aristotelian basic categories) or interdependent social organs, institutions and rules?

Schmoller's institutional theory is not easy to summarize because in his exposition a critique of the classical economists, educational elements for civil servants, his social policy views and a positive theory intermingle and cannot be described by some fundamental hypotheses; another reason may be his use of the German language (only a few pages of his *Grundriss* have been translated into English).

The starting point in *Grundriss* (1900/1904) is not a utility-maximizing individual and an abstract concept of exchange, but the fact that economic life is embedded in political and social structures. They are a manifestation of the socialization of law, custom, morals and religion. Historically, the main trend lies in the enlargement of the resulting socioeconomic organizations and institutions, from tribes in hunter and gatherer societies to little kingdoms during medieval times, right up to the European Union (EU). The intentions, problems and irritations, and the specific combination of economic, political and cultural factors that affect the introduction of the single European currency can be seen as a nice example of Schmoller's organic theory of institutional evolution as a precondition to more or less efficient markets.

Although the aim of economic action is the increase in commodities and wealth, the great miracle of the economy is the productivity-enhancing quality of cooperation and coordination. One precondition is what Schmoller calls the accumulation of symbolic capital (like language, writing and education), which socializes the individual and makes him or her able to cooperate and share the perspective of others (Adam Smith's impartial spectator). Anthropologically, man is driven by down-to-earth material pleasure and pain and the drive of survival and rivalry, but also by the drive of social approval (one of his socially integrating appetites). Schmoller holds a *homo duplex* view of the human condition (individualizing and socializing tendencies, in contrast to the one-sided *homo homini lupus* preconceptions of, for

example, game theory). The state has to set the limits for individual action, in whose compounds bilateral and multilateral negotiational compromise organs are at work to make cooperation a success.

In the social cooperation game, the guarantee and protection of individual property rights is the core of the law, but the more complicated and interdependent economies become for reasons of density and sheer numbers (and, we might add, the increase in external effects of different kinds) the more legal, negotiational and state institutions have to increase their regulating potential. To put it into the paradox formulation of Commons, they have to organize 'collective action in control, liberation and expansion of individual action' (1950, p. 21). Schmoller's view therefore was that there exist beneficial, necessary and complementary positive feedbacks of individualizing and socializing tendencies in modern economies. The necessary tension and duality of economic life and its institutions can be found not only on a macro level but also on the micro level of all private organizations and institutions, such as insurance companies, where two forces are at work: a purely private law and an egoistical tendency. But there is a human public side too, operating in the interest of the common weal. The businessman concentrates on the first aspect, the social politician on the second. To find a compromise between the two tendencies (which are not identical according to a naive and simplified invisible hand understanding) is the art of economic and political action.

Does Schmoller (1904, bk 3, pt 4, concerning value and prices) have a specific concept of markets, equilibrium, competition and so on – the core questions of present-day economics? Most economists doubt it (for an opposite view, see Priddat, 1995; Peukert, 1997; and to a certain degree Schumpeter, 1926). For Schmoller, a given price is taken as the starting point (which has some inertia because the economic agents prefer a certain predictability in an uncertain world). Supply and demand exert some pressure on it. But for Schmoller, first, 'supply and demand' are only abstractions for more important underlying forces. Second, they are not evident but have to be detected by economic agents (depending on institutional factors like information services and legal obligations to inform the public, as in the case of joint-stock companies). Third, prices must be interpreted (the question whether prices go up or down in the future is more important than the recent price as such).

In real markets in time and space without abstract auctioneers and costless full information, actors with restricted knowledge enter the stage, although they have a certain opinion about the price ranges. They first orient themselves to get a clearer picture and they may correct their hypotheses in a quasi-hermeneutical speech situation (questioning, arguing, evaluating). For Schmoller, the market process is active haggling and bargaining, depending on the prevailing thought maps which themselves influence the 'fundamentals'. For him the typical market scheme situation is the bilateral monopoly as

the ideal-type or standard case, like the supply and the demand sides, entrepreneurs and employees: a certain number of actors on two sides of market transactions, foreshadowing modern transaction cost analysis, according to which, as a result of asset specifity and lock-in effects, most market transactions lead to a fundamental transformation. 'Joined as they are in a condition of bilateral monopoly, both buyer and seller are strategically situated to bargain over the disposition of any incremental gain whenever a proposal to adapt is made by the other party' (Williamson, 1985, p. 63). Schmoller realized that the times of small business and competitive markets in the strict sense belonged to the past. He thought about economics when, after the German Reich's unification in 1871, his country was in full transformation towards a mass production economy (unfortunately, he saw cartels in a much too rosy light, as if they were economic institutions to smooth out conflicts and price volatility, although he was not an unrestricted proponent of them).

Schmoller points out that actors on micro and macro levels have different competencies and autonomy, depending on the legal framework defining non-neutral rules of the game (for example, to withhold basic services and goods for the other market side), business relations, knowledge, motivation, negotiational abilities, education, market power, stockpiles, financial reserves and transaction urgency, which means 'the abilities of the parties to take and to inflict losses during stalemates … Toughness in the sense of unwillingness to yield in a range in which the other party is expected to yield if one fails to do so' (Fellner, 1949, pp. 27–8). These are some of the manifold forces and influences which are summarized as 'supply and demand conditions', some of which cannot be expressed as numbers and equations.

Although Schmoller did not welcome or even grasp the results of more formal reasoning in his time (that of Joseph Bertrand and Francis Y. Edgeworth) he very well understood that in oligopolies and bilateral monopolies a formal derivation of the equilibrium (price) is usually not possible (Stigler, 1966). It is not passive adaptations to given market prices and the anonymous mechanical market mechanisms that take place, but strategies economic players play, which lead to Bowley, Stackelberg and other solutions. They are the essence of competition depending on the parties' relative power positions in the sense characterized above. Nor can distribution be determined by strict formal–mechanical reasoning, and the strategies the players choose cannot be derived in a straightforward way by adding up the structural factors of the bargaining situation mentioned above. This is the substantive reason for Schmoller advocating a more historical method rather than abstract models. A historical method that takes into consideration a wide range of cultural and societal factors in their ever-specific interrelatedness in real markets needs a hermeneutical–intuitive approach more than the counting of numbers, and is able to consider the widely indeterminate choice of strategies and results. In

bilateral monopolies, abstract reasoning is of little help to determine prices and equilibria.

The strategies in oligopolistic and bilateral monopoly situations can be broken down into the strategies of mutual quantity adaptations or the strategy of a monopolist and the other market side as quantity adapting, or one side is fixing options (setting price and quantity) and the other side is accepting options. Only for the extreme and somewhat unreal cases of mutual quantity adaptation and clear-cut option fixing does an equilibrium exist. Besides the non-determinate and instability arguments, the social question arises because in Schmoller's time the entrepreneurs set the prices and the workers determined their supply. Owing to the oversupply of the workforce, with no legal restrictions (a 15-hour working day), no insurance against unemployment (high transaction urgency), no public education (knowledge as a power source) and so on, the capitalist camp could more or less play the role of option fixing, that is realizing the exploitation point. As Schmoller argued, against Lujo Brentano and others, even in the case of a formal freedom of coalitions (trade unions were disadvantaged in their ability to organize labour), the capitalists were still able to realize the exploitation point. Even the more recent literature (Scherer, 1979, ch. 8; Tirole, 1988, chs 5 and 6) concludes that, in reality, in bilateral or oligopolisic market situations an unstable equilibrium or a situation without equilibrium will result, which led Stackelberg (1992) in the 1930s to the conclusion that only a strong quasi-dictatorial state (realized in Germany after 1933) could bring the anarchy of factional disputes to an end. Schmoller supported a more democratic solution: the state should set minimum rules (working hours legislation, education for all and so on) and fostered institutions like works councils (*Betriebsräte*) to bring both market sides to the round table. Instead of anonymous market processes, consensual and negotiational solutions should be sought. The logic of the market should be complemented by a sociologic of institutional compromise.

The result for Schmoller was that, to make capitalism function, markets must be regulated. The stability of market processes depends essentially on non-market interventions. In bilateral monopolies (and all continuing transactions tend to this ideal type) law plays a major role in taming the disintegrating forces and in strengthening the weaker side for the common good of productivity, thus enhancing social cooperation and social peace and trust.

> This positive law of contract is combined today with innumerable business customs, with just as many statutes and business arrangements, which include its activities. All this links up with penal and administrative law, the trade, agricultural and building law, the right of domicile, the multitude of regional and local police regulations, the by-laws of decentralized public corporations, and the constitutions of societies. All this together constitutes the dense system of social

> harmonization and standardization of the economic process and of all transactions. It is a system of norms, dams, guiding rules, orders and prohibitions, which regulate the stream of economic life, in defining this or that in agreements as being an offence or without legal force … Immoral contracts, contracts which are considered in general harmful, e.g. today hereditary employment contracts or contracts for a very long period without the right of notice, the pay of factory labour in commodities instead of cash are always forbidden or punishable. It is always the last tendency of this order to allow by customs and law a certain free exertion of the free individual will, but also to protect the common interest and the weaker, to hinder the spread of immoral activities. (Schmoller, 1904, pp. 17–18)

Economic policy along Schmollerian lines in Europe today (tax evasion by multinationals, high unemployment rates, real wage decline and so on) would not consist of decapitating political institutions and governments by surrendering their margin of autonomy in fiscal, monetary and social policy matters to the dictum of a central bank. First, and above all, the relevant policy institutions should at least catch up with the globalization of production and investment on the political plane. The level of the extension of economic activity in Schmoller's time was the national one, so he supported the national state as the relevant political centre to balance powers (collect taxes, institute rules against child labour, strengthen the weak side, introduce measures and weights, and so on). The same process of the enlargement of economic interdependency is occurring now on an international scale and even at the EU level (the developing former socialist countries notwithstanding) the challenge of globalization seen through the eyes of Schmoller hardly seems to have been understood.

References

Backhaus, J.G. (ed.) (1993), *Gustav von Schmoller und die Probleme von heute*, Berlin: Duncker & Humblot.

Balabkins, N.W. (1988), *Not by Theory Alone*, Berlin: Duncker & Humblot.

Barkai, H. (1991), 'Schmoller on money and the monetary dimension of economics', *History of Political Economy*, **23**, 13–39.

Bock, M., H. Homann and P. Schiera (eds) (1989), *Gustav Schmoller Heute*, Berlin: Duncker & Humblot.

Commons, J. (1950), *The Economics of Collective Action*, New York: Macmillan.

Dilthey, W. (1883), *Einleitung in die Geisteswissenschaften*, Leipzig: Duncker & Humblot.

Fellner, W. (1949), *Competition Among the Few*, New York: Knopf.

Fischer, W. (1968), 'Schmoller, Gustav', in D.L. Sills (ed.), *International Encyclopedia of the Social Sciences*, Vol. 14, New York: Free Press, pp. 60–63.

Giouras, A. (1994), *Historismus und Staat*, Frankfurt: Haag & Herchen.

Kaufhold, K.H. (1988), 'Gustav von Schmoller (1838–1917) als Historiker, Wirtschafts- und Sozialpolitiker und Nationalökonom', *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte*, **75**, 217–52.

Lindenlaub, D. (1967), *Richtungskämpfe im Verein für Socialpolitik*, Wiesbaden: Steiner.

Menger, C. (1883), *Untersuchungen über die Methode der Socialwissenschaften, und der Politischen Oekonomie insbesondere*, Leipzig: Duncker & Humblot.

Menger, C. (1884), *Die Irrthümer des Historismus in der deutschen Nationalökonomie*, Vienna: Alfred Holder.

Peukert, H. (1997), *Das Handlungsparadigma in der Nationalökonomie*, Marburg: Metropolis.
Priddat, B.P. (1995), *Die andere Ökonomie*, Marburg: Metropolis.
Scherer, F. (1979), *Industrial Market Structure and Economic Performance*, Boston, MA: Houghton Mifflin.
Schiera, P. and F. Tenbruck (eds) (1989), *Gustav Schmoller in seiner Zeit*, Berlin: Duncker & Humblot.
Schmoller, G. (1860), 'Zur Geschichte der national-ökonomischen Ansichten in Deutschland während der Reformations-Periode', *Journal of Institutional and Theoretical Economics*, **16**, 461–716.
Schmoller, G. (1864/65), 'Die Arbeiterfrage', *Preußische Jahrbücher*, **14**, 393–424, 523–47; **15**, 32–63.
Schmoller, G. (1870), *Zur Geschichte der deutschen Kleingewerbe im 19. Jahrhundert*, Halle: Buckhandlung der Waisenhaus.
Schmoller, G. (1879), *Die Strassburger Tucher- und Weberzunft*, Strasburg: Trübner.
Schmoller, G. (1883), 'Zur Methodologie der Staats- und Sozial-Wissenschaften', *Schmollers Jahrbuch*, **7**, 239–58.
Schmoller, G. (1900/1904), *Grundriss der Allgemeinen Volkswirtschaftslehre*, 2 vols, Berlin: Duncker & Humblot.
Schmoller, G. (1949), *Die Volkswirtschaft, die Volkswirtschaftslehre und ihre Methode*, Frankfurt: Klostermann, first published 1893 and 1910.
Schumpeter, J.A. (1926), 'Gustav v. Schmoller und die Probleme von heute', *Schmollers Jahrbuch*, **50**, 1–52.
Stackelberg, H. (1992), 'Marktform und Gleichgewicht', in N. Kloten and H. Möller (eds), *Gesammelte wissenschaftliche Abhandlungen H. v. Stackelbergs*, Regensburg: Transfer.
Stigler, G.J. (1966), *The Theory of Price*, New York: Macmillan.
Tirole, J. (1988), *The Theory of Industrial Organization*, Cambridge, MA: MIT Press.
Treitschke, H. (1875), *Der Socialismus und seine Gönner*, Berlin: Reimer.
Williamson, O.E. (1985), *The Economic Institutions of Capitalism*, New York: Free Press.

54 Adam Smith (1723–90)

Helge Peukert

The man, his, life and background

Some critics such as Schumpeter (1954, pp. 185–94) notwithstanding, Adam Smith's major achievement is the early exposition of a new type of social, cultural, economic, legal and political system after the slow but steady erosion of feudalist and agricultural societies and former theoretical concepts such as mercantilism and physiocracy. The *Wealth of Nations* (Smith, 1976) is not primarily a partisan pamphlet. Its subject matter is the understanding of the socioeconomic transformation at his time in the context of a history of civilization (the intensive debate about Smith's work is well documented and accessible in Clark et al., 1966; Skinner and Wilson, 1975; Wilson and Skinner, 1976; Glahe, 1978; Skinner, 1979; Wood, 1984–94; Jones and Skinner, 1992).

Although most of his writings were burnt by his literary executor, the complete edition of his works sheds light on his encompassing intent to develop a theory of a new type of society. This comprises his handling of rhetoric and *belles-lettres* (Smith, 1983, including his essay 'Considerations concerning the first formation of languages'), essays on philosophical subjects (Smith, 1977, including a highly original theory of science, taking astronomy as an example – see the excellent interpretation by Thomson, 1965), lectures on jurisprudence (Smith, 1982 – the original did not survive, but notes by his students did: see the introduction by Meek et al.), his first published book in 1759 about the theory of moral sentiments (Smith, 1984) and his masterpiece about the wealth of nations, first published in 1776, shortly before the American Declaration of Independence (see the survey of the literature in Muller, 1993, pp. 240–61).

Smith was born in 1723 in Kirkaldy in Scotland. His father was comptroller of customs. Adam Smith himself was commissioner of customs and charity for the last years of his life from 1778 (see West, 1969; Campbell and Skinner, 1982). When he was four years old, one of the last witch burnings took place; in the hinterland, the Highland clans were still intact. From 1660, the port of Glasgow became a main transit point and from 1740 the trade of tobacco, for example, intensified (Hollander, 1973, ch. 3). A union was established with its big British neighbour in 1707, and Scotland lost its sovereignty and independence; in exchange, the landed aristocracy was represented in the Houses of Parliament in London and the merchants were

granted a free trade zone. Scottish intellectuals vacillated between admiration and rivalry with the 'superior' British lifestyle (to copy the elegant British prose became an obsession; see Foley, 1976, ch. 1). Their intellectual challenge was in a certain way similar to the countries in transition today: to develop a free, open market and civil society.

To understand Smith, his background in Scottish philosophy (David Hume, Francis Hutcheson and, later, James Steuart and James Lauderdale, see Taylor, 1965), which is very different from the later British liberal and empirical philosophy (like that of John Locke and Jeremy Bentham), should be taken into consideration: cosmopolitanism, Calvinism, the Roman law tradition (Cicero and Seneca), French enlightenment and a teleological instead of an 'evolutionary'–objectivist perspective (Charles Darwin, Karl Marx and Sigmund Freud were still far away) were its unique combination of characteristics. According to this view, man can reasonably instal institutions which are not beyond human common design (but Smith was no meliorist: see Winch, 1978, p. 182); the 'Introduction and plan of the work' at the beginning of *The Wealth of Nations* (1976, Vol. I, pp. 1–4) bears witness to this. The Scots were social moralists, religious sceptics, anti-Benthamite, critical *vis-à-vis* the human condition and somewhat rebellious (most obvious in John Ruskin and Thomas Carlyle). The function of 'theory' is to develop a *histoire raisonnée* including psychology, ethics, law, politics and social philosophy to make sense of history. The Scots wanted to give a broad, balanced overall view from different standpoints; logical rigour, that is reasoning from some basic deductive–formal principles, was not their research ideal: theirs was to collect all relevant facts. Inconsistencies were seen as constituent parts of social life which should be reflected in theory. Smith's understanding of science deviates fundamentally from today's mainstream methodological and substantial self-understanding (the more unorthodox aspects of Smith are stressed in Ginzberg, 1977, his interventionist attitudes in Viner, 1966). Therefore he is mostly put in the antechamber of economics as a science (see, for example, Screpanti and Zamagni, 1995, pp. 54–71; Blaug, 1978, ch. 2).

Actual facts (some hundreds can be found as arguments in *The Wealth of Nations*), historical and institutional relations and individual experiences (compare Smith's descriptions of universities) are the Scots' bread and butter. Although down-to-earth (the material wealth of nations is the cornerstone of his masterpiece), the social division of labour is combined with political, ethical and legal considerations:

> They are typically curious about people, men at work, about comparative institutions … They are not concerned with logical processes or sequences, or the framing of abstract hypotheses and their analysis to their utmost limits. They wish to build a truly balanced picture of social life as they found it and the forces which controlled it. (Macfie, 1968, p. 29)

With this sceptical, realist, but optimist social science perspective of the Scottish enlightenment, Smith became a student in Glasgow during 1737–40. He went to Oxford from 1740 to 1746. From 1748 to 1751, he lectured in Edinburgh on *belles-lettres* and jurisprudence, and in 1751 he became professor in logic and moral philosophy in Glasgow, a position he held until 1763. From 1764 to 1766, he was tutor to the son of the Duke of Buccleuch. Their European travel led them to London, Toulouse, Geneva and Paris, where Smith may have met Montesquieu, Voltaire and Rousseau. His lifelong pension from the Duke made him financially independent. At the age of 40, he left university and, after staying in Kirkaldy and London, 13 years after his Central European travel, he published *The Wealth of Nations*. Smith (who never married) was personally a sceptical, detached and emotionally uncommitted personality in everyday life, but his theoretical contributions allow us to feel his bright and objective heart, his zeal for exact knowledge and fair judgement and his optimistic, but sceptical, realism.

Moral sentiments and jurisprudence

The first published book of Smith, which appeared in 1752, when he then held a chair in philosophy, was *The Theory of Moral Sentiments* (hereafter *TMS*, Smith, 1984). The underlying theme is the relationship between commercialization and morals, the old question of whether complex societies can develop sound morals (Habermas, 1986). On the surface, Smith discusses different ethical theories and tries to reconcile them: the theory of propriety (Aristotle, Plato, Zeno), the theory of prudent private individualism in the tradition of Epicurus and the theories of the inborn inclination of the happiness of others (Hutcheson, Shaftesbury). The result is much more diverse than is usually presented in today's textbooks. Smith's contemporaries, such as Hume (Smith, 1977, pp. 33 and 43) and Edmund Burke ('you are in some few places … rather a little too diffuse', ibid., p. 47) were well aware of the tensions in the book.

Smith criticizes the polar approaches of Thomas Hobbes and Bernard Mandeville (an egoistical–rationalistic and utilitarian individualism) and the authors of the 'happiness of others' approach, although he shares the starting point of their reasoning. Hutcheson remarked in 1785: 'the truest, most constant, and lively pleasure, the happiest enjoyment of life consists in kind affections to our fellow-creatures, gratitude and love to the Deity, submission to his will, and trust in his providence, with a course of suitable actions. This is the true good to our power … other pleasures seem almost to vanish when separated from them' (Hutcheson, 1989, pp. 64–5). It is not surprising that some interpreters saw a contradiction (the so-called 'A. Smith problem') between Smith's truck-and-barter man in *The Wealth of Nations* (1976, Vol. I, pp. 17–18) and this emotionally driven social man, deriving moral rules

inductively and by the sentiment of identification. But the fact that Immanuel Kant's rational and empty golden rule was influenced by Smith may serve as a first warning: the *TMS*, although having the same starting point as in Hutcheson, became something very different in the course of exposition.

Smith's analysis starts with his often recognized core statement: 'How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it' (Smith, 1984, p. 9). He talks about benevolence in the sense of Arthur Schopenhauer, 'a fellow-feeling with suffering' (ibid., p. 43) or more generally: humans have sympathy for each other in the sense that they each take the position of the other as a basis of moral conduct and see their own actions as an uncommitted impartial spectator. This is the 'mirror-glass-self' idea, which has been developed in symbolic interactionism in the tradition of William J. Cooley, Charles Horton Thomas and G.H. Mead. This indicates that Smith's book deals more with a general theory of socialization than with ethics and morals only (compare especially his statements in Smith, 1984, pp. 24–5 and 110–12), based on 'the faculty of speech, the characteristical faculty of human nature' (ibid., p. 336). Smith does not doubt the empathy concept, but as a sceptical Scottish intellectual he questions and discusses the dominant motive force of individual action. The main spear-head of his discussion is to make the benevolent assumption relative and to introduce other inclinations and driving forces to arrive at a more general, less pleasant but more realist picture of human behaviour.

He first generalizes the sympathy concept, stating that it can be combined 'with any passion whatever' (ibid., p. 10) to dissociate the capability of sympathy/taking the role of the other and benevolence. For example, torturers often seem to feel pleasure through sympathy in this generalized sense. Second, Smith frequently includes the motives of action (not their results), but they may be negative in the case of antisocial motives. Third, in the case of benevolent sympathy, a strong fellow-feeling may be counterfactual and exaggerated; this holds in the case of pity with dead persons (ibid., p. 13). Fourth, he observes that sympathy for oneself and for others has a significantly different intensity: for others it is much less and has the tendency 'to fall short' (ibid., p. 21). Fifth, people may be too affectionate and then they may be 'unfit for the world' (ibid., 1984, p. 40). Sixth, sympathy/empathy can be (mis)used in favour of strategic action to manipulate others. Seventh, if nobody watches us and we can pass over the demonstration of sympathy without recognition, we may feel no sorrow (ibid., p. 44). Eighth, ranking, and the disposition to admire, perverts and misdirects sympathy because we feel sorrow for the rich, but we evade seeing the sorrow of the poor (ibid., p. 51), which is a universal cause of moral corruption, flattery and falsehood

(ibid., p. 61). But Smith does not criticize this in his realist mood. He states that this ethically dubious habit of selective neglect is 'necessary both to establish and maintain the distinction of ranks and the order of society' (ibid.). Finally, Smith's book turns out to be less a theory of moral sentiments but a critical exposition of the complexity of human behaviour, much cooler than Hutcheson's benevolence concept, sometimes nearer to the concept of an 'indifferent spectator' (ibid, pp. 157–8).

No definite answer is given to the three main questions: why does moral behaviour evolve when I take the position of the other; is there an objective standard of motives in the sense of merit/demerit (authority, customs, fashion, reason *à la* Kant, and God – all these possibilities are taken into consideration); and how does the impartial spectator gain independence from the judgements of other people (Smith talks about the great guardian, the mind's eye and an independent inner voice in an unsystematic way)? In the tradition of the Scottish enlightenment, the raising of the question and throwing into relief its complexity is more important than definite answers. But in search of definite results we may make two assertions. First, according to Smith, human beings have a general sense of empathy/sympathy, to evaluate their own behaviour and the actions of others in a relatively detached sense as a basic requirement of social trust and cooperation. In larger societies where value generalization takes place, this is a prerequisite of the functioning of a market and civil society (often lacking in countries in transition like those of Eastern Europe; see Putnam, 1993; UNDP, 1996, chs 2 and 5). Second, human agents have positive and negative behavioural dispositions, some benevolent, some not.

At this point, Smith's structural discussion sets in and law becomes important. In larger and more complex societies like the emerging British capitalist system, which transcend face-to-face and kinship relationships which have been regulated by 'love and affection, esteem, friendship, gratitude' (like the Highland clans in Scotland), social relations are now 'among different men, as among different merchants, from a sense of its utility, without any mutual love or affection … Society cannot subsist unless the laws of justice are tolerably observed … the enforcement of the laws of justice by the punishment of those who violated them' (Smith, 1984, pp. 86–7). These laws of justice are beyond doubt and subjective interpretation. Smith mentions the case of the repayment of a loan, where the rule of law is 'precise, accurate, and indispensable', whereas the rules 'of charity, of generosity, of gratitude, of friendship, are in many respects loose and inaccurate' (ibid., pp. 174–5).

Here is the bridge between his non-contradictory theory of the human condition, morals, civil development, law and the prerequisites of a market economy. Smith gave his lectures on jurisprudence in 1762–63; they were discovered and first published by Cannan in 1896 (see Pesciarelli, 1986). The

subject is much broader than the title might suggest: 'Jurisprudence is the theory of the rules by which civil governments ought to be directed' (Smith, 1982, p. 5). Politics is not reduced to the legal framing of efficient market allocation: 'The first and chief design of every system of government is to maintain justice … To prevent the members of a society from incroaching on one another's property, or seizing what is not their own' (ibid.). Therefore police, taxes and (national) defence are necessary to secure citizens from injury (body, reputation or estate). He differentiates between property as an exclusive right and 'the giving up some part of the full right of property', such as a public road that cuts up a farm (ibid., p. 10), an attenuation of property rights he considers necessary in more complex societies. His definition of property as 'to be considered as an exclusive right by which we can hinder any other person from using in any shape what we possess' (ibid.) and his examples such as patents resemble to a high degree the institutionalist approach (Commons, 1924) and the social law school in Germany (Schmoller, 1900/1904; Stammler, 1921).

Natural rights are considered as 'evident to reason, without any explanation', for example 'that a man has received an injury when he is wounded or hurt any way' (Smith, 1982, p. 13). In contradistinction to natural rights philosophers like Locke, and typically of the Scottish approach, he mentions a very interesting exception: property! This is developed by the sociohistorical ways of occupation, tradition, accession, prescription and succession. Smith describes it here without any tendency to legitimization as natural and justly acquired:

> The only case where the origin of natural rights is not altogether plain, is in that of property. It does not at first appear evident that, e.g. any thing which may suit another as well or perhaps better than it does me, should belong to me exclusively of all others barely because I have got it into my power; as for instance, that an apple, which no doubt may be as agreable [sic] and as usefull to an other as it is to me, should be altogether appropriated to me and all others excluded from it merely because I had pulled it out of the tree. (Ibid.)

The long discussion of the apple acquisition aims at debunking all imaginable 'natural explanations' regarding to whom the apple belongs. The arguments in favour of property are due to Scottish pragmatism: property is good because the unnecessary trouble of dividing the product can be avoided, it cuts off a number of disputes.

Almost like the Historical school (for the early reception of Smith on the continent, see Palyi, 1966), Smith differentiates four stages in history: the ages of hunters, of shepherds, of agriculture and of commerce (Smith, 1982, p. 14; see Stein, 1979). Different densities of population lead to different systems of production and laws: Smith argues in an almost Marxian determinist way. Different systems require different laws and definitions of property.

For Smith, the historical trend is not a weakening of government in the process of civilization, but the other way round: 'The progress of government and the punishment of crimes is always much the same with that of society, or at least is greatly dependent on it. In the first stages of society, when government is weak, no crimes are punished; the society has not sufficient strength to embolden it to intermeddle greatly in the affairs of individuals' (Smith, 1982, p. 129). The more advanced societies are, the greater is the necessary density of regulation, according to Smith: 'The more improved any society is and the greater length the severall means of supporting the inhabitants are carried, the greater will be the number of their laws and regulations necessary to maintain justice, and prevent infringements of the right of property' (ibid., p. 16). Property therefore solves problems but it also provokes new ones, so that the legal extension of private property becomes a relative and incremental thing: 'Property is then after reaching stages with more private property, introduced, and many disputes on that head must inevitably occur' (ibid., p. 203). One consequence is the polarization of riches and the weakening of democratic decision processes: 'The rich men who have large possessions … would as I observed have many dependents who would follow their council [sic] and direction, and in this manner they would have the greatest influence over the people' (ibid.). From this perspective, the result is that the rich protect themselves from attacks by the poor, who obey them because they accept their authority and interest. Smith explicitly criticizes social contract theories. Citizens more resemble sleeping people who have been brought on ship, awakening out at sea when it is too late to decide to leave. All this does not sound like 'natural state of liberty' arguments as Smith has very often been interpreted – or did his approach undergo fundamental changes in *The Wealth of Nations*?

The Wealth of Nations and the functions of law

'The *annual labour* of *every nation* is the fund which *originally supplies* it with all the *necessaries and conveniencies of life* which it annually consumes' (Smith, 1976, Vol. I, p. 1, emphasis added). (In the following, *The Wealth of Nations* can be dealt with only very briefly and in relation to law, but see the literature mentioned at the beginning of this entry.) The per capita increase of goods as a measure of welfare in the process of civilization and in the compounds of national boundaries, customs and culture (see ibid., p. 82, as an example of Smith's implicit critique of methodological individualism) is a secular trend in many countries which needs an explanation and is the normative yardstick for Smith. The enhancement of consumer sovereignty (cheap and diversified goods) as the new major goal of economic activity can be diverted by different economic policies such as mercantilism (see Book IV of *The Wealth of Nations*) or physiocracy.

The increase in income and consumption is due to the objective and spontaneous process of the productivity-enhancing division of labour (Smith takes over the Encyclopedists' example of the pin factory) which itself leads to the introduction of machines. This unplanned, objective process is nevertheless also described in terms of a productive forces concept *à la* Friedrich List, concentrating on 'the skill, dexterity, and judgment with which its labour is generally applied' (Smith, 1976, Vol. I, p. 1).

The following chapters are an interplay between objective–structural preconditions of the division of labour (the extension of the market, the transport system and so on) on the one hand, and human faculties (thinking and language as prerequisites to making contracts and exchange goods) and institutions (such as the introduction of money) on the other, enriched by many historical and empirical illustrations.

Smith's discussion of the difference between the nominal and real price caused much debate (adding-up versus a labour or a marginalist theory of value: see, for example, the contributions in Skinner and Wilson, 1975, pt II) and Smith in fact sometimes confuses, for instance, the questions of real versus nominal terms with the existence of value as such. Seen in the Scottish tradition and from a social law perspective, he may be interpreted in pragmatic terms. Instead of describing objective economic laws, Smith develops a conjectural history of the way cooperating agents might develop adequate social rules of exchange, based on his basic assumption that only labour produces value. So in less-developed societies the necessary quantity of labour is a rough practical measure of the value of a commodity; it 'seems to be the only circumstance which can afford any rule for exchanging' (Smith, 1976, Vol. I, p. 53).

If capital is accumulated in the hands of a few, it will only be lent out for investment if a profit were to be gained for the risk. There are no productivity of capital arguments involved here. Profit is distinct from wages because it has nothing to do with 'quantity, the hardship, or the ingenuity of this supposed labour of inspection and direction' (ibid., p. 54). Profit as a non-reducible income category follows different rules. It is measured as a percentage of the capital invested. The concrete percentage depends on the general state of the economic level of production and on the specific phase of the economy in its cycle (ascending or descending). No precise rules can be formulated about the component parts of prices in the long run (contra David Ricardo). They depend, *inter alia*, on an expected level fixed by society in the past. The long-run price of labour naturally tends to a culturally defined subsistence level, but 'it is in the progressive state, while the society is advancing to the further acquisition, rather than when it has acquired its full complement of riches, that the condition of the labouring poor, of the great body of the people, seems to be the happiest and the most comfortable' (ibid., p. 91). Such a life cycle of nations takes about 200 years.

Smith has an expansion of aggregates in mind: higher employment leads to higher effective demand, which creates a larger market; new investment opportunities occur and a new division of labour; new profit expectations and the savings will go up, the loanable fund increases and higher employment will be the result in the process of a continuous upward spiral (international trade is seen in the same perspective: see Myint, 1977).

Like capital and profit, the emergence of rent depends on general social circumstances: 'As soon as the land of any country has all become private property, the landlords, like all other men, love to reap where they never sowed, and demand a rent even for its natural produce' (Smith, 1976, Vol. I, p. 56). Later on, Smith mentions rent as a remuneration of the productivity of land. As in the case of profit, Smith leaves open whether profit and rent are legitimate remunerations or only a deduction of the value-producing labour. This ambiguity has been considered one of Smith's major shortcomings. Seen in the light of the Scottish tradition, Smith only describes interpretative maps of the different social classes *vis-à-vis* themselves and each other: 'equal quantities of labour are always of equal value *to the labourer*, yet *to the person who employs him they appear* sometimes to be of greater and sometimes of smaller value' (ibid., p. 37, emphasis added).

On a short-run micro level, Smith describes in Chapter 7 the market process in a dynamic, innovative Schumpeterian or Hayekian (importance of locally bounded knowledge) and active price maker perspective (contra, Mirowski, 1991, pp. 163–71, who interprets Smith as a Newtonian at the Cartesian crossroads). If the expected demand for a produced commodity at the natural price falls short (or the other way round), price and quantity adaptations take place which are described in a less elegant and much more deterministic way in modern textbooks. In the world of Smith as a classical economist, the profit motive was an intent, an unnecessary performance; the behaviour is adaptive and often gradient-climbing instead of choice-optimizing; the agents are capable of learning and, well adapted locally, their knowledge is not assumed to be unbounded; institutions are not problematic *per se* (why do firms exist?), but are regarded as essential in guiding behaviour and in making the behaviour of others predictable. The equilibrium concept is centred on constancy (point attactors), but does not necessarily lead to the mutual consistency of plans (Leijonhufvud, 1997). The free play of market forces may lead to an efficient allocation (or not, depending on the circumstances, such as the organization of cartels and so on).

Smith sometimes states that 'the obvious and simple system of natural liberty establishes itself of its own accord' (Smith, 1976, Vol. II, p. 208), which contradicts his much more sophisticated approach in his lectures on jurisprudence. There he stated that in modern societies a junction exists

between private property and legal restrictions and that to hold property is not a natural right and one person's rights are other people's obligations.

> But the policy of Europe, by not leaving things at perfect liberty, occasions other inequalities of much greater importance. … and thirdly, by obstructing the free circulation of labour and stock, both from employment to employment and from place to place … The property which every man has in his own labour, as it is the original foundation of all other property, so it is the most sacred and inviolable. (Smith, 1976, Vol. I, pp. 132 and 136)

Unfortunately, Smith never discusses this difference in perspective between the lectures and *The Wealth of Nations*. The difference may be due to his intention to criticize mercantilist interventionism and he may have set the legal framework *ceteris paribus* for his economic exposition. Nevertheless, in many places the necessity of legal regulations to make markets work shows up. In Chapter 7, for example, he discusses the structural asymmetry of capital and labour: 'It is not, however, difficult to foresee which of the two parties must, upon all ordinary occasions, have the advantage in the dispute, and force the other into a compliance with their terms' (ibid., p. 74). Another case for legal intervention may be seen in the constant endeavour to collude and build up cartels: 'We rarely hear, it has been said, of the combinations of masters … But whoever imagines, upon this account, that masters rarely combine, is as ignorant, of the world as of the subject' (ibid.). Other instances of legal regulation may be seen in the tendency to overwork ourselves or to be overworked by the entrepreneurs (ibid., p. 92) or the tendency to conceal the long-term price and to charge an abnormally high price (ibid., p. 67). Although Smith observes a multitude of occasions for legal interventions (for his public economics, see West, 1990, ch. 7) he does not discuss them explicitly, except in some passages about the 'Expence of justice' (Smith, 1976, Vol. II, pp. 231ff.).

All in all, the juxtaposition of *The Wealth of Nations* and *Lectures on Jurisprudence* highlights two opposite but nevertheless correct insights. On the one hand, markets are instituted processes which need regulation and socially compromising laws; on the other hand, too many regulations strangle them. Arguing against mercantilism, Smith bends the bow in one direction in *The Wealth of Nations*. Including his *Lectures on Jurisprudence*, we get a more well-rounded picture of Smith's point of view, which cannot be reduced to a simple market apology.

References

Blaug, M. (1978), *Economic Theory in Retrospect*, Cambridge: Cambridge University Press.
Campbell, R.H. and A.S. Skinner (1982), *Adam Smith*, London: Croom Helm.
Clark, J.M. (ed.) (1966), *Adam Smith*, New York: Kelley.
Commons, J.R. (1924), *Legal Foundations of Capitalism*, New York: Macmillan.

Foley, V. (1976), *The Social Physics of Adam Smith*, West Lafayette, IN: Purdue University Press.
Ginzberg, E. (1977), *The House of Adam Smith Revisited*, Philadelphia, PA: Octagon Books.
Glahe, F.R. (1978), *Adam Smith and the Wealth of Nations*, Boulder, CO: Westview Press.
Habermas, J. (1986), *Theorie und Praxis*, Frankfurt: Suhrkamp.
Hollander, S. (1973), *The Economics of Adam Smith*, Toronto: University of Toronto Press.
Hutcheson, F. (1989), *Thoughts on Laughter*, Bristol: Thoemmes.
Jones, P. and A.S. Skinner (1992), *Adam Smith Reviewed*, Ames: Iowa State University Press.
Leijonhuvud, B. (1997), 'Classical and modern economics', lecture presented at the first annual meeting of the European Society of Economic Theory in Marseille.
Macfie, A.L. (1968), *The Individual in Society*, London: Allen & Unwin.
Mirowski, P. (1991), *More Heat than Light*, Cambridge: Cambridge University Press.
Muller, J.Z. (1993), *Adam Smith in His Time and Ours*, New York: Free Press.
Myint, H. (1977), 'Adam Smith's theory of international trade in the perspective of economic development', *Economica*, **44**, 231–48.
Palyi, M. (1966), 'The introduction of Adam Smith on the continent', in J.M. Clark (ed.), *Adam Smith*, New York: Kelley.
Pesciarelli, E. (1986), 'On Adam Smith's lectures on jurisprudence', *Scottish Journal of Political Economy*, **33**, 74–85.
Putnam, R.D. (1993), *Making Democracy Work*, Princeton, NJ: Princeton University Press.
Schmoller, G. (1900/1904), *Grundriß der Allgemeinen Volkswirtschaftslehre*, 2 vols, Berlin: Duncker & Humblot.
Schumpeter, J.A. (1954), *History of Economic Analysis*, London: Allen & Unwin.
Screpanti, E. and S. Zamagni (1995), *An Outline of the History of Economic Thought*, Oxford: Clarendon Press.
Skinner, A.S. (1979), *A System of Social Science*, New York: Clarendon.
Skinner, A.S. and T. Wilson (eds) (1975), *Essays on Adam Smith*, Oxford: Clarendon Press.
Smith, A. (1976), *The Wealth of Nations*, Chicago: University of Chicago Press.
Smith, A. (1977), *Correspondence*, Indianapolis: Liberty Fund.
Smith, A. (1982), *Lectures on Jurisprudence*, Indianapolis: Liberty Fund.
Smith, A. (1983), *Lectures on Rhetoric and Belles Lettres*, Indianapolis: Liberty Fund.
Smith, A. (1984), *The Theory of Moral Sentiments*, Indianapolis: Liberty Fund.
Stammler, R. (1921), *Wirtschaft und Recht nach der materialistischen Geschichtsauffassung*, Berlin: de Gruyter.
Stein, P. (1979), 'Adam Smith's jurisprudence: Between morality and economics', *Cornell Law Review*, **64**, 621–38.
Taylor, W.L. (1965), *Francis Hutcheson and David Hume as Predecessors of Adam Smith*, Durham, NC: Duke University Press.
Thomson, H.F. (1965), 'Adam Smith's philosophy of science', *Quarterly Journal of Economics*, **74**, 212–33.
United Nations Development Programme (UNDP) (1996), *Latvia: Human Development Report*, Riga.
Viner, J. (1966), 'Adam Smith and laissez faire', in J.M. Clark (ed.), *Adam Smith*, New York: Kelley.
West, E.G. (1969), *Adam Smith*, New York.
West, E.G. (1990), *Adam Smith and Modern Economics*, Aldershot, UK and Brookfield, US: Edward Elgar.
Wilson, T. and A.S. Skinner (eds) (1976), *The Market and the State*, Oxford: Oxford University Press.
Winch, D. (1978), *Adam Smith's Politics*, Cambridge: Cambridge University Press.
Wood, J.C. (ed.) (1984–94), *Adam Smith: Critical Assessments*, 7 vols, London: Routledge.

55 Werner Sombart (1863–1941)

Günther Chaloupek

The German Historical school of which Werner Sombart was a member had developed in the nineteenth century as a school of economic thought alternative to the classical school, which it criticized for its abstract theoretical approach. Sometimes recognizing and sometimes questioning the scientific legitimacy of deductive economic laws, the Historical school put its main emphasis on the changeable and changing conditions which constitute the reality in which economic laws operate. Its research efforts were therefore devoted, to a large extent, to institutions and, especially, to the evolution of institutions in time.

Werner Sombart, who is generally considered the leading member of the third and last generation of the Historical school – together with the sociologist Max Weber – has produced the most comprehensive synthesis of the enormous research work of the Historical school with his book *Der moderne Kapitalismus* which was completed by the third volume in 1927. An earlier version had been published in 1902, but the critical reception it had encountered, especially from Sombart's teacher, Gustav von Schmoller, had convinced Sombart of the necessity of both a more detailed empirical foundation and a more thorough analysis of the subject.

Born in 1863 in Ermsleben (Prussia), Sombart studied political economy at the universities of Berlin, Pisa and Rome. His doctoral dissertation on tenancy and labour relations in the Roman *campagna* was completed under the supervision of Gustav Schmoller and published in 1888. The main focus of this book is on the causes of the unsatisfactory level of economic development of this part of Italy, which Sombart considered typical of numerous other areas of southern Europe. His conclusion was that aristocratic land ownership in combination with tenant farming which used a growing part of the land in units of expanding size for sheep farming constituted a poverty trap for the sizeable rural population. The author's reform proposals were in the spirit of the famous German 'Verein für Socialpolitik', of which Sombart's father had been a co-founder: the government should initiate a land reform through which the shepherds should be turned into small farm holders who would devote themselves to a more intensive cultivation of the fertile soil.

In 1890, Sombart was awarded an extraordinary professorship at Breslau University. In the following years he took a strong interest in the social movement and in social policy. *Sozialismus und soziale Bewegung* became

his most famous book. First published in 1896, its wide circulation (nine editions by 1923) earned him true popularity among the intellectual public. If Sombart had been sympathetic towards the Social Democratic Party in his early years, he turned away from socialism around the turn of the century and became its fierce critic later, as his book, *Der proletarische Sozialismus* (1925) testifies. Political authorities, though, viewed him with suspicion for a considerable time, so that he was given an ordinary professorship at the University of Berlin only in 1917, from which he retired in 1933. In his later years, Sombart tried to exert some influence on the National Socialist movement through his *Deutscher Sozialismus* (1934) which was his own version of a genuinely German socialism, but the book was completely rejected by the party, mainly because of the author's negative attitude towards modern technology. The circulation of Sombart's last book, *Vom Menschen* (1938) was restricted by the government. Sombart died in political and social isolation in 1941.

Sombart's scientific reputation rests mainly on his magnum opus, *Der moderne Kapitalismus*. The book is intended to be a paradigmatic representation of 'theoretical historicism', that is, Sombart's ideal of a synthesis between historical empiricism and theoretical economics. It gives a typology of the stages of capitalism as an economic system and of its evolutionary development since the Middle Ages: '*Frühkapitalismus*' – early capitalism, '*Hochkapitalismus*' as the period from the Industrial Revolution to 1914, and '*Spätkapitalismus*' ('late' capitalism) as the period during which capitalism is gradually being transformed into a non-capitalist system. Sombart envisaged some kind of mixed economy in which private ownership of businesses would still dominate, but where government would control the economy through a system of overall and indicative economic planning. Among Sombart's scientific achievements mention must also be made of his *Die drei Nationalökonomien* (1930), an important methodological work.

According to his own self-understanding, Werner Sombart's greatest scientific–theoretical achievement was the introduction of the concept of 'economic system' ('*Wirtschaftssystem*'). Economy for Sombart meant provision of livelihood ('*Unterhaltsfürsorge*') for which many different settings are theoretically possible and practically feasible. The basic elements of an economic system are its spirit ('*Geist*'), its pattern model or 'form' ('*Form*') and its technology. Any type of 'economic order' can be meaningfully defined only in the context of a distinct economic system.

In his booklet, *Die Ordnung des Wirtschaftslebens* (1925), Sombart identifies three principal aspects of economic order: regulation, organization and systematization. In the context of the relationship between law and economics, regulation is the most important subject heading under which we have to look for Sombart's contributions. However, we always have to bear in mind that in

Sombart's view the formal aspects, the pattern models of an economic order, cannot be properly understood without placing them in the context of a specific economic system. Thus Sombart rejected any purely formal classification of an economic order. He did not give abstract definitions of the basic elements of the economic system as such but referred to rather concrete categories which he listed under these elements. In this sense, for example, profit orientation versus provision for needs ('*Bedarfsdeckung*') and solidarism versus individualism are categories of 'spirit', freedom versus constraint or private versus social economy are categories of 'form', and empirical versus scientific or stationary versus revolutionary are categories of technology.

Regulation is the system of rules which govern economic behaviour of the actors. Sombart distinguishes between the legal order, the conventional order and the ethical order. A regulatory order derives its orientation from a certain spirit. 'Insofar as this spirit is incorporated into the economic order by the law-giving power, we can speak of a specific system of economic policy (*wirtschaftspolitisches System*)' (Sombart, 1925, p. 3). Parallel to the evolution of the economic system, the system of economic policy has evolved from the pre-capitalist economic policy of medieval cities to mercantilism, that is, the economic policy of early capitalism, and to liberalism, which is the economic policy corresponding to 'high capitalism'.

In the first volume of *Der moderne Kapitalismus*, Sombart gives a separate and systematic treatment to economic policies that promoted development towards the capitalist system. Essentially, capitalism evolved gradually through a transformation of the production apparatus from the traditional handicraft system organized in guilds into capitalist enterprises operating in a market-regulated exchange economy ('*Verkehrswirtschaft*') with the state playing a decisive role as initiator and promoter of this transformation process. Sombart also describes and evaluates a multitude of legal measures concerning the regulatory framework of the economy intended to promote expansion and growth of the 'commercialized', that is, capitalist sector. In this context, he elaborates in great detail on subjects such as the gradual dilution and abolition of traditional guild orders restricting competition and protecting small handicraft shops, unification of markets at the level of national states, or expanding public demand caused by the centralization of government functions.

Sombart's most original contributions to the study of law and economics are the sections in which he elaborates in detail on the evolution of legal forms of enterprise and of business transactions which have become standard practices of business life, mostly during the nineteenth century, so that we take them as self-evident elements of economic life today without being aware that these forms are products of complex and sometimes difficult evolutionary processes. Sombart (1916/1927, Vol. II/1, pp. 66ff.) distinguishes four different forms of economic organization: apart from the 'organization

of power' under which conventional matters of economic policy are subsumed, Sombart also mentions the 'organization of wealth and property' ('*Vermögensorganisation*'), the 'organization of work' and the 'professional organization'.

The evolution of a capitalist organization of wealth and property is essentially characterized by businesses becoming entities of their own (*Verselbständigung des Geschäfts*) separate from the persons who own and operate them (ibid., pp. 101ff.). Sombart cites evidence for the origins of this process from Italian city states in the fourteenth century (*firma*). The development of a formal accounting system is an important step in the process of *Verselbständigung*, as is the development of trading companies. Joint-stock companies are the final step in the process towards limiting personal liability of the owner to the amount of capital which he has contributed to the share capital of the enterprise which is a legal entity totally separate from its owners. To establish these various legal forms of businesses within the legal orders of states was a complicated process which was not completed before the nineteenth century. By devoting considerable attention to describing and analysing their evolution, Sombart places great emphasis on their availability as a precondition for the real take-off of capitalism at the end of the eighteenth century.

Another example of legal forms constituting a basic precondition of capitalist development is the separation of obligations from personal relationships: what later became known as the 'Berle–Means hypothesis'. As a means of payment without using cash, the bill of exchange dates back to the late Middle Ages (ibid., pp. 520ff.) Originally, mere accommodation bills without an underlying trade transaction were used to facilitate payment, especially at international trade fairs. The innovative breakthrough in the development of the bill of exchange as a generally accepted instrument of payment came from the possibility of endorsing such bills. However, this was not possible as long as the obligatory relationship between the creditor and the debtor was considered a personal one, so that a third person could not enforce his claim without the help of the original creditor. Especially in countries with a strong tradition of Roman law, which considered obligatory relationships as strictly personal, the idea of multiple endorsement encountered considerable difficulties in gaining general and legal acceptance (Sombart, 1911, pp. 66ff.).

These are two examples of aspects of the history of modern capitalism which are barely dealt with in modern literature. Economic histories of Europe may have separate chapters on money and banking but they do not succeed in integrating these aspects properly into their overall view in which usually the 'real sector' dominates the stage. In his writings, Sombart does not devote special attention to the actual and concrete interlocking of the real and of the financial spheres. None the less, his analysis has a unifying point of view: for development towards capitalism, 'objectivization' ('reification' is

the usual English word, for the German '*Versachlichung*' which is difficult to translate in the particular meaning Sombart assigns to it) and 'rationalization' of all kinds of social relationships is a fundamental characteristic. Relationships of all kinds increasingly take on abstract, institutionalized forms which exist independently of men acting within these forms (Sombart, 1916/1927, Vol. II/2, pp. 1076ff.). This 'mechanization of society', as Sombart also calls it, is at the same time a powerful driving force of economic development and also of cultural decay. Sombart's cultural pessimism, which dominated human sciences in Germany before and after the First World War is perhaps the main reason why he made little effort to apply his sophisticated insights into the interdependence of law and economics and into the working of institutions to the practical problems of his times.

If 'institutions form the incentive structure of a society, and the political and economic institutions … are the underlying determinants of economic performance' (North, 1994, p. 359), then differences in the institutional set-up are important for the explanation of success and failure in the economic performance of nations, and also for the conclusions to be drawn for improvement of such performance. Sombart's principal interest was in the great tendencies of capitalist evolution, including the evolution of its institutions in time. His preference was to assemble evidence for facts and causes that were supporting and promoting these tendencies. He thought that, sooner or later, all countries would follow a similar path of development. He was not really interested in comparative analysis, whose main purpose is to find out about the differences between countries.

Sombart has not dealt with the relationship between law and economics in a general and abstract mannner as he did with methodology of economics in *Die drei Nationalökomien*, and as Max Weber has done in *Wirtschaft und Gesellschaft*. If Weber points out that (personal–individual) valuations of goods are particularly inaccessible to the influence of law, because they derive from the consequences of economic actions emanating from primary origins' (Weber, 1980, p. 197), this is an aspect in the relationship between law and economics which is not pursued by Sombart. Sombart's emphasis is always on the production side of the economy. Demand is considered under structural aspects (for example, increasing consumption of luxuries) or as an aggregate (for example, in the context of cyclical fluctuations) but not in the form of individual schemes of preference.

In dealing with questions of legal forms of doing business and of capitalist enterprise, Sombart draws on works of economic history and also on the history of law as his sources:

> It is a precondition of any investigation of the origins of an institution, no matter whether it is an investigation in the history of law or in economic history, to have

> a clear concept of this institution in its ideal-type perfection. … The sociologist usually faces the task of having first to formulate such a concept towards which he can orientate his subsequent investigation, because it is a synthetic concept; whereas the jurist finds this concept in the law in force at the time. (Sombart, 1916/1927, vol. II/1, p. 142)

But the meaning of an institution can be understood and determined only in a socioeconomic context which does not coincide with the logic of juridical argumentation and investigation. Sombart therefore implicitly rejected approaches put forward by some of his contemporaries (R. Stammler and Karl Diehl) which made legal institutions appear as the determinant of economic evolution (Pribram, 1983, p. 224). With respect to the traditional orders for handicraft and small business (*Gewerbeordnung*) Sombart maintains that they were a nuisance for expanding capitalist enterprises but 'no essential obstacle to progress towards capitalism' (Sombart, 1916/1927, vol. II/2, p. 1119). While Sombart did not deny that there is some kind of logic of its own inherent in the development of legal forms, in his work the 'law' of the 'law and economics' paradigm primarily lies in what he calls the 'spirit' of the system.

Bibliography

Backhaus, Jürgen (ed.) (1996), *Werner Sombart (1863–1941) Social Scientist*, 3 vols, Marburg: Metropolis.

Lenger, Friedrich (1994), *Werner Sombart 1863–1941. Eine Biographie*, Munich: C.H. Beck.

North, Douglass (1994), 'Economic performance through time', *American Economic Review*, **3** (84), March, 359–68.

Nussbaum, Frederick Louis (1933), 'A history of economic institutions of modern Europe', *Introduction to Der moderne Kapitalismus of Werner Sombart*, New York: F.S. Crofts & Co.

Pribram, Karl (1983), *A History of Economic Reasoning*, Baltimore, MD and London: Johns Hopkins University Press.

Sombart, Werner (1888), 'Die römische Campagna' ['The Roman Campagna'], in Gustav Schmoller (ed.), *Staats- und sozialwissenschaftliche Studien*, **VIII** (3), Leipzig.

Sombart, Werner (1896), *Sozialismus und soziale Bewegung im neunzehnten Jahrhundert*, Jena: Gustav Fischer; English translation, *Socialism and the Social Movement in the 19th Century*, New York: G.P. Putnam's Sons, 1898.

Sombart, Werner (1911), *Die Juden und das Wirtschaftsleben*, Leipzig: Duncker & Humblot; English translation, *The Jews and Modern Capitalism*, London: T.F Unwin, 1913.

Sombart, Werner (1916/1927), *Der moderne Kapitalismus* [Modern Capitalism], vols I–III, Munich and Leipzig: Duncker & Humblot.

Sombart, Werner (1925), *Die Ordnung des Wirtschaftslebens* [The Order of Economic Life], Berlin: Verlag Julius Springer.

Stammler, Rudolf (1925), 'Recht' ['Law'], in J. Conrad and E. Loening (eds), *Handwörterbuch der Staatswissenschaften*, 4th edn vol. 6, Jena: G. Fischer, pp. 1187–1202.

Weber, Max (1980), *Wirtschaft und Gesellschaft* [Economy and Society], Fünfte revidierte Auflage besorgt von J. Winckelmann, Tübingen : J.C.B. Mohr (Paul Siebeck).

56 Lorenz von Stein (1815–90)

Heinz Grossekettler

A survey of von Stein's life and work

Lorenz von Stein was born on 15 November 1815 in Eckernförde, a small town in the present-day German *Land* of Schleswig-Holstein. He was the illegitimate son of von Wasner, an officer in the Danish army, and the wife of a sergeant by the name of Stein.[1] His father assumed responsibility for his education, and sent him to a school founded by the Danish king for soldiers' sons. As a result of his exceptional talent, Lorenz von Stein was introduced to the King of Denmark who awarded him a scholarship. He won further scholarships and thus he was subsequently able to register as a law and philosophy student at the University of Kiel. He also spent some time at the University of Jena, where he devoted much thought to J.G. Fichte's views on philosophy and economics, as well as his quest for a rational legal and political system. He then returned to Kiel, where he took his final examination in law in 1839.

While serving a period of articled clerkship in Copenhagen, he worked on a dissertation about the history of Danish civil procedure (von Stein, 1841) and in 1840 he was made a doctor of law in Kiel. The King of Denmark thereupon awarded him a travelling scholarship which enabled him to finance a stay in Berlin before going on to Paris. In Berlin, he moved in neo-Hegelian circles and made an intensive study of the political philosophy of G.W.F. Hegel, whose work *Grundlinien der Philosophie des Rechts* ('The Philosophy of Right') had been published in 1821 and had exerted a profound influence on the prevailing theory of public law. In Paris, he studied the doctrines expounded by the socialists and communists and had discussions with Louis Blanc, among other people. The sojourn in France left its mark in two books – one about socialism and communism in France (von Stein, 1842) and the other about local government law in that country (von Stein, 1843). After his return to Kiel, he applied for a post as a senior lecturer and obtained the desired position in 1845. His main lectures dealt with the history of law and theories of the state.

In the meantime, Schleswig-Holstein had become the bone of contention in a dispute between Denmark and the German *Länder*. In the Middle Ages, Schleswig was a Danish fief, while Holstein was German. In 1460, a union was effected under Christian I, who reigned at the same time as a Danish king in Schleswig and as a German duke in Holstein, and who had vowed upon his election to leave the two regions 'eternally undivided'. This kind of thing was

possible as long as there were feudal states, but when nation-states began to evolve disputes were virtually inevitable. Although von Stein had received generous financial aid from the Danish king, he made common cause with other jurists in Kiel who in a juridical report argued on the strength of historical documents that Schleswig-Holstein ought to be considered part of Germany. When, in 1852, owing to the pressure exerted by other European powers, and in spite of the Prussian successes in the first war for Schleswig-Holstein, this region was annexed to Denmark,[2] von Stein was dismissed on account of his contribution to the above-mentioned report.

There followed three years of privation in which von Stein worked as a freelance journalist. In this period, he applied to various universities in the hope of obtaining a professorship. In 1855, he was appointed to a chair of political economy at the University of Vienna. He rose to fame while he lectured there for the next 35 years, until his death on 23 September 1890. He published coursebooks which were held in high repute; they dealt with subjects ranging from statistics to political economy, the science of administration and public finance. In 1868, a hereditary knighthood was conferred on him. In 1878, he became a member of the Viennese Academy of Sciences. In the same year, he was awarded an honorary doctorate by the University of Bologna and was appointed adviser to the Japanese government on questions relating to the framing of a constitution and the setting up of an administration. This was part of the modernization programme the Japanese were then endeavouring to implement. As far as the symbiosis of law and economics is concerned, special mention should be made of von Stein's manuals on the history of the social movement in France (von Stein, 1855), his manual on political economy (von Stein, 1887), his multi-volume work on the science of administration (von Stein, 1884), and his manual on public finance (von Stein, 1885/86). A survey of his work is given in the obituaries referred to in von Beckerath and Kloten (1959, pp. 89ff.). An appraisal made from a present-day point of view can be found in a volume containing a commentary on a facsimile edition of the first edition of von Stein's book on public finance in the series 'Klassiker der Nationalökonomie' (Hax et al., 1998).

How von Stein became a precursor of the law and economics approach

We have to bear in mind three points if we want to understand how von Stein gradually came to embrace ideas which justify our considering him as a precursor of the law and economics school:

1. the law and economics approach has points of contact with the areas of study to which von Stein devoted most of his energies;
2. German public finance is rooted in cameralism and the initial economic conditions which prevailed in cameralistic Germany; and

3. von Stein's thinking was influenced by the way in which the economies of the German-speaking countries evolved during his lifetime.

The *points of contact* between von Stein's ideas and those of the proponents of the law and economics school become apparent if we consider the characteristic features of the school in question (Schäfer and Ott, 1995, p. 10). First, the legal system is viewed not as a datum, but as a variable. Second, the history of the origins of property rights is analysed. Third, the influence of legal structures on the efficiency of economic processes is investigated. Fourth, efforts are made to determine the circumstances in which efficient legal structures would develop. Fifth, the requisite properties of such efficient legal structures are described. Von Stein did not ask such questions about civil law, but he did consider the implications that such questions have for public law and the optimal way to develop public law in order to render it capable of fulfilling meaningful state functions.

As far as the first two features of the law and economics school are concerned, von Stein's approach was strongly influenced by his study of Fichte and Hegel, as well as by his stays abroad. As a result, he considered political and economic history as the outcome of a battle of ideas waged by various pressure groups – a combat that would lead to concrete improvements in the legal systems of states which were at different stages of development. By adopting a comparativist approach to law, he believed, one could therefore ascertain the characteristics of an efficient legal structure towards which history was advancing (Kamp, 1950). This is why, in his manual of public finance, von Stein considered it so important to describe the relevant laws of all the major European states and compare their legal systems in terms of efficiency. In this instance, however, the term 'efficiency' should not be understood as a sort of Pareto optimum interpreted in a static sense. What von Stein had in mind was a kind of evolutional efficiency. He believed that the comparison of various legal systems would make it possible to set up rules for national debt and lay down principles for organizing administrative structures and tax systems.

The *cameralistic origins of German public finance* are important because the cameralists' insights resulted from the observation of economic processes which took place under initial conditions quite different from those prevailing in countries such as France and England. In the wake of the Thirty Years' War, the Holy Roman Empire had broken up into a multiplicity of independent states ruled by absolutist princes who had the right to levy taxes directly. This had consequences which we may explain with the aid of an admittedly somewhat daring comparison. Let us imagine a region where several bands of mafiosi are fighting for predominance in a protection racket. This state of affairs is in some respects comparable to the situation that existed during the

Middle Ages in feudal states where the property rights of the various classes of the nobility had not yet been clearly defined. The war and marriage policies pursued by the higher nobility led to the establishment of absolutist states ruled by princes whose right to levy taxes was undisputed; and the upshot of all this was a state of affairs which presents certain points of resemblance to the situation which arises once the claims of rival mafia bands have been staked out. Those who have acquired the right to levy taxes or extort protection money participate, so to speak, in the commercial success of their subjects or 'protégés'. If they set store by long-term profit maximization they will be careful not to kill the goose that lays the golden egg. In other words, they will be well advised to promote the economic welfare of the people who pay them taxes or protection money.

In the German *Länder*, this now came about in circumstances quite different from those which existed in relatively closed nation-states such as France and England. It was hard to enforce border controls between a great number of small states, and there were virtually no linguistic or geographical barriers to mobility. As a result, factors of production (labour and capital) were able to move freely (*de facto*, though not *de jure*) between the numerous small states, and there was keen competition between the *Länder*, which vied with each other for enterprises and workers (Backhaus and Wagner, 1987). German princes were therefore very interested in advice concerning a practicable site development policy. The upshot of all this was a comprehensive doctrine of political economy. It was this kind of doctrine that was one of the major distinctive features[3] of von Stein's work, and it was this kind of doctrine that he sought to establish by carrying on the cameralistic tradition and combining public finance with the science of administration.

Finally, *nineteenth-century German economic history* was marked by German unification and Germany's transformation into an industrial state. These events unfolded between 1840 and 1890 – a crucially important period in von Stein's career. As a result, they reinforced his ideas about efficiency and his quest for suitable legal structures for the public sector.

Specific contributions: von Stein's influence on the science of administration, the doctrine of a rational tax system and the mapping out of rules for public debt

If we look up 'Stein, Lorenz von' in the first edition of the *Dictionary of Political Economy* (Palgrave, 1899, p. 474), we find an article by J.K. Ingram in which emphasis is placed on the fact that von Stein always stressed the close links between 'philosophy, law and economics' and that 'he and R. von Mohl are regarded as the creators of the modern science of administration'. *Von Stein's treatises on the science of administration* are appraised in similar terms in the relevant German jurisprudential literature of the past 30 years

(Forsthoff, 1972; Schnur, 1978). Owing to their excessive period specificity, however, von Stein's reflections are largely irrelevant to an economically viable science of administration of the type that is reflected at the end of the twentieth century in the nisus towards new public management. None the less, we now propose to deal briefly with his *doctrine of social monarchy* (von Stein, 1855, pp. 9ff.), not only for historical reasons, but also because the doctrine in question probably had political repercussions.

In order to understand the doctrine of social monarchy, we have to bear in mind the fact that, in the wake of the French Revolution and the subsequent uprisings, all the royal houses then reigning in Europe were in mortal fear of further revolutions. If, in such a situation, an economist was aware that with respect to economic development there was a fundamental identity of interest between the ruler and his subjects,[4] the obvious thing was to search for other potential identities of interest which might be exploited in the interest of social and economic development. Von Stein perceived such a possibility in the reconciliation of the antagonisms between the propertied and the non-propertied classes and between the state (the prince) and society at large (the working population). The basic idea is brilliantly simple. In order to avoid class warfare and stabilize his dominion, the king has to offer *both* classes something that *only he* can offer: he has to offer the non-propertied class a minimum amount of social security and at least some career prospects, and he has to offer the propertied class the assurance that they will not be dispossessed as a result of a class struggle. Von Stein used the catch-phrase 'social monarchy' to propagate social measures aimed at facilitating the attainment of such an objective and protecting the vested interests of all those concerned. Similar reflections presumably contributed to a not inconsiderable extent to the fact that in the late nineteenth century, Bismarck effected social reforms which were then universally acclaimed as exemplary and which really did have a stabilizing effect. In the juristic literature, these reflections about social reforms were developed further and took solid shape in the idea that the state had an obligation to provide welfare services. Some jurists believed that this obligation ought to be enshrined in the constitution, and later on the obligation in question actually *was* enshrined in the German constitution. However, under the entirely different conditions prevailing in the twentieth century, this has resulted in a welfarist tendency whose effects are becoming increasingly destabilizing. This trend is no longer an expression of the law *and* economics approach. In actual fact, it is a manifestation of law *without* economics.

Like many nineteenth-century specialists in public finance, von Stein initially considered the state's responsibilities and expenditure as elements of the science of government in the narrower sense. This system of ideas is set against the theory of revenue as the theory of public finance in the narrower

sense. In the first edition of von Stein's manual on public finance (von Stein, 1860), for instance, the disquisition on government expenditure constitutes the first part of the science of government, the discussion of the state budget makes up the second part, and the commentary on public revenue and the appropriate fiscal administration forms the third part of the science of government in the wider sense, as well as the theory of public finance proper. It was only in the fifth edition of his manual on public finance (von Stein, 1885/86) that he abandoned the distinction between the theory of public finance in the narrower sense and the theory of government in the narrower sense. It was no longer possible to draw such a distinction because modern states, unlike the budgets of medieval principalities, are not interested in revenue as such. Their main concern is (or should be) to perform duties which are in the interest of every citizen. What such duties require is not the raising of revenues as such, but rather the raising of revenues for the purpose of performing these duties.

This basic idea, which is implicit in the first edition of the manual on public finance, informs von Stein's *theory of a 'rational tax system'*. In von Stein's view, such a system must fulfil two main conditions. First, it must comprise a large number of taxes which are levied at all the appropriate points in the economic process, and it must be organized in such a way that the capacity for forming real and human capital is preserved. In other words, nobody should be obliged to draw on his capital reserves in order to pay tax. Second, it is important to ensure that a substantial proportion of tax revenue is used 'productively', as von Stein puts it. In other words, the capital in question is to be employed for 'directly productive' (that is profitable) investments or for indirectly productive activities such as the development of the physical infrastructure, health care or the formation of human capital. It is not proposed to go into the minutiae of von Stein's theory of an economically oriented system of tax laws. His reflections revolved around the special difficulties attendant on an economy whose national product was still largely generated by day labourers, on the one hand, and by non-bookkeeping farmers, artisans and small businesses, on the other.[5]

Much of von Stein's taxation theory is now of purely historical interest and can only be recommended to readers whose curiosity is whetted by the prospect of discovering in tax legislation a mirror of nineteenth-century trade and industry. By contrast, von Stein's theory of public debt comprises ideas which even today would be worth pondering. He refined and elaborated these ideas in the various editions of his manual on public finance, and people still quote a maxim which can be found in somewhat different forms in the second to fourth editions of this work. In the third edition, the aphorism in question reads as follows: 'A state with no national debt either does too little for its future or demands too much from its present' (von Stein, 1875, p. 716). In

this apophthegm, the author is referring to intergenerative externalities, and he says he would like them to be internalized by pay-as-you-use financing aided by overseas borrowing. The sentence under discussion is absent from the fifth edition of the manual on public finance, and much more space (the entire fourth volume) is now given over to the theory of public debt, which is treated in a much more subtly differentiated manner. If one wishes to bring von Stein's main ideas into relief, one must bear in mind that from an economic viewpoint it is pointless to consider borrowing activities in isolation since decisions concerning public sector borrowing must always be placed in relationship to three things, namely the budget, the balance of payments and the development of an economy's production potential.[6]

As far as the *budget* is concerned, it is essential to bear in mind that, from a technical viewpoint, the budget, as a whole, is always balanced. If a debt is contracted, an offsetting entry therefore has to be made opposite it. In the present instance, we may distinguish three cases, which are denoted by different terms in the literature:

1. If the contracting of a loan is set against an increase in government expenditure, one speaks of *deficit spending* or the *budgetary effect of public debt*.
2. If, on the other hand, borrowing serves to reduce taxes and the volume of expenditure remains unchanged, one speaks of a *deficit without spending* or a *tax-debt differential effect*.
3. If debt category *i* is replaced by debt category *j* and the volume of expenditure remains unaltered, there is a *debt–debt differential effect*. The means whereby such an effect is produced is described as *debt management*.

In order to bring such budgetary questions into sharper focus, von Stein sets up rules whose main purpose is to produce desirable budgetary effects and ensure proper debt management. When we look at the *balance of payments*, it is important to draw a distinction between domestic and overseas borrowing. In so doing, one should not consider those who subscribe to a loan, but the effects on the balance of payments on current account. If, as in the wake of German reunification, state indebtedness leads to a current account deficit, the home country seeks to encourage present consumption and/or present investment by borrowing goods and factors from other countries, thereby burdening future generations with liabilities to foreign countries. Domestic borrowing *may* also be a burden on future generations if, for instance, capital borrowed by the government squeezes out private investments and is used by the state for purposes of consumption, but this is not always necessarily the case.

When von Stein analyses these problems associated with the balance of payments, the focus of his analysis is on overseas borrowing. This kind of borrowing, he says, is characteristic of the German *Länder*, where there is fierce competition between different locations, and factors can move freely across open borders. In France and England, by contrast, domestic borrowing is the rule, owing to natural barriers to mobility and the way in which loans are traditionally handled by banks or administrative bodies.

Turning now to *the development of the production potential*, we have to draw a distinction between two types of effects: (1) the cyclical effects of the public debt on the capacity utilization rate of a given production potential, and (2) growth effects on the development of this potential. Unlike John Maynard Keynes, von Stein here takes growth effects as his focus of attention. In his analysis of the various types of public debt, an analysis tailored to the situation in German *Länder* in the nineteenth century, von Stein maps out allocation- and distribution-oriented *borrowing rules* designed to ensure that this financing instrument is always used correctly. The maxims in question may be described as reproductivity, pension, burden-sharing and interest cost-minimization rules.

The ultimate aim of the *reproductivity rule* is the best possible utilization of capital throughout the world. Let us suppose there are good grounds for believing that a backward country will yield above-average returns on investment if an appropriate policy is implemented. In such a case, the reproductivity rule requires that the state in question should speed up the development of its infrastructure by borrowing abroad. It is, however, crucially important that the borrowed capital should not be used for purposes of consumption, but solely for purposes of investment. This utilization may be either 'directly productive' (so-called 'profitable investments') or 'indirectly productive' (that is, likely to raise the productivity of private investments by creating better infrastructural conditions). According to this rule, borrowing is only justified if it leads sooner or later to an increase in production potential which boosts tax revenue to such an extent that tax receipts are sufficiently high to cover the cost of servicing the loan. In such circumstances, the debt service can be 'reproduced', so that the debt might, so to speak, be said to finance itself.

The *pension rule* is distributively oriented and emphasizes the state's responsibility for integration and insurance. When managing the national debt, the administration is to take into consideration the fact that citizens want sound investment opportunities, particularly in view of their old age pensions. In order to take account of this need, the government should therefore ensure that the national debt consists partly of liabilities due to domestic borrowing. The national debt should, moreover, be denominated, made negotiable and promoted by price support operations.

The *burden-sharing rule* relates to the internalization of intergenerative externalities. This objective is to be attained by pay-as-you-use financing of

long-term investments to improve the infrastructure by means of overseas borrowing. The second to fourth editions of von Stein's manual on public finance contain an embryonic form of this rule, namely the aforementioned judgement on a state without a national debt.

The *interest cost-minimization rule* aims – as the name suggests – at minimizing the state's interest costs. Von Stein says that the interest rate at which a state can run up debts depends, among other things, on how other countries judge its creditworthiness. He adds that its position in the international rating order is influenced to a very large extent by the regularity of its administration, the terms of an issue and the redemption commitments. The public character of the budget is of the utmost importance, and so are a sensible organization of the fiscal system and the conscious fostering of investors' confidence. In order to reinforce this confidence, von Stein proposes, among other things, the setting up of a *State Loan Commission*. This is not supposed to be responsible for debt management in the modern sense. It is a sort of Council of Economic Experts whose incumbent duty is to decide whether a plan to borrow capital is legitimate in the sense of the reproductivity rule, and whether the state is in a position to service the loan. Even today it would be desirable to set up such a body in order to ensure that state borrowing is limited to a sustainable level. A State Loan Commission could form part of a practicable public sector borrowing system (Funke, 1995).

These are all proposals for the construction of a legal framework for national debt. They seem eminently worthy of consideration at a time when European and even global locational competition has taken on new dimensions because of the sharp decline in transport and transactions costs. At any rate, there is little doubt that competition is now much fiercer than when von Stein mapped out his rules, which were geared to the needs of small open countries competing for capital and labour some 150 years ago.

From what has been said about von Stein's work in general and his manual on public finance in particular, it should be clear that his way of thought can be aptly described as a European law and economics approach. He sought to establish a kind of efficiency-oriented law, though the term 'efficiency', as has already been pointed out, must not be misapprehended as a sort of Pareto-optimum, but viewed rather as a species of evolutional efficiency geared towards the kind of dynamism on which Schumpeter laid primary stress. Von Stein, moreover, was a pioneer in the comparative analysis of legal structures designed to find out optimal solutions to given economic problems.

Notes

1. See Schmidt (1956) and Taschke (1985) for these and other biographical details.
2. Schleswig-Holstein remained part of Denmark until it was reconquered in 1864–65.

3. The other main distinctive feature was von Stein's pursuit of dynamic efficiency.
4. This identity of interest was described by the cameralists and by our comparison with the situation of mafiosi 'ruling' undisputed claims.
5. A good impression of von Stein's general position on taxation is conveyed by an extract from the fifth edition of his manual on public finance. This extract was translated from German by Jacques Kahane and published in Musgrave and Peacock's (1967) *Classics in the Theory of Public Finance*. An in-depth analysis of von Stein's taxation theory can be found in Heilmann (1984).
6. What now follows is a summary of ideas set forth in Grossekettler (1990). The relevant references and illustrative quotations can be found in the same article.

References

Backhaus, R. and R.E. Wagner (1987), 'The cameralists: a public choice perspective', *Public Choice*, **53**, 3–20.

Beckerath, E. von and N. Kloten (1959), 'Lorenz von Stein', in E. v. Beckerath, H. Bente, C. Brinkmann, Gutenberg, Harberler, Jecht, Jöhr, Lütge, Predöhl, Schaeder, Schmidt-Rimpler, Weber and v. Wiese (eds), *Handwörterbuch der Sozialwissenschaften*, Vol. 10, Stuttgart, Tübingen and Göttingen: Fischer, Mohr, Vandenhoeck & Ruprecht.

Forsthoff, E. (ed.) (1972), *Lorenz von Stein: Gesellschaft–Recht–Staat* [*Lorenz von Stein: Society, Right, Government*], Frankfurt am Main: Ullstein.

Funke, St. (1995), *Die Verschuldungsordnung. Ein Beitrag zur finanzwirtschaftlichen Ordnungspolitik* (national debt principles), Berlin: Duncker & Humblot.

Grossekettler, H. (1990), 'Die Staatsverschuldung aus der Sicht Lorenz von Steins' (von Stein's theses concerning the national debt), in H. Scherf (ed.), *Studien zur Entwicklung der Ökonomischen Theorie IX*, Berlin: Duncker & Humblot.

Hax, H., A. Heertje, K.-D. Grüske and B. Schefold (eds) (1998), *Vademecum zu einem Klassiker der Staatswissenschaft. Kommentarband zur Faksimileausgabe von Lorenz von Stein, Lehrbuch der Finanzwissenschaft (1860)* (companion to von Stein's *Manual on Public Finance* from 1860), Düsseldorf: Verlag Wirtschaft und Finanzen.

Hegel, G.W.F. (1821), *Grundlinien der Philosophie des Rechts. Naturrecht und Staatswissenschaft im Grundrisse* (philosophy of right), Berlin: Nicolai.

Heilmann, M. (1984), *Lorenz von Stein und die Grundprobleme der Steuerlehre. Ein Beitrag zur Geschichte der Finanzwissenschaft* (von Stein and the principles of taxation), Heidelberg: Decker.

Ingram, J.K. (1899), 'Stein, Lorenz von', in Palgrave (ed.), p. 474.

Kamp, M.E. (1950), *Die Theorie der Epochen der öffentlichen Wirtschaft bei Lorenz von Stein* (von Stein's stages of political economy), Bonn: Röhrscheid.

Musgrave, R.A. and A.T. Peacock (1967), *Classics in the Theory of Public Finance*, London, Melbourne and Toronto: Macmillan/New York: St Martin's Press.

Palgrave, R.H.I. (1899), *Dictionary of Political Economy*, London: Macmillan.

Schäfer, H.-B. and C. Ott (1995), *Lehrbuch der ökonomischen Analyse des Zivilrechts* (Economic analysis of law), 2nd edn, Berlin: Springer.

Schmidt, W. (1956), *Lorenz von Stein. Ein Beitrag zur Biographie, zur Geschichte Schleswig–Holsteins und zur Geistesgeschichte des 19. Jahrhunderts* [*Lorenz von Stein. Comments on his Biography, the History of Schleswig–Holstein and 19th-century History of Ideas*], Eckernförde: Schwensen.

Schnur, R. (ed.) (1978), *Staat und Gesellschaft. Studien über Lorenz von Stein* [*Government and Society. Studies on Lorenz von Stein*], Berlin: Duncker & Humblot.

Stein, L. von (1841), 'Die Geschichte des dänischen Zivilprozesses und das heutige Verfahren. Als Beitrag zu einer vergleichenden Rechtswissenschaft' (von Stein's thesis about Danish civil procedure), Kiel.

Stein, L. von (1842), *Der Socialismus und Kommunismus des heutigen Frankreichs. Ein Beitrag zur Zeitgeschichte* (socialism and communism in France today), Leipzig: Wigand.

Stein, L. von (1843), *Die Munizipalverfassung Frankreichs* [*Local Government Law in France*], Leipzig: Wigand.

Stein, L. von (1855), *Das Königtum, die Republik und die Souveränität der französischen Gesellschaft seit der Februarrevolution 1848 (Geschichte der Socialen Bewegung in Frankreich von 1789 bis auf unsere Tage, 3. Band*) (kingdom, republic and sovereignty in France), third volume of 'History of the Social Movement in France', 2nd edn, Leipzig: Wigand.

Stein, L. von (1860), *Lehrbuch der Finanzwissenschaft. Als Grundlage für Vorlesungen und Selbststudium* (manual on public finance), Leipzig: Brockhaus.

Stein, L. von (1875), *Lehrbuch der Finanzwissenschaft. Als Grundlage für Vorlesungen und Selbststudium mit Vergleichung der Finanzsysteme und Finanzgesetze von England, Frankreich, Deutschland, Oesterreich und Rußland, 3., vielfach verbesserte und vermehrte Auflage* (manual on public finance), 3rd edn, Leipzig: Brockhaus.

Stein, L. von (1884), *Die Verwaltungslehre* [*Science of Administration*], Stuttgart: Cotta.

Stein, L. von (1885/86), *Lehrbuch der Finanzwissenschaft.* Mit specieller Vergleichung Englands, Frankreichs, Deutschlands, Oesterreichs, Italiens, Rußlands und anderer Länder, 5., neubearbeitete Auflage [*Manual on Public Finance*], 5th edn, Leipzig: Brockhaus.

Stein, L. von (1887), *Lehrbuch der Nationalökonomie* [*Manual on Economics*], 3rd edn, Vienna: Manz.

Taschke, H. (1985), *Lorenz von Steins nachgelassene staatsrechtliche und rechtsphilosophische Vorlesungsmanuskripte. Zugleich ein Beitrag zu seiner Biographie und zu seinem Persönlichkeitsbegriff* (von Stein's notes on lectures on constitutional law and the philosophy of law), Heidelberg: Decker.

57 George Joseph Stigler (1911–92)

Peter R. Senn

George Stigler's 1982 Nobel Prize for Economics was awarded to him for his seminal studies of industrial structure, the functioning of markets and the causes and effects of public regulation. He thought 'much the most' important contribution was his theoretical work on the economics of information (Stigler, 1986, p. 105). Gary S. Becker (b. 1930) (1993, p. 763) agreed but also felt that 'Stigler's main scientific contributions were to the history of economic thought and to microeconomics, with a special emphasis on industrial organization' (ibid., p. 762). Ronald Harry Coase (b. 1910; Nobel Prize, 1991), in his sympathetic memoir, wrote that Stigler was 'seen at his best' in his studies of the history of economic thought (Coase, 1991, p. 472). He was also a pioneer in the development of 'public choice economics'.

In addition to his Nobel prize, Stigler received many other honours. Among them were the National Medal of Science, 1987, and the position of President of the American Economic Association, 1964, President of the Mont Pelerin Society (of which he was a founding member), 1976 and President of the History of Economics Society, 1977. He received eight honorary degrees and was elected to the American Philosophical Society in 1955 and to the National Academy of Sciences in 1975. None of these honours was specifically related to his contributions to the subdiscipline of law and economics.

There is no mention of law and economics in the 1993 issue of the *Journal of Political Economy* (Anon, 1993) dedicated to his memory. His complete bibliography by Vicky M. Longawa (1993) does not have any article titles with the phrase 'law and economics' in it.[1] Indeed, the term 'law' appears in the titles of only three of his articles. This is despite the fact that he had many publications in legal and law and economics journals, listed at the end of this entry, including 'The economies of scale' in the first issue of the *Journal of Law and Economics* (1958).

Most of his publications about the law were on the subjects in the field that gave him one of his several claims to fame, economic concentration and public regulation. Much of what he published outside the legal and law and economics journals was closely related to the field as it now stands. Only one of the many obituaries, that by C.R. McCann, Jr. and Mark Perlman (1993, p. 996) mentions law and economics. They discuss Stigler's contribution under the rubric 'Public policy, political economy and regulation' (ibid., pp. 1003ff.). By way of examples, Peter Passell (1991), in the *New York*

Times, and Robert L. Barro, in the *Wall Street Journal*, have nothing to say of his contributions in law and economics.

Stigler wrote no important book on law and economics, yet his was a decisive influence in shaping the approach and many of the methods of the subdiscipline of law and economics during its early development. His innovative contributions were necessary for the advancement of the field. If Stigler had not made them, someone else would have had to. The way they were made has obscured their significance. To understand this, some history and description is required.

From the most ancient pre-scientific writings on economic subjects it is clear that people knew that laws affected the economy. From its very beginnings as a social science at the hands of Adam Smith, economists have always recognized that there were profound and close relationships between law and economics. By the middle of the 1800s, economists in Europe and the United States had investigated the effects of laws on many subjects: slavery, child labour, interest rates, the value of money – to mention but a few.

The older writings on law and economics were mostly confined to obvious situations where the law affected economic activity. A conspicuous characteristic of this literature is that it was almost always based on explicit value judgements about the social effects of the laws and their economic consequences. Most of these writings concerned the effects of the law on economic development, with the role of the state in social welfare a powerful secondary theme. The economic analysis was weak compared with the strength of the calls for changes in the law based on moral fervour. Faith that changing the law would improve welfare was widespread.

Before the late 1800s, the main tools for economic analysis were solid logic, sound judgement and common sense. There were relatively few specifically economic tools, such as diminishing returns and comparative cost. With the coming of Alfred Marshall's analytic engine and the triumph of marginalist analysis, the range of problems amenable to economic analysis was expanded.

Around the middle of the 1930s, a group of economists began theoretical research that transformed microeconomics into the mathematical framework of today. About the same time, or shortly after, there was a rapid development of mathematical techniques useful for economic analysis. The changes that followed brought greater precision and the ability to deal with a wider range of economic phenomena. It also meant results that were usually more abstract and less representative of the social world. Older economic analysis was closely tied to the circumstances of the problems with which it dealt.

From its early beginnings as a distinctive subdiscipline in the late 1950s, to the developed and expanding field of the mid-1990s, analysis in law and

economics came to rely on a number of basic concepts and tools. The most important and distinctive of these are transaction costs; some hypotheses about maximization, for example of wealth, income or health; concepts of ownership, such as property rights and the optimal division of these; consideration of risk and uncertainty; and the effects of legal and economic changes. These are used within the framework of microeconomics which rests upon such ideas as scarcity, the valuation of alternatives and opportunity costs.

The main methods of law and economics are quantitative and empirical. An important methodological requirement is that results should be testable by means of falsification. Predictions should be verifiable on the basis of empirical evidence. A hallmark of the subdiscipline is the application of economic theory, mostly microeconomic, to subjects that now encompass the entire legal system. The scope of economics is virtually coextensive with that of the law. Public choice economics makes government one of the subjects of economics. Never abandoned was the faith that changing the law, by adding new ones, or eliminating or adjusting old ones, could improve social welfare.

The subdiscipline of law and economics can be seen as having two, often overlapping, approaches. One is in the application of economic analysis to the workings of the legal system. The other is the study of the way the legal system influences the performance of the economic system. With this picture of the field, it is now possible to get some idea of Stigler's influence on both approaches during its formation. He drew attention to important costs that had not been fully recognized – the costs of obtaining information are an important component of transaction costs.

Stigler was a total empiricist in the sense that he had a passion for measurement coupled with the belief that both theory and practice had to have verifiable evidence of their effects. An empirical orientation and commitment to quantitative evidence for the testing of hypotheses were outstanding characteristics of his work. Both of these characteristics are now firmly entrenched in the field of law and economics. Stigler's influence was very important in establishing them as standards for the field. Today there are few journals that would accept an article without a falsifiable hypothesis based on an empirical foundation.

Law and economics owes another of its characteristics to Stigler. Work in the field often avoids the complexities of the interdependence of the parts of the legal system. In this it follows Stigler's treatment of the oligopoly problem. It often assumes that the actors on the legal scene – legislators, judges, lawyers, regulators and policemen – will solve the details of their problem because it is in their interests to do so. This was a forerunner of the 'efficiency' assumption, now a common one.

Stigler's influence was important in establishing the necessity for a strong foundation in microeconomics as a requirement for the development of work

in all the fields in which he was interested. The necessity for a solid microeconomic foundation is now recognized widely in studies of law and economics.

Another of his contributions stemmed from the way he changed how economists analyse governmental regulations. His work in this field opened new vistas for intellectual inquiry. Government regulation begins with laws. The questions that Stigler posed about the way regulations actually worked, and the technical sophistication with which he answered them, were unprecedented.

Stigler changed the way economists looked at the role of government. For a long time, most economists thought that the primary economic role of government was to correct private market failures. Stigler showed that other things were going on and ended the tendency of economists to accept conclusions about the role of government without empirical verification. It was a short step from expanding these ideas about regulation to the entire field of law. He extended the economists' rational behaviour model to regulation. He argued that political behaviour was rational and needed more sound analysis: 'The failure to analyze the political process – to leave it as a curious mixture of benevolent public interest and unintentional blunders – is most unsatisfactory' (1981, p. 9). There is now a large body of literature based on the assumption of rationality for subjects far different from those of utilities which were his early preoccupation.

Stigler insisted that sound theory had to generate empirically refutable implications if it was to be taken seriously. Work that did not meet these specifications he immediately exposed and sometimes ridiculed. Along with his colleagues at the University of Chicago (hereafter Chicago), he set new standards of analytical rigour for the profession. Not quantifiable, but a requirement if a scientific field is to advance, are scientific standards. These Stigler embodied and advanced. His work exemplified the standards of intellectual integrity, analytical rigour and a respect (perhaps 'devotion' is a better word) for evidence. Every topic he studied was subject to acute intellectual analysis. Stigler insisted upon, and his work also reflects, the practical uses of economics with empirical support for theory. He stood for the idea that, if economic theory is taken out of law and economics, you remove the skeleton which supports the rest.[2]

Stigler was always guided by the idea that freedom – an important requirement for work in the social sciences – was the greatest good. Both his professional work and his activities that had political overtones were influenced by a commitment to a free society. His political views on government intervention were driven by factual investigations. He, like Milton Friedman (b. 1912, Nobel Prize 1976), was one of those rare specimens of economists in his understanding of the fundamental ideas of science.

Stigler began his autobiographical lecture: 'It is a good rule that a scientist has only one chance to become successful in influencing his science, and that is when he influences his contemporaries' (Stigler, 1986, p. 93). His personal life, a near-perfect example of his rule, casts some light on both his influence and his development. It gives clues to the answers to two questions: 'By what means did Stigler affect the early development of law and economics?' and 'Why is there not more direct recognition of his contributions?'.[3] How Stigler affected the early development of law and economics is explained by one of the ways he did it, through close personal association with those who were important in shaping the field. Why there is not more direct recognition of his contributions is directly related to this: many were made in face-to-face interactions.

His public fame came from the usual sources. He influenced his own students through ideas and personality and generations of others through his popular textbook. His colleagues were also primarily influenced by his personality and ideas. Prolific and high-quality writings, presence at professional meetings and editorial activities made him widely known and respected among economists. He edited the *Journal of Political Economy* for 19 years, beginning when he was 61 years old. He was an imposing, for some an awe-inspiring, figure, more than six feet tall. Quick and witty, he wrote with a rare grace.

His influence on law and economics had a somewhat more private and personal origin. It centred on Chicago, where many of the methods, analytical techniques and concepts important in the study of law and economics were developed. Stigler began his graduate studies at Chicago in 1933, at the age of 22, just at the exciting time when the modern transformation of economics was in its infancy. Important influences came from some of the most significant economic minds of the time. Among them were Frank Hyneman Knight (1885–1972), under whom he wrote his doctoral dissertation, Henry Calvert Simons (1899–1946), who was a close friend, and Jacob Viner (1892–1970). He received his doctorate from Chicago in 1938.

For most of his academic life he was closely associated with people who shaped the new field: Aaron Director (1901–2005), Coase and Friedman. Stigler and Friedman were friends from the time they were at Chicago in the mid-1930s. Stigler took a class in statistics from Director in the 1930s, but only became a close friend from the time of the first meeting of the Mont Pelerin Society, in 1946. Coase joined the faculty in 1964 and also became a close friend.

Stigler joined the Chicago faculty in 1958. The economics community was then in 'a stage of high prosperity'. Indeed, an argument could be made, on the basis of Nobel prizes alone, that no other university had a comparable economics faculty during Stigler's time. He acknowledged the importance of

social relationships when he wrote that 'it is difficult to do creative work if one is not in a congenial intellectual environment' (Stigler, 1986, p. 101). He might have added that being well financed helped.

There are many testimonies to the importance of the stimulating and collegial environment the early workers found at Chicago. Both Coase and Stigler tell the story, in more than one place, of how the Chicago economists, including Friedman, met at the home of Aaron Director to hear Coase, whom they had invited to defend his work (Stigler, 1988, p. 75; Coase, 1996, p. 107; Kitch, 1983, p. 221). One outcome was Coase's famous article on the problem of social cost – one of the most cited in the literature. Another was the convincing of a group of brilliant economists of the possibilities inherent in the new idea.

Director's fame as one of the founders of law and economics was also a result of personal interaction. All of those around him recognized his brilliance, but he did not write. He left that to his colleagues and students. One consequence is that many of his ideas are found in the writings of others. An example, from probably hundreds, is Stigler's article, 'Director's law of public income redistribution' (1970). In 1981, many of the key figures in the early development of law and economics met in Los Angeles to pay tribute to Coase and Director as founders of the field. The transcript of that meeting was edited by Edmund W. Kitch. It is full of evidence of the force of the personal interactions, including the constant discussion and correction of ideas.

Kitch makes the point that innovation came from the fringe, not the mainstreams of economics and the law (Kitch, 1983, p. 233). Stigler was part of the mainstream in the fields of industrial organization, the functioning of markets, the causes and effects of public regulation, the economics of information and the history of economic thought. He was part of the fringe in his work on economics and the law. Stigler's contributions were of a unique kind and made in a highly personal way. Both the contributions and the way they were made were shaped by circumstances at Chicago at the time.

Although many scientists are dogmatic – Stigler was fond of the quotation to the effect that progress does not happen until the old professors die off – he was not. This trait, along with his curiosity, undoubtedly affected both the range of his interests and his influence. He was always able to change his mind on the basis of evidence and logic, not a conspicuous characteristic of his, usually vocal and second rate, detractors. Two examples of change must suffice. In his *Memoirs* he tells how Coase changed his mind (Stigler, 1988, pp. 73ff.). The other example is his shift from an activist antitrust policy early in his career to a minimalist antitrust policy. In his early career he believed that regulated industries would and could avoid effective regulation. He later revised this as he became convinced that regulators and those in charge of

antitrust enforcement used weak economic analyses and responded to political pressures which did not have the desired results of increasing competition. In short, they often captured the regulation process and used it for their own purposes.

Like his personality, much of what he wrote was provocative, in the sense that it forced people to think about what he said. His mordant wit made him difficult to like for those who did not know him well. Although often called 'conservative' by his enemies or those who did not understand him, he was really non-partisan. The political positions he took were derived from his approach to economics. Historians of economic thought are not always kind to those whose work gets integrated into a field. Sometimes, as in Stigler's case, when influence comes the way it did, their work is virtually forgotten. When an approach and methods become part of the fabric of a discipline, the names of the pioneers are likely to be neglected.

Stigler was one of the great economists of his generation. He left his mark on everything he touched. The things he insisted upon, sound theory and logic and empirical verification for work on practical problems, are now some of the hallmarks of the modern study of law and economics. In many ways, each new application of law and economics is a tribute to the value of Stigler's contributions.

When answers are found to the questions Stigler raised in his work on the history of economic thought about whether there are laws or regularities that shape the growth of knowledge itself, then it will be possible to put his contributions to the development of law and economics into a more precise framework. Until that time, the matter must be left with the recognition that his approach, the standards he represented, the methods he used and the subjects he chose to study are now firmly embodied in the field.

Notes

1. One of his most cited articles is not in the bibliography. 'The scientific uses of scientific biography: with special reference to John Stuart Mill' in *The Economist as Preacher and Other Essays* was reprinted. The note on the title page (p. 86) says that the source was *James and John Stuart Mill: Papers of the Centenary Conference*, ed. John M. Robson and Michael Laine (Toronto: University of Toronto Press, 1976).
2. It is interesting that some of the features of the study of law and economics are the basis for hostility in many welfare state countries. Some of the hostility stems from misunderstandings. Law and economics is not equivalent to microeconomic theory or private enterprise, nor is it against government intervention. Stigler's approach was to use the best theory possible to find out, on an empirical basis, what the economic results of laws and regulations are. He felt that it was next to criminal for economists to advocate policies for which results were not obtainable or known.
3. Obvious explanations for the lack of direct recognition of his contributions to the early development of law and economics are that he made none, or those that he made were inconsequential. These can be dismissed in the light of the evidence.

Stigler's publications in legal and law and economics journals

1948 'Review of Eugene V. Rostow's book, *A National Policy for the Oil Industry*', *Yale Law Journal*.
1951 'Review of Richard B. Tennant's book, *The American Cigarette Industry*', *Yale Law Journal*.
1955 'Mergers and preventative anti-trust policy', *University of Pennsylvania Law Review*.
1958 'The economies of scale', *Journal of Law and Economics*.
1961 'Private vice and public virtue', *Journal of Law and Economics*.
1962 'What can regulators regulate? The case of electricity' (with Claire Friedland), *Journal of Law and Economics*.
1963 '*United States* v. *Loew's Inc*. A note on black-booking', *Supreme Court Review, 1963*.
1965 'The dominant firm and the inverted umbrella', *Journal of Law and Economics*.
1965 'The economic effects of the antitrust laws', *Journal of Law and Economics*.
1967 'Comment' (on 'The application of an entropy theory of concentration to the Clayton Act'), *Yale Law Journal*.
1969 'Alfred Marshall's lectures on progress and poverty', *Journal of Law and Economics*.
1970 'Director's law of public income redistribution', *Journal of Law and Economics*.
1971 'The law and economics of public policy: a plea to scholars', *Journal of Legal Studies*.
1971 'Henry Calvert Simons', *Journal of Law and Economics*.
1974 'Law enforcement, malfeasance and compensation of enforcers' (with Gary S. Becker), *Journal of Legal Studies*.
1975 'Determinants of participation in presidential elections: comment', *Journal of Law and Economics*.
1975 'The effects of economic policies on votes for the presidency: some evidence from recent elections: comment', *Journal of Law and Economics*.
1976 'The sizes of legislatures', *Journal of Legal Studies*.
1978 'Wealth and possibly liberty', *Journal of Legal Studies*.
1980 'An introduction to privacy in economics and politics', *Journal of Legal Studies*.
1983 'The literature of economics: the case of Berle and Means' (with Claire Friedland), *Journal of Law and Economics.*
1983 'What does an economist know?', *Journal of Legal Education*.
1985 'The extent of the market' (with Robert A. Sherwin), *Journal of Law and Economics*.
1985 'The origin of the Sherman Act', *Journal of Legal Studies*.
1989 'Address' (to the 65 Annual Meeting of the American Law Institute, 19 May), *Proceedings*.
1989 'Two notes on the Coase theorem', *Yale Law Journal*.
1992 'Law or economics', *Journal of Law and Economics*.

Full publication details for the items above can be found in the Longawa bibliography cited below.

References

Anon, (1993), 'In Memoriam: George J. Stigler', *Journal of Political Economy*, **101** (5), October, 759.
Barro, Robert L. (1991), 'An unregulated economist', *Wall Street Journal*, col. 3, p. 14, s. A, Monday, 9 December.
Becker, Gary S. (1993), 'George Joseph Stigler: January 17, 1911–December 1, 1992', *Journal of Political Economy*, **101** (5), October, 761–7.
Breit, William and Roger W. Spencer (eds) (1986), *Lives of the Laureates: Seven Nobel Economists*, Cambridge, MA: MIT Press.
Coase, Ronald (1991), 'Stigler, George J.', in Edward Shils (ed.) (1991), *Remembering the University of Chicago: Teachers, Scientists and Scholars*, Chicago: University of Chicago Press.
Coase, Ronald (1996), 'Law and economics and A.W. Brian Simpson', *Journal of Legal Studies*, **25** (1), January, 103–19.

Kitch, Edmund W. (ed.) (1983), 'The fire of truth: a remembrance of law and economics at Chicago, 1923–1970', *Journal of Law and Economics*, **26** (1), April, 163–234.
Longawa, Vicky M. (1993), 'George J. Stigler: a bibliography', *Journal of Political Economy*, **101** (5), October, 849–62.
McCann, C.R., Jr. and Mark Perlman (1993), 'On thinking about George Stigler', *Economic Journal*, **103** (419), July, 994–1014.
Passell, Peter (1991), 'George Joseph Stigler dies at 80; Nobel prize winner in economics', *New York Times*, col. 1, p. 12, s. B, Tuesday 3 December.
Stigler, George J. (1970), 'Director's law of public income redistribution', *Journal of Law and Economics*, **13** (1), April, 1–10.
Stigler, George J. (1976), 'The scientific uses of scientific biography: with special reference to John Stuart Mill'; reprinted in *The Economist as Preacher and Other Essays*, Chicago: University of Chicago Press/Oxford: Blackwell, 1982.
Stigler, George J. (1981), 'The economist as preacher'; reprinted in *The Economist as Preacher and Other Essays*, Chicago: University of Chicago Press/Oxford: Blackwell, 1982.
Stigler, George J. (1982), *The Economist as Preacher and Other Essays*, Chicago: University of Chicago Press/Oxford: Blackwell.
Stigler, George J. (1986), 'George J. Stigler', in William Breit and Roger W. Spencer (eds) (1986), *Lives of the Laureates: Seven Nobel Economists*, Cambridge, MA: MIT Press.
Stigler, George J. (1988), *Memoirs of an Unregulated Economist*, New York: Basic Books.

58 Pietro Trimarchi (1934–)

Giampaolo Frezza and Francesco Parisi

Introduction

In 1961, Pietro Trimarchi published *Rischio e Responsabilità Oggettiva* (Risk and strict liability) (Trimarchi, 1961), a monograph that has become a 'milestone' in the economic analysis of the law in Italy. Meanwhile, in the United States, the writings of Ronald Coase, Richard A. Posner, Guido Calabresi and Gary Becker were popularizing the so-called 'new law and economics'. Trimarchi, however, studied the delicate question of strict liability separately. His independent study focused on the function and the structure of strict liability in microeconomic terms.

Trimarchi's works include other publications concerning the economic approach to law. Particularly relevant, and quoted in Trimarchi (1961), is *Il 'caso fortuito' quale limite della responsabilita per il danno da cose* ('Fortuitous event' as a limit of the liability for damage caused by things) (Trimarchi, 1959). Another monograph, titled *Causalità e danno* (Causality and damage) (Trimarchi, 1967), synthesizes the content of his theory on strict liability for enterprise risk.

This chapter first assesses Trimarchi's theory and also devotes particular attention to the economic analysis of the principles of contractual liability that Trimarchi takes pains to develop and contrast with widespread and respected theories from the United States. According to Trimarchi, a legal system, not a conventional system of contractual liability, provides an optimal distribution of productive resources, as supported in a famous essay by Coase (1960). In 'L'analisi economica del diritto: tendenze e prospettive' (Economic analysis of law: tendencies and perspectives) (Trimarchi, 1987), originally presented at a congress in Padua, 21–23 May 1987, the author gave Italian scholars an overview of the history, methodology and goals of the economic analysis of law, emphasizing tendencies and practical aspects through a critical analysis. The concluding section will consider Trimarchi's approach to liability in contracts, with particular attention to his article 'Sul Significato economico dia criteri di responsabilite contrattuale', (On the economics of liability in contracts) (Trimarchi, 1977).

Risk and strict liability

Trimarchi sets out to determine whether and how the Italian legal system recognizes a principle of strict liability for 'risk of the enterprise', in his

monograph *Rischio e Responsabilità Oggettiva* (Trimarchi, 1961),[1] where such liability is analysed according to its function and structure. After an enquiry into the foundation of strict liability, Trimarchi devotes himself to a deeper study of typical cases of no-fault liability. He develops his analysis following a strict conceptual order, which can be schematized as follows:

1. function of the rules concerning the strict liability for the risk of the enterprise in the Italian legal system;
2. scope of applicability; and
3. parameter of the risk.

Although Italian legislators did not intend to adopt a general system of strict liability (see the illustrative commentary of the secretary of state in the 1942 civil code), the Italian legal system introduced the concept of strict liability in some special cases of civil responsibility. According to Trimarchi, two different perspectives explain its introduction: on one hand, a detailed examination of some rules of the Civil Code (articles 2049, 2051, 844), of the Navigation Code (articles 965, 978) and of the special legislation (29 July 1927, no. 1433, articles 10, para. II, and 31) may explain it, on the other, 'the actual economic and social needs and their intensity, and the fact that the strict liability for the risk of the enterprise has been affirmed by most industrialized nations', could also provide an explanation. Trimarchi studies the function and structure of strict liability in relation to the various concepts formulated by the modern legal doctrine, and proceeds to establish which of them may be considered realized or realizable and in harmony with the rules of the Italian legal system.

The economic function of liability

First Trimarchi highlights that in the nineteenth century the principle generally accepted by the legal systems was expressed by the phrase 'no liability without fault'. Nevertheless, the concept of vicarious liability was accepted by the Napoleonic Code and also by the Italian Civil Code of 1865, which was based on the French framework.

In Trimarchi's view, acceptance of the liability without fault principle between the nineteenth and twentieth centuries was of fundamental importance, because the period was characterized by an intense process of industrialization. Trimarchi held that industrial plants create unavoidable dangers despite maximum care being taken. Sometimes, a worker's imprudence causes an injury, and the principles of fault liability are sufficient to address liability. When the injury is caused by the operation of the plant, however, principles of fault liability are insufficient. Accordingly, a rule connecting the risk to the profit assigns liability to the plant owner because he can obtain

insurance, or, in modern law and economics language, internalize the costs of the injuries by raising the price of the goods he produces.

Since the concept of risk liability spread, Italian legal doctrine has attempted to examine it in detail. Therefore, an analysis and criticism of the relevant theories was essential before Trimarchi could build an economic analysis of risk liability. Thus, he analyses and criticizes the theories founding liability on pure causation, fault, profit, fairness, insurance or mobile combinations of these elements. He undertakes a careful analysis of such theories, and elaborates a model that constitutes the milestone of strict liability which operates everywhere with substantial unity of foundation:

> This system considers that the risk introduced into the society is part of its social cost and therefore must be borne by the entrepreneur as part of his production cost. The theory, in simple terms, is that the benefits of an enterprise must cover its costs.

Thus, an enterprise's strict liability for risk plays an economic role connected to the economic theory of distribution of costs and profits, an essential element in the strategic choices of an enterprise. Trimarchi suggests that the cost of running an enterprise includes not only the costs of labour, raw materials, and ordinary wear and tear of machines, but also the costs of the physical and economic injuries caused to third parties. If the legal system does not place the burden of the cost on the entrepreneur who creates the risk, a marginal sector or industry may appear profitable to the entrepreneur, while from a social point of view these marginal sectors only survive because the cost of the risk they introduce is borne by third parties. In the language of modern law and economics, the creation of risk without liability allows the entrepreneur to externalize some costs of production. On the other hand, under a negligence rule, the potential injurers may escape liability for part of the harm they cause, with a socially inefficient outcome, because it is not practicable for courts to compare benefits and costs on a case-by-case basis and check the profitability of the activity in order to ascertain whether the risk was economically justified: the standard of care is therefore determined by the courts only in general and abstract terms. In lieu of a more costly, limited and uncertain judicial assessment of social benefits and costs, the system of strict liability automatically places on the entrepreneur a socially efficient incentive to control the risk created by his production activity. The entrepreneur may adopt supplemental security measures, substitute a safer method of production, or increase productivity. In extreme cases, the marginal section of the enterprise would have to evolve or be terminated. (See also Trimarchi, 1967, pp. 134–8.) As Trimarchi notes:

> It is important to highlight that this discussion applies not only to particularly dangerous enterprises, but to any organized and continuous enterprise that is likely to cause injuries (both physical and economic) to third parties, no matter the size, frequency, or predictability of these injuries, because they can always be translated into cost.

This pressure, when exerted on those who have general control of the risk, may be more efficient in certain cases than the incentive created by a fault-based liability system that is exerted on those who participate in a risky activity. (Some research conducted in the United States on psychology of incidents supports such statements.)

The relationship between risk liability and the possibility of insuring against such risks thus assumes particular relevance. Although Trimarchi does not believe it appropriate to understand the theory of liability only as an 'insurance' mechanism, he deems the insurability of the risk to be very important in the design of a liability system. Thus, an entrepreneur should be liable only for injuries caused by ascertainable, calculable risks. Only then can the entrepreneur include the risks as costs and protect against liability by insurance or self-insurance. Accordingly, as Trimarchi notes, 'there must be a constant (not occasional) relationship with the risk'. Furthermore, according to the author, strict liability promotes insurance and this, in the absence of a general social security system, implies the additional benefit of preventing the further social harmful repercussions of uncompensated injuries. (See Trimarchi, 1961, p. 39.)

The structure of strict liability

Trimarchi suggests that the risk liability concept plays its role only if applied to economic activities, and not biological ones. Biological activities are characterized by the element of necessity, not by economic rationality, thus calculating costs and benefits in relation to them makes little sense. Therefore, Trimarchi concludes that liability must be based on an activity that 'is the result of an economic decision, that has a minimum of continuity and organization, and presents a considerable risk' if the concept of risk liability is to fulfil its function. Most biological activities (that is, the non-economic activities that are not organized or risky) are subject to the 'no liability without fault' principle.[2]

According to Trimarchi, the structural study of strict liability also requires an analysis of the differences between the strict liability system and the so-called strict liability for 'avoidable risk'. The latter category encompasses situations where the law's formula subjects a tortfeasor to liability unless 'he proves that he adopted all measures and took all proper steps to avoid the injury (both physical and economic)'.[3] According to Trimarchi, the function of strict liability for avoidable risk is limited compared to strict liability that

encompasses all kinds of typical risks. Strict liability for avoidable risk requires only the adoption of security measures for the system of production currently adopted by the entrepreneur, while general strict liability may also induce a radical change of production system or the partial or total termination of the enterprise. Thus, the former operates only within the competition among firms that produce the same type of goods utilizing similar technologies. The latter, meanwhile, also operates on competing firms that produce the same products with a different technology, or that produce different but competing goods or products. More specifically, strict liability creates an incentive to discover new technologies that will reduce accident losses, while a negligence rule, or a rule of strict liability for avoidable risk, merely apply a pressure to take precautions that are reasonable, or possible, given the existing technology, with a benchmark that does not induce evolution and technological improvement.

Considering the subjective approach (that is, the determination of the person who is objectively liable), unless specifically explained by the rules of the civil code (such as articles 2053 and 2054), strict liability is only imposed on the entrepreneur, because he controls the general conditions of the risk, can calculate its cost, and include the costs in the firm's balance sheet through insurance. Trimarchi notes that 'there must be a constant (not occasional) relationship with the risk' to create the practical incentive to seek insurance.

The employer's vicarious liability

The author criticizes the legal doctrines that base the employer's vicarious liability for the torts of servants and employees on a 'presumption of fault'. These doctrines attempt to maintain the unity of the system based on the 'no liability without fault' principle. By studying the initial drafts of the Napoleonic Code and by analysing the nineteenth-century common law cases, the author confirms that the employer's vicarious liability existed with conceptual autonomy, independent of the criterion of fault. The need for vicarious liability, which increased along with the development of the modern industrial society, is also evident in cases of damage caused by machines. Thus, Trimarchi purports to identify the foundation of vicarious liability. In particular, the criteria identified by the legal doctrine and the case law are:

1. The control criterion;
2. the economic dependence criterion; and
3. other criteria elaborated by the French and American doctrines.

Trimarchi's analysis begins by noting that people have no practical interest in insuring against a risk unless they can foresee and calculate the risk, and as long as they are judgement-proof' because of a lack of assets. Thus, vicarious

liability must be placed on the owner of a business, because repeated exposure to the risk allows for calculation, and the owner is not without assets. Only the owner can reduce or distribute the risk costs through insurance or modification, substitution or termination of the risky activity. By placing liability on the person who can translate the risk into cost, the rule provides an economically efficient incentive while guaranteeing recovery for injured third parties and keeping total damages within tolerable limits.

Having clarified the function of article 2049 of the Italian Civil Code, Trimarchi analyses its scope of application. The summary of the factual situation considered can be represented as follows: B performing a service on behalf of A causes unjust damage to C. The problem here is whether A is liable to C. Trimarchi stresses the fact that the legal relationship between A and B may vary with the nature of the actual relationship: relationships between A and B can be contractual (for example, a private contract or employer/employee relationship); corporative (relationships within a partnership); legal; or based on mere friendship. He takes account of each type of relationship separately.

The first type of relationship (contractual relationship between A and B) received particular attention in the legal doctrine because it occurs so frequently. At the outset, if A and B have an employer/employee type of relationship, the torts of B are within the scope of vicarious liability because B's torts are among the risk of A's enterprise (when A is both employer and entrepreneur). However, if the relationship between A and B is not defined by the parties as an employer/employee relationship, two elements must be clarified. First, for the operating party to act without implying the vicarious liability of the other party it is necessary to have capital to operate the business enterprise and there must be enough capital to cover the damages caused by the risks created. Second, there must be enough capital to give some economic incentive to take steps towards prevention. In such cases the judge has to decide whether the capital is large enough to cover the risk (on an empirical basis).

When the activity does not require significant capital, the criterion of the ownership of capital cannot be applied and continuity and exclusivity of the relationship between A and B becomes determinative. In these cases, if the relationship between A and B is occasional, the word 'risk' may not be appropriate because, as defined, risk requires static regularity and continuity over time. Here the tort committed by B against C is considered *casus fortuitus* to A, and does not fit within the concept of vicarious liability.

Trimarchi notes that the law places limits on attempts to immunize the master/employer by burdening the employee with the risk. Such immunity is allowed only when the performing party can effectively perform at his own risk, so that the principal cannot profit by, in modern economic language,

externalizing the cost of the risk by contract. The determination that an immunizing contract is not valid is made on objective criteria, without reference to the American doctrine of bad faith. This is sensible to Trimarchi, because where the risks are distributed according to objective economic criteria, intent (or lack of intent) to breach the law is irrelevant.

This approach is confirmed by Italian law 1369, of 23 October 1960, enacted contemporaneously to Trimarchi's work on the subject.

According to Trimarchi, only the contractual nature of the underlying relationship is relevant for risk sharing, while the criteria adopted by the legal doctrine and the courts[4] have only persuasive value. The immediate and practical effect of determining the criteria is that Trimarchi can identify the scope of the rule's application. Thus, he devotes several pages of his monograph to considering whether vicarious liability, identified according to the economic analysis of the law, can apply in the following cases:

1. satellite enterprises and subsidiaries;
2. subcontractors;
3. small businessmen;
4. professionals and artists;
5. managers and administrators;
6. agents and attorneys;
7. athletes;
8. regular partnerships, *de facto* partnerships and similarly structured contracts;
9. non-contractual relationships;
10. accessory service; and
11. complimentary service.

The issue involving employees borrowed by a different employer also deserves analysis. What happens when A lends to B his employee C (borrowed servant) for a certain period of time, and C causes injuries (both physical and economic) to an unrelated party D while performing his work for B? Who is liable for this injury, A, or B, or both? The Italian courts have traditionally resolved such problems through the imposition of joint and several liability on A and B. Meanwhile, the French and English courts apply the control criterion. According to Trimarchi, these approaches are not satisfactory.

The problem, once again, is the accurate application of the concept articulated by article 2049 of the Italian Civil Code. As analysed under a functional (economic) approach, the vicarious liability of the employer provides economic sharing of the risk created by the enterprise. Seen in this light, the Italian cases fall into the following categories:

1. A is an entrepreneur; B is not an entrepreneur. The vicarious liability will only apply to A, who is able to bear the risk of the enterprise by translating such risks into costs and by controlling the risks and associated costs.
2. A is an entrepreneur; B is also an entrepreneur. If the service is performed by C only for the benefit of one of the two enterprises, the problem is easily resolved: liability is attached only to the enterprise that is benefiting from the service.
3. A is an entrepreneur; B is also an entrepreneur; the service is performed by the worker for the benefit of both enterprises. In this case, if A and B have contractually allocated the risk, their agreement should assign the risk. If such a contract does not have clauses regulating liability towards third parties, the problem may be solved by stipulating an insurance contract that covers the risk. If the problem remains unresolved, the author assumes that both A and B are jointly and severally liable, with equal shares in the internal apportionment of the loss.

Trimarchi notes that the solution to (3) suits both the structure and the function of the rule in article 2049 of the Italian Civil Code:

> It suits the function of risk distribution because, in our hypothesis, it is strictly related to both enterprises and can be predicted, calculated, insured, and translated into cost by either of the two entrepreneurs. It suits the preventive function, because both entrepreneurs can control the employee, though in a different manner.

Before vicarious liability is established against an employer, the rule of vicarious liability requires that the employee's injury-causing act was related to the performance of his duties. Indeed, the employee is not always a mere cog in the industrial and commercial machine. Employees are capable of independent actions that should not impose vicarious liability on the entrepreneur. Trimarchi therefore determines the scope of employee actions that can result in vicarious liability of the employer.

After determining the function, structure and scope of the rule of vicarious liability, Trimarchi proceeds to the essential study of the concept of fortuitous event, understood as limit of liability for damage caused by things. The study of the doctrine of fortuitous event suggests the possibility, recognized by the Italian legal system, that the fortuitous event represents key evidence in favour of the entrepreneur. Again, Trimarchi's discussion connects the rule to its function.

First, the fortuitous event is characterized by its *relativity*. In other words, the determination of the fortuitous character of an event varies according to the approach adopted. In this sense there are no fixed rules establishing a

benchmark for determining which events may qualify as fortuitous. The determination depends on the practical objectives of the rule to which it is applied. Second, Trimarchi demonstrates through a variety of arguments that it is inadequate to equate a fortuitous event with lack of fault. Having reached these conclusions, he argues that further analysis of the concept of *fortuitous event* requires a re-examination that refers to the function of the rule in which it appears. In the context of the complementary rules of strict liability for enterprise risk and vicarious liability, events that create risks which are not ascertainable or calculable should be considered fortuitous. In the context of enterprises, even very low-risk activities present ascertainable, calculable risks because of repetition. At the same time, repetition increases the social utility of the activity because of learning and economies of scale. Thus, strict and vicarious liability provide an incentive for the entrepreneur to internalize costs, while those events that cause injury but which do not allow the entrepreneur to internalize costs should be considered fortuitous. Thus, entrepreneurs are liable for the risks that are usually related to their activities and their liability shall not extend to the unforeseeable harmful consequences occasioned by their activities. As such, Trimarchi finds no connection between liability for injury losses occasioned by things and the idea of fault. The former relates to the economic distribution of costs and profits, which is inherent to the enterprise. Such economic distribution of profits and costs requires that whoever organizes a business for his own profit must bear the risk of injuries to third parties. According to this view, the fortuitous event doctrine constitutes a limit to the so-called 'risk of enterprise' or, in general, 'risk of economic activity'.

Next, Trimarchi analyses the relationship of fortuitous event and *force majeure*. These terms are often confused in legal doctrine and dominant jurisprudence (both Italian and foreign). As with all of Trimarchi's analyses, his discussion of *force majeure* is based on the function of risk liability. Therefore, the definition of *force majeure* as 'unavoidable event' seems unconvincing. If it is understood as an unavoidable event, and if the function of strict liability, as mentioned earlier, is to create an economic incentive to eliminate enterprises (or parts of enterprises) that are socially unproductive because of the connected unavoidable risks, then damage caused by *force majeure* must be borne by the entrepreneur. Because inevitability is an essential aspect of *force majeure*, it is inadequate to limit the risk of the enterprise.

In sum, Trimarchi's approach does not exclude unavoidable risks from the risk of enterprise. *Inevitability*, narrowly interpreted, however, may justify the exclusion of an injury (both physical and economic) from the concept of risk when the risk would not be reduced even if the business activity is curtailed. For example, liability could be excluded when the damage is caused by an external factor that is not facilitated, amplified or extended by the product.

Thus, Trimarchi characterizes the concept of fortuitous event by reference to the following factors. The first is the level of predictability. If an event is such that it cannot be considered part of the typical risk of the activity, it may be deemed *fortuitous.* The concept of fortuitous events, indeed, 'refers only to events in which the risk is lower than such rare probability, because the fortuitous nature of an event constitutes a limitation to risk liability (which results from lawful activities), and not fault liability'.

Second, even damages caused by activities with low risk may not be exempted from liability as fortuitous when they are repeated in a business undertaking. Risk liability is normally connected to the performance of a business activity and not to single acts, thus liability may be attached even if the act that caused the injury (both physical and economic) is very safe, for even very safe acts may be risky when repeated.

The third factor is related to the knowledge available to the entrepreneur, according to which the foreseeability must be measured. The standard must be objective and thus connected to the science and technology available at the time when the entrepreneur should have insured himself against the risk. The function of strict liability is not to punish, but to attach liability to whoever objectively creates the risk that can be translated into cost and be economically manageable through the ordinary methods of a good manager. An enterprise which is mismanaged by the entrepreneur and his assistants must suffer the economic consequences of the mismanagement.

The concept of a fortuitous event can also be clarified by Trimarchi's hypothesis that article 2051 of the Italian Civil Code is not only applicable to dangerous activities, but is also applicable to business activities with risks that can be translated into costs. Strict liability is always connected to 'business activities that present a risk' and not to 'things that present a risk'. Trimarchi finds the legal foundation for his hypothesis in the legal ruling that only a fortuitous event excludes liability. In other words, proving the existence of a fortuitous event means proving that the loss is not connected to a manageable risk. Strict liability must be applied in the field of business activities and the same incident may constitute a fortuitous event to common citizen, but not to the entrepreneur.

Thus, Trimarchi concludes that the function of the rule according to which there is no liability for a fortuitous event is to exclude the entrepreneur from a non-manageable risk that cannot be taken into consideration in a cost–benefit analysis prior to its occurrence. Also, the fortuitous event category is narrowed by the operation of the rule of large numbers: the more developed the business organization to which the risk is inherent, the narrower the category of the fortuitous event.

After examining the relationship between *force majeure* and fortuitous event, Trimarchi briefly addresses the question of loss caused by the activity

of third parties unrelated to the business activity. The concept cannot be connected to the idea of fault, as some suggest, thus he states that the loss caused by a third party (intentionally or not) does not constitute another ground of exemption of the entrepreneur's liability, unless the third party's activity presents the characteristics of the fortuitous event.

The damage caused by objects: allocating liability

After examining the foundation of the rules on liability for harm occasioned by things, (articles 2051, 2052, 2053, 2054, subsection 4 of the Italian Civil Code and articles 965 and 978 of the Admiralty Code), and explaining the principle of risk liability that these rules develop along with the principle of vicarious liability, Trimarchi considers the criteria for allocating the liability among the parties which own, use or control the object that has caused the harm.

Again, Trimarchi tries to formulate principles for allocating liability that are adapted to the functions of the risk liability already described. Imagine that A rents a machine to B. If A is liable for damage caused by the machine, A will charge B more than otherwise, and use the money to insure himself against the risks of damage. If B is liable for damage caused by the machine, B will not pay as much, and will use the savings to purchase insurance. In such a situation the allocation of liability, if known in advance, might matter less since, through the market mechanism the cost of the accident would eventually fall on the party controlling the risk. However, Trimarchi analyses why placing liability directly upon this party would be more efficient; if only A or B has knows what risks the machine presents, then only the informed party would appreciate the need for insurance. Similarly, if neither A nor B has any assets and is judgement-proof, the party without assets would have no incentive to buy insurance. Under these circumstances, allocation of liability is very important. Thus, if only A knows of the risks and has assets because he is in the business of renting the machine, the risk must be allocated to A. If it is not, neither A nor B has an incentive to insure against the risk, and the unfortunate party to whom risk is finally attached will have made no provision to account for it. On the other hand, if the risk is appropriately allocated to A, then A will have to operate as safely as his competition or face reduction in his profits or unprofitability. Thus, risk liability functions to reduce risk through technical advancement yielding a more efficient organizational structure, better security measures and more frequent controls. When these measures are adopted, risk is reduced to a level compatible with the economics of the enterprise, and its methods of production. Moreover, risk liability requires that the liable party be in a position to predict the risk and to translate the risk into a cost that is evenly distributed over time. Only in this way is risk liability incorporated into the economic process without interrupting the equilibrium and planned flow of production.

Taking this economic and functional analysis into account, Trimarchi approaches the fundamental problem of allocating liability. Distinguishing between two situations is important. On the one hand, use of some objects, to a certain extent, is not related to risk. For example, a house can collapse and cause damage because of a defect in its structure unrelated to a lessee's use. On the other hand, some objects are rendered dangerous through their use. A typical example is the use of machines and instruments in a plant managed by an entrepreneur. In the first situation, the owner, not the tenant is always the person responsible. In the second situation, the entrepreneur, not the owner is responsible.

Trimarchi finds support for this analysis in the rules of articles 2051–54, subsections 3 and 4 of the Italian Civil Code, and formulates the generally applicable principle that liability should be placed on the person who is in the best position to predict and calculate the risk, and who is thus in the best position to control and modify the risk in response to liability. He notes that identifying this person depends on the facts of a case. In the case of a bailor/bailee, for example, liability is usually placed on the bailor, unless the bailee's relationship to the risk is of a sufficiently long term that the bailee can predict and assess the dimensions of the risk and take efficient precautions against it. Therefore, the bailee is generally responsible if he uses the object that causes the loss professionally.

In sum, according to Trimarchi, it is essential that liability attaches to a person with a sufficiently continuous relationship to the risk. Other aspects, such as the payment of the rent, maintenance of the object, or the consensus of the parties regarding these issues are merely persuasive, not determinative, factors.

Concurring of risks

The concurring of risks concerns the distribution of liability in cases where two or more activities subject to the no-fault liability regime cause a single loss together. Such losses result in the joint and several liability of those to whom the law attributes the risk.[5] The law creates a fundamental problem, however, because it does not establish any criteria for the distribution of the loss between the persons liable. According to Trimarchi, article 2055, subsections 2 and 3 of the Italian Civil Code only relate to fault liability, and therefore may not be applied by analogy in the context of no-fault liability. In attempting to solve this problem, Trimarchi distinguishes between situations where two activities subject to the no-fault regime cause an injury together on the one hand, and situations where one activity subject to the no-fault regime and one subject to the fault regime cause a loss. We shall examine each situation separately.

In discussing the issue of concurrent risks subject to the no-fault regime, Trimarchi avoids founding the discussion on the causality principle because,

in his view, it obscures, rather than clarifies, the real issues involved. The legal doctrine that divides liability among those who contributed to the injury in proportion to their degree of fault is meaningless. On the contrary, the injury may be attributed entirely to several causes, not to one cause more than another. According to Trimarchi, the concept of 'cause' never allows a definitive quantification of relative harm.

The proper solution to the problem of concurrence of liability without fault apportions the burden of the damages in a manner consistent with the function of risk liability. Again, the function of risk liability is to put economic pressure on the person who controls the activity to reduce the risks in accordance with the total costs of the activity. Given this function, Trimarchi suggests that the liability for a loss in a case of concurrence of risk without fault should be apportioned according to the relative risks created by the activities involved. Thus, he looks for a method of quantifying the risks according to the circumstances of the actual case. In his method, the use of statistics is indispensable for the evaluation of the risk because statistics prevent reliance on arbitrary, subjective elements. Accordingly, Trimarchi's method has two tiers: first, one determines statistically a general index of the danger inherent in the activities of the same type; and second, the index is modified according to the particular characteristics of the activity that causes the harm.

This analysis of joint and several liability is not exhaustive, however, because it does not take account of the fact that some activities cause small losses frequently while others cause large losses infrequently. If such different risks cause an injury, then calculating the amount of loss attributable to each risk is problematic. In such a case, the entrepreneur whose activity caused a risk of a small injury may find himself faced with joint liability for a greater loss than the economics of his activity had prepared him for. For example, he may have purchased insurance with a low coverage per claim and thus be forced into bankruptcy by a single, large claim. Trimarchi thinks that such liability could have been avoided if the entrepreneur had obtained an adequate insurance coverage. Given the economic function of strict liability as discussed above, Trimarchi suggests that the loss be shared by the entrepreneurs according to the amount each could have expected as a result of his activity.

In discussing situations where loss is caused by activities subject to both the risk and the fault regimes, Trimarchi identifies several possible types of cases. Liability may be caused concurrently by two or more persons who are liable both for risk and fault; by several persons, some liable for risk and fault, others only for risk, and others only for fault; or by several persons, some liable only for risk, others only for fault. He starts his discussion with the last case and bases his solution on the different functions of each type of liability.

Suppose that a smoker disposes of a cigarette butt in a large quantity of flammable material and so causes a fire. In this case, the fire is caused by the smoker's negligence, and by the pre-existing risk created by the person who was responsible for the container of flammable material. Risk liability, as discussed, forces entrepreneurs to internalize all the costs of their enterprise and compensates victims. Trimarchi suggests that the rule of contribution among joint tortfeasors should burden the person liable for fault with a sum which is an adequate punishment and burden the other party with the remaining loss. Fault liability, according to Trimarchi, functions to punish tortfeasors, to prevent harm and to compensate victims. Since the damages for fault (negligence) are broader than those for risk (strict liability) he suggests that the person liable for fault should pay the total damages, but be indemnified by those strictly liable for the amount of the damage caused by the risk they each created. As Trimarchi notes, there is a valid basis for the rule of indemnification in the Italian Civil Code.[6] Moreover, one can use the principle of indemnification according to the above criteria to solve all of the potential combinations of risk and fault liability identified.

Contractual liability and economic analysis of the law

According to Trimarchi, contractual liability has three fundamental objectives:

1. to improve the distribution of productive resources;
2. to distribute the economic burden of losses without substantially endangering the financial status of the enterprise; and
3. to reduce the cost of judicial distribution of losses.

The threat of contractual liability induces compliance with contractual obligations, and thus liability should be imposed on the party who can avoid the costs of breach at least expense. If the least-cost avoider rule dictates liability, parties can be forced to adopt socially efficient precautions to avoid breach; they will thus maximize social wealth, and achieve the objective of improving the distribution of productive resources. Trimarchi expresses his argument with mathematical notations, a noteworthy methodological step for traditional continental scholarship.

Imagine a contractual relationship between a group of suppliers, A and their clients, B. If the supplier A breaches the contract, he will not be paid and both A and B sustain a loss. The supplier's loss (lost profit and waste of labour, raw materials and capital spent attempting performance) is *a*, while the buyer's loss (caused by the disturbance of their economic plans) is *b*. Assume that the supplier can eliminate all risk connected to breach of contract by taking better precautions, with cumulative cost of $\forall$. Meanwhile, the

buyers B can also take precautions to avoid or reduce the costs of breach by investing ∃ (for example, by warehousing raw materials necessary to avoid a delay in production caused by breach). If the buyer takes precaution ∃, he eliminates the cost of breach *b*.

Clearly, A should take precautions if the cost of doing so is less than the cost of the risk avoided. Formally: precaution ∀ can only be efficient if $\forall < a + b$. Similarly, B should take precautions only if the cost is less than the cost of the risk avoided. Formally: ∃ can only be efficient if $\exists < b$. If only one of these inequalities is true, then only that precaution should be taken.

If both inequalities are true, then one must choose between precautions ∀ and ∃, because taking precautions eliminates the risk of losses *a* and *b*, and therefore precaution ∀ renders precaution ∃ useless. If $\forall < a$, precaution ∀ should be taken, and precaution ∃ would be superfluous. If, however, $\forall > a$, the correct choice for maximizing social wealth depends on the relative efficiency of each precaution. The net social benefit of each precaution is the value left when one subtracts the cost of the precaution from the total risk avoided by taking the precaution. Formally, the net social benefit of precaution ∀ will be equal to $a + b - \forall$, while the net social benefit of precaution ∃ will be equal to $b - \exists$. Thus, precaution ∀ should be taken if $a + b - \forall > b - \exists$, while precaution ∃ should be taken if $b - \exists > a + b - \forall$.

In all these situations, the question whether A should pay damages to B is resolved by establishing which party can take the socially efficient precaution. Thus, when precaution ∀ is socially efficient, A should be liable to B for the loss *b*, because then, as an effect of liability, A will have an incentive to take precaution ∀. (Recall that $\forall < a + b$.) Meanwhile, when it is socially efficient to take precaution ∃, A should not liable to B for cost *b*, so B will have an incentive to take the socially efficient precautions.

Trimarchi notes that some may object that A will avoid breach even without liability to retain the right to consideration. However, if A is not liable to B for damages, A will only take precaution ∀ when $a - \forall > 0$, while economic efficiency requires that account be taken of B's potential losses as well. Of course, if A is not liable to B, then B may choose its supplier more carefully, and thus create incentives for A's adoption of the efficient level of care. Clearly, however, in many cases the buyers B cannot determine the efficiency of their suppliers beforehand, so this pressure will not be sufficient. In these cases, the threat of contractual liability is the only wealth-maximizing pressure that can be imposed on A.

Trimarchi then examines the second function of contractual liability, that is, the efficient distribution of losses between the parties. Here, Trimarchi uses an example that calls to mind the efficiencies of scale, as well as a concept akin to risk aversion versus risk neutrality. His example is of deciding whether a hotel or its guest should purchase insurance against the risk of loss or theft of

baggage. He believes that the economically efficient result is for the hotel to insure against the loss. The hotel can take precautions more cheaply because it need only negotiate one insurance contract, and thus it is in a better position to prevent loss and theft through monitoring, personal-insurance or third-party insurance – whichever is cheapest. Finally, the last objective of contractual liability is the reduction of the judicial costs of allocating the losses among the parties. According to Trimarchi, reduction of judicial costs may be achieved by adopting simple and certain rules and procedures.

After investigating the economics of contractual liability, Trimarchi considers two theories of contractual liability developed by Italian courts, and determines which one maximizes wealth. Italian courts base liability on fault, connected to the violation of rules of diligence, or according to the principle of strict liability, understood as liability for mere breach unless excused by impossibility caused by a *force majeure*. Trimarchi demonstrates that in situations where performance is rendered by an enterprise, wealth will be maximized by adopting a strict liability regime.

Trimarchi begins by identifying the effect of strict liability on the distribution of resources. Strict liability for all breaches due to the breacher's lack of organization or due to external causes that are not of a catastrophic nature, forces the breacher to take account of all risks under his control or ability to predict. Using the same symbols as above, the breacher, A, will take any precautions, $\forall$, that satisfy the inequality $a + b - \forall > 0$. These precautions are wealth maximizing except in the extreme case where $a - \forall < 0$, and $b - \exists > a + b - \forall$, so a strict liability regime that accounts for this situation will be wealth maximizing. According to Trimarchi, a strict liability regime with a rule of comparative or contributory negligence, if correctly interpreted and applied, can achieve this result.

Trimarchi notes that strict liability is also efficient because it accounts for downstream losses. Suppose that A supplies the mechanical parts for the construction of machines to B, and B supplies such machines to C, who produces consumer goods. If A breaches its contract with B, then C will also suffer a loss: A will suffer the loss *a*, B will suffer the loss *b*, and C will suffer the loss *c*. A will take precautions against breach if $a + b - \forall > 0$. This is not the socially optimal level of precaution, however, because the total loss resulting from the breach is $a + b + c$. Therefore, the expense $\forall$ should be taken if $a + b + c - \forall > 0$. Therefore, it is necessary to burden A also with the loss *c*, by use of strict liability. Strict liability will impose liability on B for the loss *c*, and A, in turn, is liable to B for $b + c$. Trimarchi is not persuaded that this is excessive liability for A. He notes that, if B's enterprise merged with C's, A would then have to pay $b + c$ damages. Therefore, there is no reason to reduce liability towards independent enterprises which would occur if there was one large, vertically integrated enterprise.

Trimarchi notes that a fault liability regime, based on due diligence, could be wealth maximizing. It is not, however, because the entrepreneur may expect that some economically beneficial precautions will be considered outside the boundaries of due diligence by the judge. Rather, unless the judge compares the costs of the precautions required by due diligence with the precautions required to avoid the total costs of breach illustrated above, wealth will not be maximized under a system of fault liability. Judges, however, simply impose traditional precautions along with those that seem necessary according to common sense and fair dealing. Further, requiring judges to make these calculations imposes substantial dead-weight costs of judicial decision making on society. Consequently, the entrepreneur's incentives to take precautions under a fault liability system will not force him to take wealth-maximizing precautions, and thus Trimarchi concludes that strict liability is a superior foundation for enterprise contractual liability.

Having established his view on the economics of contractual liability, Trimarchi criticizes theories which suggest that the legal or conventional regime of contractual liability has no influence on the distribution of productive resources. For example, Trimarchi criticizes Branca's theory (Scialoja and Branca, 1967 at 94 et seq.), which suggests that distribution of losses caused by a breach is irrelevant, because the parties will modify the contract price depending on the contract rule. Trimarchi argues, on the contrary, that the market would fail to take account of the contract rule quickly, and would fail to supply sufficient information regarding individual suppliers' precautions. Thus, suppliers would have no incentive to take socially efficient precautions.

Similarly, Ronald Coase's article, 'The problem of social cost' (Coase, 1960, pp. 1ff.) suggests that the legal distribution of damages resulting from breach is irrelevant for the optimal distribution of productive resources. If one party can avoid the loss more efficiently than the other, both would be better off if the least-cost avoider accepts liability and the contract price compensates him for taking the risk. As Trimarchi notes, however, the Coase theorem is valid only in perfect markets, specifically markets without transaction costs. Additionally, psychological factors, free riding and asymmetric information may prevent the bargaining process from reaching Coase's efficient outcome.

Notes

1. Unless otherwise stated, all the following quotations are taken from this book.
2. There are two exceptions for this rule contained in articles 2045 and 2047(2) of the Italian Civil Code, however. On the other hand, cases of organized and continuous activity are subject to the strict liability principles. For example, the provisions contained in articles 2049, 2050, 2053 and 2054, subsection 4 of the Italian Civil Code; articles 965 and 978 of the Italian Admiralty Code; and articles 10 subsection 2 and 3 of law No. 1433 of 29 July 1927.

3. Article 2050 of the Italian Civil Code.
4. For example, the method of retribution, the burden of the risk on whoever profits from the enterprise, the criterion of control, the technical competence of the master/servant related to the time and place of activity, the social position of the parties, the duration of the relationship between the parties, the actual continuity of the relationship, and the exclusivity of the relationship.
5. See article 2055 of the Italian Civil Code.
6. See article 2055, subsection 2.

References

Coase, Ronald (1960), 'The problem of social cost', *Journal of Law and Economics*, **3**, 1–44.
Scialoja, A. and G. Branca (1967), 'On the breach of obligations', *Comment on the Civil Code*, Bologna, Rome:
Trimarchi, Pietro (1959), 'Il "case fortuito" quale limite della responsibilità per il danno da cose' ['Fortuitous event' as a limit of the liability for damage caused by things], *Rivestra Trimestrale di Diritto e Procedura Civile*, pp. 808–62.
Trimarchi, Pietro (1961), *Rischio et Responsibilità Uggettiva* (Risk and strict liability), Milan: Giuffrè.
Trimarchi, Pietro (1967), *Causalità e danno* (Causality and damage), Milan: Giuffrè.
Trimarchi, Pietro (1977), 'Sul significo economico dei criteri di responsabilità contrattuale' [On the economics of liability in contracts], *Rivestra Trimestrale di Diritto e Procedura Civile*, pp. 512–31.
Trimarchi Pietro (1987), 'L'analisi economica del diritto: tendenze e prospettive' [Economic analysis of law: tendencies and perspectives], 2 *Quadrimestre*, 563–82.

59 Thorstein Veblen (1857–1929)

Heath Pearson

Introduction

Veblen was born to Norwegian immigrant parents in rural Wisconsin; he was raised there and in Minnesota, and did not become fluent in English until his late teens. He attended Carleton College (BA in philosophy, 1880), Johns Hopkins (no degree) and Yale (PhD in economics, 1884). Thereafter he taught economics at Chicago, Stanford, Missouri and the New School for Social Research. In addition to his scholarly endeavours, he served as editor of the *Journal of Political Economy* and occasionally contributed to such popular periodicals as the *Dial*. Veblen's greatest posthumous fame is as the progenitor of the American school of 'institutional economics', better known to outsiders as 'old institutionalism'. This entry will illuminate those aspects of his thought which bear on the concerns of modern 'law and economics', sometimes termed the 'new institutionalism'. First, we shall show that Veblen's thought was grounded in the same jurisprudential revolution that has issued in the economic analysis of law. Having established this, succeeding sections will stress how he took that common impulse in an exotic direction that has marked him ever since as a pole of heterodoxy.

From natural law to legal realism

Veblen's thought matured in a broader intellectual ferment that predisposed him to the same style of social thought that has motivated law and economics ever since. Had he entered university 30 years earlier, Veblen would no doubt have imbibed a social philosophy grounded in natural law; in other words, he would have been taught to approach the concepts of philosophy, law and political economy as emanations of the coherent logic of the universe, distinct from, or even contradistinct to, worldly interests. In the event, though, his philosophical training included a healthy dose of pragmatism, that novel and distinctively American doctrine which stood four-square against all that natural law represented. Whereas traditional jurisprudence had sought to analyse legal principles by reference to their function in a metaphysical whole, the pragmatist spirit demanded that they be understood as none other than an instrument of man's *purposes*. It was this impulse which led Veblen to argue for 'the adaptation of all human institutions by a process of selective elimination of the economically unfit' (Veblen, 1897, p. 462). Such an assertion would have been shocking not long before Veblen was writing; not long

after, it was commonplace. In large part, it is to Veblen, his teachers and his students that the ascendancy of legal realism is ultimately due.

Legal irrationalism

If Veblen is not remembered as a founder of 'law and economics', and typically he is not, this is because the concern with rational *purpose* was ultimately supplanted by *instinct* and *habit* as the explanatory keys to all institutional structure. To some extent, Veblen's concern to diminish the cognitive status of jurisprudence must reflect his deep-seated antipathy to lawyers, thanks to whose machinations his father had once been forced to forfeit his farm. But there were higher-minded causes as well, most conspicuously the anti-Utilitarian backlash in turn-of-the-century social science. Veblen absorbed this especially through his teacher, William Graham Sumner, through William James, and through his Chicago colleague Jacques Loeb. This insight, combined with Veblen's flair for the voguish terminology of behaviourist psychology, set the stamp on his rhetoric for decades to come:

> For mankind as for the other higher animals, the life of the species is conditioned by the complement of instinctive proclivities and tropismatic aptitudes with which the species is typically endowed. … Human activity, in so far as it can be spoken of as conduct, can never exceed the scope of these instinctive dispositions, by initiative of which man takes action. (Veblen, 1914, p. 1)

This was of central import for legal evolution:

> Any large and persistent change in the material conditions – such, e.g., as has been taking effect in the scale and methods of industry during the past one hundred years – will necessarily be followed in due course by more or less pronounced changes in the established order of human relations and principles of conduct; but it need not follow that the resulting changes in law and morals will be of such a nature as to enhance the facility of life under the new order of material conditions which has induced these changes. … any resulting revision of the principles of conduct will come in as a drift of habituation rather than a dispassionately reasoned adaptation of conduct to the circumstances of the case. (Veblen, 1923, pp. 18–19)

In sum: Veblen's economics was heterodox, and so too was his economic analysis of law bound to be.

It is on account of this very scepticism that Veblen came to define 'institutions' in a fairly novel way, not as explicit rules, but rather as 'settled habits of thought common to the generality of men' (Veblen, 1909, p. 239). (He did, however, allow that ingrained proclivities 'not only become a habitual matter of course, easy and obvious, but they come likewise to be sanctioned by social conventions', Veblen, 1914, p. 7). Whereas the new

institutionalism tends to construe social norms as (i) the result of rational choices, and (ii) effective by constraining rational choices in turn, the Veblenian will rejoin that norms (i) are just as likely to be the precipitates of habit or instinct, and (ii) operate largely by structuring the preferences that guide individual choice. Hence there remains a danger that old (Veblenian) institutionalists and new institutionalists will talk at cross-purposes. With that caveat in mind, let us turn to the evolutionary schema through which Veblen believed law to develop.

Legal evolution

Veblen's conception of institutional change bears comparison to Auguste Comte's, in that social evolution was seen as rooted in psychic evolution through several epochal stages. But Veblen, as might be expected of an economist, insisted that the human mind's development was driven ultimately by the imperatives of livelihood in general, and of technology in particular.

The first stage of material life was that of 'savagery', roughly what we today would term the hunter–gatherer economy. As Veblen outlined in his article, 'The beginnings of ownership', at this undeveloped and undifferentiated stage the right of property (and, we are led to surmise, law as a whole) does not exist: 'such meager belongings of the savage as would under the nomenclature of a later day be classed as personal property are not thought of by him as his property at all; they pertain organically to his person' (Veblen, 1898b, p. 357). This pre-legal state of affairs began to change as 'savagery' gave way to 'barbarian' society, which he associated with the agricultural and pastoral modes of production. It was at this juncture that the use of valuable tools, the accumulation of wealth and the scope for predatory acquisition set the stage for social differentiation between 'honorific' employments (war and its civilian analogues) and 'humiliating' ones. This economic change issued in the psychic urge for 'emulation', by which Veblen meant the desire to demonstrate – or simulate – the status that attends social dominance. The emulative urge, in its turn, ramified in *law*, first and foremost in the principle of chattel, that is the ownership of womenfolk, slaves and, of course, inanimate objects, all as tokens of martial prowess (Veblen, 1898a, p. 59; 1898b, p. 364). The whole Hammurabic Code, indeed, Veblen adduced as bespeaking the ideological revolution wrought by 'militant pastoral culture' (Veblen, 1905, p. 526).

It is worth pausing for a moment to underline, first, the difference between Veblen's account of legal evolution and the now-standard new institutionalist ones. Whereas the latter tend to point to the rational logic of private property over things and persons, Veblen stressed that it was, and remained, 'a cultural fact and has to be learned; it is a cultural fact which has grown into an

institution in the past through a long course of habituation, and which is transmitted from generation to generation as all cultural factors are' (Veblen, 1898b, p. 360). Second, Veblen believed that the barbarian, 'emulative' impulse had survived down to his own day as a hardy vestige, despite the fact that the economic forms which once nourished it had long since expired. This held particularly true of lawyers themselves:

> [N]o taint of usefulness, for other than the competitive purpose, attaches to the lawyer's trade … The lawyer is exclusively occupied with the details of predatory fraud, either in achieving or in checkmating chicane, and success in the profession is therefore accepted as marking a large endowment of that barbarian astuteness which has always commanded men's respect and fear. (Veblen, 1899, p. 231).

Here, it seems, Veblen had found a theoretical apparatus upon which to hang his long-standing distaste for jurists and their guild.

A second epoch-making change in mentality and law accompanied the incipient industrialization of early modern Europe. Regularization of work effort, a more precise division of labour and at least a vague familiarity with the achievements of the scientific revolution instilled in engineers, operatives and merchants a heightened sense of the underlying mechanical unity of the universe. This new mind-set, which we might term 'Newtonian' (Veblen himself called it 'supernatural'), found its basic legal expression in the doctrine of natural rights, which 'grew up and found secure lodgment in the common sense of the community, as well as with its lawgivers and courts, under the discipline of the small industry and petty trade ("domestic industry") whose development culminated in the eighteenth century' (Veblen, 1904, p. 270; 1923, pp. 40–49). In the realm of constitutional law, the mind-set ushered in the age of democratic rule to both England and France.

Veblen believed that the historical conjuncture through which he himself was living was fairly intricate. The regnant principles of private law remained those of Enlightenment natural rights doctrine, but its practice was dominated by attorneys who, as we have seen, were avatars of a more barbaric age. Similarly, the Western powers had seen industrialization and constitutional democracy reinforce one another, but in the new century they now faced a rising German state in which industrialization had been effectively turned to the purposes of a predatory autocracy (Veblen, 1915). And as if this were not complex enough, it was currently becoming clear that the technical and organizational implications of the Second Industrial Revolution were changing the mind-set of the workforce yet again, towards 'mechanistic' habits of thought. As he put it, 'the discipline of daily life, from which the common-sense notions of the vulgar are in good part derived, is no longer in full accord with the natural-rights conceptions handed down from the eighteenth century' (Veblen, 1904, pp. 282–3). Patriarchal monogamy ('ownership-mar-

riage') was already in crisis among the industrial proletariat (Veblen, 1898a, pp. 59–60); the doctrine of absolute freedom of contract had been in decline for some time ('obsolescent, of course, not in point of law, but in point of fact', Veblen, 1904, pp. 274–5); and it was only a matter of time before the engineers and their allies declared war on the outmoded conceptions of private property that served only the pecuniary interests of capital (Veblen, 1917, pp. 363–6).

Though he was far too delphic to say it unambiguously, the clear implication of Veblen's evolutionary scheme was that the future, including the future of law, belonged to engineers and administrators. Technocracy, or rule by 'a Soviet of technicians', would ensure that all social elements were combined scientifically and efficiently. With power and reason finally yoked together, law itself – as a system of constraints binding the knowledgeable and the ignorant alike – could be given a decent burial.

Conclusion

Veblen died in 1929, long before the economic analysis of law took on anything like the scale and shape it knows today, and it is difficult to estimate his influence on its later developments. A handful of practitioners in the generation immediately succeeding Veblen – including Walton Hamilton, Thurman Arnold and Robert Lee Hale – acknowledged his influence on their work; but the direct legacy seems to have been broken there. If there has been a Veblenian factor in the latter-day evolution of the movement, its efficacy has been more diffuse. On the one hand, Veblen may be seen as a prophet, one who did not live to see the promised land of 'law and economics', but who performed a vital function in demystifying a profession that tended to take itself altogether too seriously. On the other hand, he may plausibly be blamed for having retarded the field's advance with his intemperance, the absurdist excesses of his rhetoric, and the all-round opacity of his writing style. If the old institutionalism and the new institutionalism have missed some opportunities for fruitful cross-fertilization, perhaps Veblen must bear some responsibility.

Bibliography

Works by Veblen

1897 'Review of Antonio Labriola's *Essais sur la conception matérialiste de l'histoire*', *Journal of Political Economy*, **5**, 390–91; reprinted in Veblen (1973, pp. 461–3).

1898a 'The barbarian status of women', *American Journal of Sociology*, **4**, 503–14; reprinted in Veblen (1934, pp. 50–64).

1898b 'The beginnings of ownership', *American Journal of Sociology*, **4**, 352–65.

1899 *The Theory of the Leisure Class: An Economic Study of the Evolution of Institutions*, New York: Macmillan.

1904 *The Theory of Business Enterprise*, New York: Scribner's.

1905 'Review of Robert Francis Harper's *The Code of Hammurabi*', *Journal of Political Economy*, **13**, 319–20; reprinted in Veblen (1973, pp. 525–7).
1909 'The limits of marginal utility', *Journal of Political Economy*, **17**; reprinted in Veblen (1919, pp. 231–51).
1914 *The Instinct of Workmanship and the State of the Industrial Arts*, New York: Macmillan.
1915 *Imperial Germany and the Industrial Revolution*, New York: Macmillan.
1917 *An Inquiry into the Nature of Peace and the Terms of Its Perpetuation*, New York: Macmillan.
1919 *The Place of Science in Modern Civilization and Other Essays*, New York: Huebsch.
1923 *Absentee Ownership and Business Enterprise in Recent Times: The Case of America*, New York: Huebsch.
1934 *Essays in Our Changing Order* (ed. Leon Ardzrooni), New York: Augustus M. Kelley.
1973 *Essays, Reviews and Reports* (ed. Joseph Dorfman), New York: Augustus M. Kelley.

Further reading

Riesman, David (1953), *Thorstein Veblen: A Critical Interpretation*, New York: Scribner's.
Rutherford, Malcolm (1994), *Institutions in Economics: The Old and the New Institutionalism*, Cambridge: Cambridge University Press.

60 Max Weber (1864–1920)

Helge Peukert

The man, his life and sociology

Max Weber was born in 1864 in Erfurt (Thüringen) and died in Munich in 1920. His father came from a family of industrialists and tradespeople. He was a lawyer (and after 1866 became a city adviser in Berlin) and without doubt stimulated his son's early studies in the history of commercial law and his emergence as one of the major personalities in a new generation of historical political economists in Germany in the 1890s. In 1892, he became extraordinary professor in commercial and German law at Berlin University. In 1894, a switch from law to economics took place: he was appointed to a chair in political economy at Freiburg, the town where, in 1882, he had begun to study law, economics, philosophy and some theology; his special interests as a student were already history of late antiquity, modern commercial law and contemporary history of constitutional law.

Although Weber is mainly considered as a founding father of sociology (a term he began to use not long before the 1910s), his writings deal with the interpenetration of law, economy and society. Turner and Factor (1994) put forward an interpretation of Weber as being mainly a translator of Rudolf von Ihering's legal philosophy into sociology, see also Loos (1970), Breuer and Treiber (1984), Rehbinder and Tieck (1987), Zippelius (1991), Marra (1992), and the introductions by Rheinstein (1954a) and Winckelmann (1960); a general overview about the debate on Weber in the secondary literature has been put together by Hamilton (1991); as a valuable introduction, see Käsler (1988).

After his retreat from academic teaching (resuming full-time teaching as professor in Munich, in 1919 only shortly before his death) owing to a longlasting nervous breakdown beginning in 1898, his interdisciplinary orientation can also be found in his editorship of the *Archiv für Sozialwissenschaft und Sozialpolitik* and his further publications. Weber can be considered as one of the last universalists in the social sciences, trying to integrate law, economy, politics, culture and what later would be called 'social system' by T. Parsons, the translator of his best-known book, *The Protestant Ethic* (1985). Although criticizing the older and younger Historical school (Karl Knies, Wilhelm Roscher and Gustav von Schmoller) his approach and methodology can best be described as belonging, like Werner Sombart's, to the youngest historical school, where comparative historical inquiries are at the core of research.

The speciality of Weber lies in his reference to historical research (he never plunged into the upcoming literature of neoclassical economics) and a concept of understanding (comparable to today's hermeneutical approaches: see Lavoie, 1991), but at the same time he tried to integrate understanding and explaining (for example, by the application of statistical methods). At the beginning of his opus magnum posthumum, *Economy and Society* (1978), he therefore defines sociology as

> a science concerning itself with the interpretive understanding of social action and thereby with a causal explanation of its course and consequences. We shall speak of 'action' insofar as the acting individual attaches a subjective meaning to his behavior – be it overt or covert … Action is 'social' insofar as its *subjective meaning takes account of the behavior of others* and is thereby oriented in its course. (Weber, 1978, p. 4, emphasis added; this part was written from 1918 to 1920)

An analysis of the understanding paradigm in the German Historical school, in Austrian economics and in American 'old' institutionalism has been undertaken in Peukert (1997).

As in the historical approach, understanding involves the interpretative grasp of the actually intended meaning of concrete individual action. But Weber continues, 'or (b) as in cases of sociological mass phenomena, the average of, or an approximation to, the actually intended meaning; or (c) the meaning appropriate to a scientifically formulated pure type (as ideal type) of a common phenomenon' (Weber, 1978, p. 9).

Weber's method of understanding therefore goes far beyond a mere understanding of concrete, individual meaningful actions. It does not investigate objectively correct or metaphysical true, but rationally meaningful, constructions by the social scientist (ideal types), that is, an instrumentally rational ('*zweckrational*') reconstruction of actions excluding errors or emotional factors, if actions were completely directed to a single end or a set of coherent ends.

Weber distinguishes four non-reducible, possible orientations of action:

> (1) *instrumentally rational* (*zweckrational*), that is, determined by expectations as to the behavior of objects in the environment and of other human beings; these expectations are uses as 'conditions' or 'means' for the attainment of the actor's own rationally pursued and calculated ends; (2) *value rational* (*wertrational*), that is, determined by a conscious belief in the value for its own sake of some ethical, aesthetic, religious, or other form of behavior, independently of its prospects of success; (3) *affectual* (especially emotional), that is, determined by the actor's specific affects and feeling states; (4) *traditional*, that is, determined by ingrained habituation. (Ibid., pp. 24–5)

Historically, the universal process of rationalization can be described in terms of these four pattern variables, but structures can also be analysed vertically with them: a society is composed of different levels of social relationships which can be described according to their relative mixture of the four action types along the axis traditional/affectual, value rational and finally instrumentally rational, like usage ('*Brauch*'), including fashion ('*Mode*'), and custom ('*Sitte*'), which have to be distinguished from convention ('*Konvention*') and law ('*Recht*'), the last category having the highest degree of instrumental rationality. On a micro level, Weber then analyses social relationships such as communal and associative, open and closed, voluntary and compulsory associations.

Furthermore, the action types are used to demarcate four ideal types of legitimate political order. According to Weber, there are no relatively stable political units without a certain belief in their existence and legitimacy, transforming power relations in domination ('*Herrschaft*'), characterized by the probability that a command with a specific content will be obeyed and accepted by a group of persons. 'The actors may ascribe legitimacy to a social order by virtue of: (a) *tradition*: valid is that which has always been; (b) *affectual*, especially emotional, *faith*: valid is that which is newly revealed or exemplary; (c) *value-rational* faith: valid is that which has been deduced as an absolute; (d) positive enactment which is believed to be *legal*' (ibid., p. 36). Three ideal type legitimate orders are distinguished: the rational, the traditional and the charismatic order (divided, for example, into hereditary and charisma of office). Charismatic order, first based on devotion to the exceptional sanctity, heroism or exemplary character of an individual person, consequently leads to a traditional or legal order; that is, a routinization of charisma takes place; the administrative staff and his followers develop vested interests, the necessity of the economy as a going regular concern, the problem of succession and so on.

Weber's underlying world view as a universal historian was a tragic one: affectual charismatic new beginnings always fade into bureaucratic routines or, put more generally, the increasing instrumental rationality of the modern world comes at the cost of affectual anthropological impulses being left out, an iron cage opening up and the loss of liberty occurring as the result of mechanization, the increase in formal rules and procedures, and an instrumental efficiency without further general value directions, and the privatization of the questions of ultimate concern in the face of the disenchantment of the world take place. For Weber, the adoption of a complex technical division of labour and a hierarchical structure of administration fostered an overarching bureaucratic mode of organization (and, Weber would add, a new method of domination) which is the most characteristic trait in the modernization process from the traditional to the rationalized pattern. It dominates even in the

political sphere, resembling Joseph Schumpeter's and Anthony Downs's theories of competitive leadership democracy, according to which the leaders of mass parties and their bureaucratic machines compete for support, temporarily deossified by creative individuals at the top, the new charismatic leaders.

As mentioned, Weber tried to integrate the concepts of understanding and explaining, that is the level of meaning ('*Sinnadäquanz*') and causal adequacy ('*Kausaladäquanz*'). Sociology should also be an empirical science, transcending a distinction between natural (empirical testability) and social sciences (or '*Geisteswissenschaften*'), made, for example, by Wilhelm Dilthey and his concept of vague intuition. His attempt to mediate tried to take into account the crisis of historicism, the critique of (logical) positivism, a lost sense in history, the debate on value judgements and the liberation of these questions from a narrow economic science background. A causal explanation depends on the probability of an observable set of action clusters; sociological generalizations like ideal-type constructions therefore can be tested as statistical uniformities. For Weber, between a hermeneutical concept of the social sciences and a more empirical/statistical/econometric/fallibilistic approach a complementary relationship exists which also cross-cuts the different schools in economic science. Weber even mentions his first example of a hermeneutical constructivist ideal-type method:

> [T]he concepts and 'laws' of pure economic theory are examples of this kind of ideal type. They state what course a given type of human action would take if it were strictly rational, unaffected by errors or emotional factors, and if, furthermore, it were completely and unequivocally directed to a single end, the maximization of economic advantage. In reality, action takes exactly this course only in unusual cases, as sometimes on the stock exchange; and even then there is usually only an approximation of the ideal type. (Ibid., p. 9; for a more radical understanding paradigm of the stock exchange, see Soros, 1994)

Although never discussing more formal neoclassical approaches in detail, and criticizing their ahistorical universalizing aspirations on an epistemological level, Weber accepts and interprets their research as ideal type reasoning. He probably would have criticized a certain 'fallacy of misplaced concreteness' (Alfred V. Whitehead), that is, taking constructivist ideal types as quasi-empirical real types and deriving practical policy conclusions from them. The question whether behaviour can be identified as 'strictly rational' outside very narrow confines of a static maximization under given constraints will not be discussed here.

Weber refers to a strict standard of value neutrality in economic research. This was one of the reasons why he was critical of the *Verein für Socialpolitik* and became co-founder of the *Deutsche Gesellschaft für Soziologie* in Berlin in 1909, and which he left in 1912 for the same reasons – the problem and

principle of *Wertfreiheit* – which did not prevent him from making utterances as a natural *zoon politicon*; at least he tried to separate academic analysis and political practice and zeal. In 1895, he gave his inaugural academic lecture on 'The national state and economic policy'. He accepted *Realpolitik* and German imperialism and supported the First World War with national enthusiasm and as a volunteer, quickly criticizing the German annexation policy and the harsh submarine war. But even in 1918 he was still regarding a constitutional monarchy as ideal and possible. He joined the Deutsche Demokratische Partei (DDP) of Friedrich Naumann and considered constitutional–parliamentary reform as necessary, but he never became a republican of the heart and was strictly opposed to the left-wing workers', soldiers' and peasants' councils. His position can be described as that of a more or less progressive national–liberal. He was opposed to the supposed paternalism of the socialists of the chair and the one-sided assumptions of class conflict in Marxism (Weber thought that status group cleavages are important too), but his writings can also be regarded as a critique of the ahistorical theorizing of Carl Menger and his Austrian school and neoclassical economics in general and perhaps of their untrammelled support of *laissez-faire* capitalism and the spread of market values into all areas of life.

In Weber's opinion the strictly goal-oriented, instrumentally calculating, welfare-maximizing behaviour of individuals in neoclassical theories was not a universal characteristic of human rationality as such, but a product of modern Western rationalism. He shared the critique of socialist planning of Ludwig von Mises: an efficient system of allocation is impossible without markets indicating scarcity. In conjunction with his thesis of the spread of bureaucracy, he saw socialist planning mainly as another force to increase the degree of bureaucratic serfdom and stagnation. The hitherto separate economic, legal and political hierarchies would be fused and the dictatorship of the official and not of the worker would ensue – a prediction which came true in the former 'socialist' countries (Nove, 1990).

With the Historical school, Weber shared the main research topic: the explanation of modern Western capitalism or what Weber called 'occidental rationalism', where capitalism is only one facet of a more general rationalization process. His sociology of religion, that is the study of the major world religions like Hinduism, Confucianism, Christianity, Buddhism and ancient Judaism, came to be understood as an implicit critique of basic neoclassical preconceptions (Veblen, 1919) in underlining the distinctiveness and specificity of economic action in different cultural settings (compare Edgar Salin's dictum that different economic stages need different economic theories).

> [His studies] showed that, while instrumental rationality was a universal category of social action, only in the modern West had the goal-maximizing calculation of

> the most efficient means to given ends become generalized. And while other cultures had attempted to make the world intelligible through the development of elaborate theodicies, or to create internally consistent systems of ethics or law, the distinctive feature of western rationalism was the scientific assumption that all things could be comprehended by reason, together with the attitude of practical mastery which sought to subject the world to human control rather than merely adjust oneself to it. (Beetham, 1994, p. 887)

Weber (1985) first developed his conception in his most famous book about the spirit of capitalism, written in 1904/5. As an empirical starting point he mentions that, among entrepreneurs, a disproportionate percentage of Protestants (mainly Calvinists, Pietists and Methodists) can be found. According to Weber, this is not due to a weakening of religious bonds, but to the contrary: a systematic increase in religious demands in everyday life as innerworldly asceticism took place, furthermore it supported the specific orientation of non-traditional economic rationalism which spread in the Netherlands, France and England in the sixteenth and seventeenth centuries. Capitalism, the profit motive, existed everywhere, but the specific trait was a new ethos in a disenchanted world.

Analysing theological writings (for example those of Baxter), Weber tried to demonstrate the connection between religious beliefs and ideas, between the puritan idea of calling and asceticism and the maxims of everyday life, that is the capitalist spirit. The lonely individual in a disenchanted cosmos tries to find out if he or she belongs to those chosen by God (predestination in Calvinism). The successful accumulation of wealth was interpreted as an indicator of being chosen:

> The worldly Protestant asceticism … acted powerfully against the spontaneous enjoyment of possessions; it restricted consumption … On the other hand, it had the psychological effect of freeing the acquisition of goods from the inhibitions of traditionalistic ethics … What was condemned as covetousness … was the pursuit of riches for their own sake. For wealth in itself was a temptation … the inevitable practical result is obvious: accumulation of capital through ascetic compulsion to save. (Weber, 1985, pp. 170–72)

Weber on law and economics

Decoupled from its original religious content, worldly asceticism became a part of modern civilization:

> For when asceticism was carried out of monastic cells into everyday life, and began to dominate worldly morality, it did its part in building the tremendous cosmos of the modern economic order. This order is now bound to the technical and economic conditions of machine production which today determine the lives of all the individuals who are born into that mechanism … Since asceticism undertook to remodel the world and to work out its ideals in the world, material

> goods have gained an increasing and finally an inexorable power over the lives of men as at no previous period in history. Today the spirit of religious asceticism … has escaped from the cage. But victorious capitalism, since it rests on mechanical foundations, needs its support no longer. (Weber, 1985, pp. 181–2)

Weber knew that the influence of ideas on the economy is only one side of the medal, asserting that 'I consider the influence of economic development on the fate of religious ideas to be very important' (ibid., p. 277, n. 84).

The striving for maximum profit is not the main characteristic of capitalism: it can be found in waiters, robbers, corruptible officials and so on. Capitalism is based on the exploitation of possibilities of exchange, of formally peaceful gain, sometimes presupposing the taming of the irrational instinct of the rampant desire to acquire. Typical of Western capitalism are rational bookkeeping, substantive freedom of contract, mechanically rational technology, market freedom, formal rational order of the monetary system, the separation of household and business, rational business forms which are oriented towards the possibilities of the consumer market, the capitalistic organization of formally free work and – as already mentioned – the disposition of people to lead their lives in a certain systematic and rational way. Weber's economic sociology was not intended to be a new fully-fledged economic theory, he only tried to develop some basic categories, such as money as chartal and natural. Weber emphasizes the motives of the actors, such as material interests, the transition from traditional to rational orientations and the impersonality of the market. He shows the implications of capital accounting in the sense of shop discipline and appropriation of the means of production as a system of domination and he discusses the appropriation of the workers. Weber underlines the specific occidental form of capitalism proper and the overlap of value spheres; the modern rational enterprise capitalism is, among other factors, highly dependent on the rational and calculable law. Both spheres underlie the trend to formal rationalization with a tendency towards material irrationality, highlighting once again Weber's tragical secular world view:

> The term 'formal rationality of economic action' will be used to designate the extent of quantitative calculation or accounting which is technically possible and which is actually applied. The 'substantive rationality'… is the degree to which the provisioning of given groups of persons … with goods is shaped by economically oriented social action under some criterion (past, present, or potential) of ultimate values (*wertende Postulate*), regardless of the nature of these ends. (Weber, 1978, p. 85)

Formal rationality depends on its capability of being expressed in numerical, calculable terms; substantive rationality depends on the possibility that economic results are measured on a scale of value rationality or substantive goal orientation.

Weber's somehow neglected sociology of law (see Weber, 1978, pp. 641–900; on the interrelationship of legal and economic order, see ibid., pp. 311–38) deals with the interrelationship of law, religion, the economy and politics and its internal antinomies and contradictions. In an intercultural (from the most 'primitive' societies to the world's greatest civilizations) and diachronical way (from magical and irrational forms to the rational law techniques of today) he develops some general typological terms to understand the empirical validity of law as social conduct in a legitimate order and its likelihood of being obeyed, including its social holders, the impact of material interests of social groups, the influence of law education and political institutions as a source of law enforcement and so on in a comparative perspective without making any simple cause-and-effect assumptions (see, for the striking parallels between his sociology of law and religion, Treiber, 1984).

Already in his early dissertation about the history of trading companies in the Middle Ages (Weber, 1924, pp. 312–443) and in his *habilitation* on Roman agricultural history and its significance in public and private law (1891; see also the introductions in Weber, 1984, 1986, 1993), Weber focused on the constant interrelations with political, economic, legal and religious developments (such as the influence of the Church on legal thinking – he argued that Jewish–Christian religions foster a higher degree of rational thinking and institutionalization) until his last theoretical contributions such as the *Wirtschaftsgeschichte* (1958, first published in 1923), understanding sociology as the science of society in general.

Weber asks 'What *actually* happens in a group owing to the *probability* that persons engaged in social action (*Gemeinschaftshandeln*) especially those exerting a socially relevant amount of power, subjectively consider certain norms as valid and practically act according to them, in other words, orient their own conduct towards these norms?' (Weber, 1978, p. 311). Law is distinguished from custom and conventions by a specific coercive apparatus which enforces the legal rules, the legal coercion by violence as the monopoly of the state only being a specific modern case. 'The empirical validity of a norm as a legal norm affects the interests of an individual in many respects. In particular, it may convey to an individual certain calculable chances of having economic goods available or of acquiring them under certain conditions in the future' (ibid., p. 315). For Weber, the originating form of law has to be seen in contracts (first being status contracts) which are also essential for the exchange mechanism in modern, money-based industrial capitalism, which needs (in a certain contrast to trade capitalism) the calculable chance of economic exchange guaranteed by legal rules.

The relationship between law and economics is not an easy and causal one, but is characterized by relative autonomy, mutual indifference and functional complementarities and alternatives:

> (1) Law … guarantees by no means only economic interests but rather the most diverse interests ranging from the most elementary one of protection … to such purely ideal goods as personal honor … (2) Under certain conditions a 'legal order' can remain unchanged while economic relations are undergoing a radical transformation … (3) The legal status of a matter may be basically different according to the point of view of the legal system from which it is considered. But such differences … need not have any relevant economic consequences … (4) Obviously, legal guarantees are directly at the service of economic interests to a very large extent … (5) Only a limited measure of success can be attained through the threat of coercion supporting the legal order. This applies especially to the economic sphere … (6) From the purely theoretical point of view, legal guaranty *by the state* is not indispensable to any basic economic phenomenon … But an economic system, especially of the modern type, could certainly not exist without a legal order with very special features which could not develop except in the frame of a public legal order … The universal predominance of the market consociation requires on the one hand a legal system, the functioning of which is *calculable* in accordance with rational rules. On the other hand, the constant expansion of the market … has favored the monopolization and regulation of all 'legitimate' coercive power by *one* universalist coercive institution through the disintegration of all particularist status-determined and other coercive structures which have been resting mainly on economic monopolies. (Ibid., pp. 333–7)

In accordance with his evolutionary minimal programme, he develops comparative stages and general tendencies. Like his three types of dominance (charismatic, traditional and rational) he distinguishes

> the general development of law and procedure … as passing through the following stages: first charismatic legal revelation through 'law prophets'; second, empirical creation and finding of law by legal honoratiores, i.e., law creation through cautelary jurisprudence and adherence to precedent; third, imposition of law by secular or theocratic powers; fourth and finally, systematic elaboration of law and professionalized administration of justice by persons who have received their legal training in a learned and formally logical manner. (Ibid., p. 882)

To analyse lawmaking and lawfinding in an evolutionary and comparative perspective, Weber used as main categories the differences between rational/irrational and formal/informal (the latter referring to the technical apparatus of legal practice). The historical evolution went (according to the reconstruction by Schluchter, 1979, pp. 129–31) from material–irrational law (traditional law) to the formal–irrational law (reveiled law), followed by material–rational law (derived law from 'out of law' perspectives such as ethics) and led finally to the formal–rational law (modern continental legal law), which is most compatible with rational capitalism.

> A body of law can be 'rational' in several different senses, depending on which of several possible courses legal thinking takes toward rationalization. Let us begin with the seemingly most elementary thought process, viz. generalization, i.e., in

> our case, the reduction of the reasons relevant in the decision of concrete individual cases … 'systematization', which has never appeared except in late stages of legal modes of thought … it represents an integration of all analytically derived legal propositions in such a way that they constitute a logically clear, internally consistent, and, at least, in theory, gapless system of rules, under which, it is implied, all conceivable fact situations must be capable of being logically subsumed lest their order lack an effective guaranty … Both lawmaking and lawfinding may be either rational or irrational. They are *formally* irrational when one applies in lawmaking or lawfinding means which cannot be controlled by the intellect, for instance when recourse is had to oracles or substitutes therefor … [They] are *substantially* irrational on the other hand to the extent that decision is influenced by concrete factors of the particular case as evaluated upon an ethical, emotional, or political basis rather than by general norms … Law … is 'formal' to the extent that, in both substantive and procedural matters, only unambiguous general characteristics of the facts of the case are taken into account. (Weber 1978, pp. 655–7)

Leaving out Weber's tragic view that formal rationalization undermines from a certain phase its own rationality and the relapses of modern law into substantive concepts (rematerialization), the more formal rational law is, the more it is compatible with capitalism on an ideal-type level of reasoning. But the judge-centred practical training and special case-oriented common law of Great Britain belongs to the substantial irrational type, although Great Britain, where the attorneys were organized like guilds which could dominate and fill the law courts, was the first country to industrialize. The academic models of formal general law have been developed in more backward countries (perhaps as a legal attempt to foster the missing political and economic unification?) like Germany, where conceptual jurisprudence (*Begriffsjurisprudenz*) which strongly influenced Weber dominated in the nineteenth century.

British common law shows that logical formalism is not the only compatible form of law in modern exchange economies, but Weber nevertheless believed in the greatest congeniality ('*Wahlverwandschaft*') between modern economies and societies and formal rational law. Some criticism has been raised against his opinion. One point is that he overstated the difference between the common law and the academic conceptual jurisprudence because case law is based on general principles too (making a case as a comparable variant of something: see Rheinstein, 1954b, p. lv) and that the academic model contains manifold elements of 'irrationality' (like the lay courts' decision-finding process). So the ideal-type character of the distinction should not be forgotten.

On the other hand, Rehbinder (1987) argues that Weber even misunderstood the signals of his time in social and labour law, arguing (somewhat paradoxically as a founder of sociology) against a sociological jurisprudence. He sees the prime function of the lawyer to be a guardian of formal contract law, on which Weber concentrated one-sidedly. He did not recognize the

importance of the law of the welfare state and that manifold inclusions of substantive reasonings like the antiformal social law have not undermined the inner logic of law and its functioning in modern society, although it has passed over the liberal–formal, contract-centred approach of a mere deducing according to systems of notions. Weber's neo-Kantian epistemological viewpoint on the purity of method in different social sciences led him to believe that, for example, the law of the *Sozialstaat* as a non-formal substantive orientation is a step backwards and not forwards in the process of internal rationalization. Probably he was wrong with this assumption, but on the other hand it illustrates that the process of rationalization may be more heterogeneous and contradictory then even he himself thought.

References

Beetham. D. (1994), 'Weber, Max (1864–1920)', in J. Eatwell, M. Milgate and P. Newman (eds), *The New Palgrave*, London and Basingstoke: Macmillan, pp. 886–8.

Breuer, S. and H. Treiber (eds) (1984), *Zur Rechtssoziologie Max Webers*, Opladen: Westdeutscher Verlag.

Hamilton, P. (1991), *Max Weber: Critical Assessments*, 8 vols, London and New York: Routledge.

Käsler, D. (1988), *Max Weber: An Introduction*, London: Polity Press.

Lavoie, D. (ed.) (1991), *Economics and Hermeneutics*, London and New York: Routledge.

Loos, F. (1970), *Zur Wert- und Rechtslehre Max Webers*, Tübingen: Mohr.

Marra, R. (1992), *Dalla comitá al diritto moderno*, Turin: Giappichelli.

Nove, A. (1990), *An Economic History of the USSR 1917–1991*, London: Penguin Books.

Peukert, H. (1997), *Das Handlungsparadigma in der Nationalökonomie*, Marburg: Metropolis.

Rehbinder, M. (1987), 'Max Weber und die Rechtswissenschaft', in Rehbinder and Tieck (eds), pp. 127–49.

Rehbinder, M. and K.-P. Tieck (eds) (1987), *Max Weber als Rechtssoziologe*, Berlin: Duncker & Humblot.

Rheinstein, M. (1954a), 'Introduction', in Rheinstein (ed.), *Max Weber on Law in Economy and Society*, Cambridge, MA: Harvard University Press, pp. xxv–lxxii.

Rheinstein, M. (ed.) (1954b), *Max Weber on Law in Economy and Society*, Cambridge, MA: Harvard University Press.

Schluchter, W. (1979), *Die Entwicklung des okzidentalen Rationalismus*, Tübingen: Mohr.

Soros, G. (1994), *The Alchemy of Finance*, New York: John Wiley & Sons.

Treiber, H. (1984), '"Wahlverwandschaften" zwischen Webers Religions- und Rechtssoziologie', in Breuer and Treiber (eds), pp. 6–68.

Turner, S.P. and R.A. Factor (1994), *Max Weber: The Lawyer as Social Thinker*, London and New York: Routledge.

Veblen, T. (1919), *The Place of Science in Modern Civilization and Other Essays*, New York: Huebsch.

Weber, M. (1891), *Die römische Agrargeschichte in ihrer Bedeuting für das Staats- und Privatrecht*, Stuttgart: Enke.

Weber, M. (1924), *Gesammelte Aufsätze zur Sozial- und Wirtschaftsgeschichte*, Tübingen: Mohr.

Weber, M. (1958), *Wirtschaftsgeschichte*, Berlin: Duncker & Humblot.

Weber, M. (1978), *Economy and Society*, Berkeley: University of California Press.

Weber, M. (1984), *Die Lage der Landarbeiter im ostelbischen Deutschland*, ed. M. Riesebrodt, *Max Weber Gesamtausgabe*, I/3, Tübingen: Mohr.

Weber, M. (1985), *The Protestant Ethic and the Spirit of Capitalism*, trans. T. Parsons, London: Counterpoint.

Weber, M. (1986), *Die römische Agrargeschichte in ihrer Bedeutung für das Staats- und Privatrecht*, ed. J. Deininger, *Max Weber Gesamtausgabe*, I/2, Tübingen: Mohr.
Weber, M. (1993), *Landarbeiterfrage, Nationalstaat und Volkswirtschaftspolitik*, ed. W.J. Mommsen and R. Aldenhoff, *Max Weber Gesamtausgabe*, I/4, 2 vols, Tübingen: Mohr.
Winckelmann, J. (ed.) (1960), *Max Webers Rechtssoziologie*, Neuwied: Luchterhand.
Zippelius, R. (1991), *Grundbegriffe der Rechts- und Staatssoziologie*, Munich: Beck.

61 Christian Wolff (1679–1754)

Wolfgang Drechsler

Introduction

Christian Wolff is the most eminent German philosopher between Leibniz and Kant. His main achievement is a complete set of work on practically any scholarly subject of his time, displayed and unfolded according to his demonstrative–deductive, mathematical method, which perhaps represents the peak of Enlightenment rationality in Germany. Wolff is also the creator of German as the language of scholarly instruction and research, although he published also in Latin, so that an international audience could, and did, read him. A founding father of, *inter alia*, economics and public administration as academic disciplines, he concentrated especially in these fields on advice, on practical matters for members of government, and on the professional nature of university education. Although he was a quintessentially continental thinker, both in form and in content, his work is said to have had a strong impact even on the American Declaration of Independence (Goebel, 1920).

Biography[1]

Wolff was born on 24 January 1679 in Breslau, Silesia. Coming from a modest background, he received his degrees from the University of Leipzig and spent his entire life as a university professor of mathematics, sciences, philosophy and, later, public law at the Universities of Halle, Prussia (1706–23 and 1740–54) and Marburg, Hesse-Cassel (1723–40).

In 1723, Wolff was ousted from his first chair at Halle in one of the most celebrated academic dramas of the eighteenth century. By decree of the King of Prussia, initiated by the pietist Divinity Faculty, which saw in Wolff a dangerous rival and enemy of the faith, Wolff had to leave Prussia within 48 hours or be hanged. The immediate cause had been his farewell address as vice-rector in 1721 (Wolff, 1985), in which Wolff describes Confucianism as ethically rather admirable.

Wolff immediately found refuge at the Hessian university of Marburg. As one of the most popular and fashionable university teachers in Europe, he increased matriculation figures within five years by about 50 per cent. However, he returned to Halle in 1740 after the accession of the new King of Prussia and admirer of his, Frederick II (the Great).

When Wolff died on 9 April 1754, he was a very wealthy man, almost entirely as a result of his income from lecture fees, salaries and royalties. He

was also a member of many academies and probably the first scholar to have been created hereditary Baron of the Holy Roman Empire on the basis of his academic work. His school, the Wolffians – the first 'school' in our sense that any German philosopher had – dominated Germany until the rise of Kantianism.

Law and economics

Wolff himself was not a lawyer in our sense; after 1740, he held a chair of public and international law, but this was quite different from 'positive' (material) law which corresponds to the twentieth-century understanding of the term. His direct contribution to (positive) law is slight (but see Nettelbladt, 1754); his indirect influence via his legal students, especially Johann Ulrich von Cramer (see Cramer, 1742–67; Drechsler, 1996), is much more important, but neither appears to be particularly focused on law and economics. Rather, Wolff's importance for this area, as has very recently been argued, lies elsewhere.

In what constitutes the key essay on Wolff's significance for law and economics, Jürgen Backhaus (1997) has said that 'Wolff conceived of both law and economics still largely as one discipline and therefore was able to integrate naturally what has today to be integrated conscientiously, and with effort' (p. 129). Backhaus goes on to argue:

> [I]n economic policy analysis[,] next to problems of allocative efficiency, issues of distributive justice or equity play an important role. However, these economic discussions of equity issues typically do not even refer to the legal framework in which they ultimately have to find their application. This creates the possibility that the results of an analysis of welfare economics can be in direct conflict with legal realities. Ultimately, the latter prevail and the ignorance of the legal framework in which economic policy needs to be implemented leads to the impossibility of applying the results of economic theoretical work in practice. Wolff already talked about this problem which plagues modern economics to the extremes, and he insisted on the need to do the legal and economic analysis *in tandem*. (p. 133)

The aim of the state is, for Wolff, the self-sufficiency of individual houses, or households. This seems to go both against all classical economic theory, which focuses on a smaller unit (the individual) and against continental *Nationalökonomie*, which focuses on the state, but Wolff's approach of considering the household, whence the term 'economics' stems in the first place,[2] allows him to develop a model of the welfare state, including a sophisticated approach to subsidiarity (see Backhaus, 1997, pp. 135–9). This model is all the more important today because the built-in subsidiarity approach is one of the few effective insurances against the overinvolved state which any welfare state, to which there seems to be no realistic alternative in Western countries, tends to become (see Drechsler, 1995, pp. 227, 230; 1997b).

> Public policy is exclusively defined in terms of social welfare, and it is limited by the ability of the households to take care of their own affairs. The primary task of the welfare state is the support of the households in such a way that they can take care of their own matters in finding the means for their existence. … For this purpose, Wolff already argues for the modern economic policies of full employment, the stability of the currency and even environmental policy. (Backhaus, 1997, p. 134)[3]

The possibility of using the implications of Wolff's 'pure' philosophy as a countermodel to (neo-)classical economics has been raised by Reinert and Daastøl (1997). On the basis of Wolff's so-called 'German ethics' (1733), they make the case for the growth-centredness and dynamism of this approach, and of Wolff as a founding father of evolutionary economics: 'The English classical economists focused on the incentives of the market place as an external driving force for man. Wolff, although acutely aware of the role of markets and contracts, emphasizes Man's inner drive for learning, creating and innovating' (1997, p. 264). Reinert and Daastøl conclude by stating that 'in order to understand the knowledge-based society[, the] time has come to focus again on the Wolffian tradition' (ibid.).

Bibliography

Wolff's own writings are easily accessible in the form of a Collected Works edition, published since 1962 in three series (German texts, Latin texts, and materials and documents, the latter including texts by disciples as well as enemies) (Wolff, 1962–), the first two of which are now complete. Almost all of these are reprints of the most important original editions, many with substantial introductions. Wolff (1740, and especially 1754) are the basis for the arguments by Backhaus; Reinert and Daastøl focus on Wolff (1733). Wolff's strictly legal work is discussed in Nettelbladt (1754).

In English, the literature on Wolff, let alone by him, is quite meagre. The *European Journal of Law and Economics*, **4** (213), June 1997, comprises the above-mentioned essays by Backhaus and Reinert and Daastøl, as well as a biographical sketch by Drechsler (1997a) and Peter Senn's (1997) important study of the impact of Wolff on the social sciences, especially in the English-speaking world. There is a very good survey article on Wolff in the *Encyclopaedia Britannica*'s celebrated 11th edition (Pringle-Pattison, 1910). The translation of Wolff's *Jus gentium* (1934) is the most important Wolffian original text available in English. Among the German publications, a collection of essays on Wolff and his influence (Schneiders, 1986), including a comprehensive bibliography of more recent works on him, deserves special mention.

Notes

1. The biographical segment is based on Drechsler (1997a), where detailed references are available.
2. Fuhrmann (1997, pp. 40–44), presents a good synopsis of the concept of the house in the history of economic thinking.
3. This focus on the self-reliance of the smallest functioning unit, the household, is probably more than just homologous to communitarian (see Reese-Schäfer, 1994) as well as to neoconservative (see Drechsler, 1995, pp. 217–18) approaches today, in that it is arguably based on a similar view of society and individual.

References

Backhaus, J.G. (1997), 'Christian Wolff on subsidiarity, the division of labor, and social welfare', *European Journal of Law and Economics*, **4** (213), June, 129–46.

Cramer, J.U. von (1742–67), *Opuscula*, 5 vols; now also reprinted in Wolff (1962–) 3rd ser., 34 (1–5) (1996).

Drechsler, W. (1995), 'Benjamin Disraeli's social Toryism. Sanitas sanitatum, omnia sanitas reconsidered', *International Review of Comparative Public Policy*, **6**, 217–43.

Drechsler, W. (1996), 'Einführung', in Cramer (1742–67), 34 (1), pp. vii–liv.

Drechsler, W. (1997a), 'Christian Wolff (1679–1754): A Biographical Essay', in *European Journal of Law and Economics*, **4** (213), June, 111–28.

Drechsler, W. (1997b), 'Public administration as "state science"', *Foundations of Public Administration. A European Collection*, Tartu: at the University Press; in Estonian.

European Journal of Law and Economics, **4** (213) June; special issue on Christian Wolff, (ed.) J.G. Backhaus, now also reprinted in Wolff (1962–), 3rd ser., 45 (1996).

Fuhrmann, H. (1997), 'Die Lehre vom Haus und das Haus des Gelehrten', *Überall ist Mittelalter. Von der Gegenwart einer vergangenen Zeit*, 2nd edn, Munich: Beck.

Goebel, J. (1920), 'Christian Wolff and the Declaration of Independence', *Deutsch–Amerikanische Geschichtsblätter. Jahrbuch der Deutsch-Amerikanischen Gesellschaft von Illinois*, 18/19 (Jg. 1918/19), Chicago: Deutsch-Amerikanische Gesellschaft von Illinois.

Nettelbladt, D. (1754), 'Von den Verdiensten des Freyherrn von Wolf um die positive Rechtsgelahrtheit', *Hallische Beyträge zu der Juristischen Gelehrten Historie* 1[1755](2)I, 207–49.

Pringle-Pattison, A.S. (1910), 'Christian Wolff (less correctly Wolf)', in *The Encyclopaedia Britannica. A Dictionary of Arts, Sciences, Literature and General Information*, 11th edn, Cambridge: Cambridge University Press.

Reese-Schäfer, W. (1994), *Was ist Kommunitarismus?*, Frankfurt am Main: Campus.

Reinert, E.S. and A.M. Daastøl (1997), 'Exploring the genesis of economic innovations: The religious *gestalt*-switch and the *duty to invent* as preconditions for economic growth', *European Journal of Law and Economics*, **4** (213), June, 233–83.

Schneiders, W. (ed.) (1986), *Christian Wolff, 1697–1754. Interpretationen zu seiner Philosophie und deren Wirkung. Mit einer Bibliographie der Wolff-Literatur*, 2nd edn, Hamburg: Meiner.

Senn, P.R. (1997), 'Christian Wolff in the pre-history of the social sciences', *European Journal of Law and Economics*, **4** (213), June, 147–232.

Wolff, C. (1733), *Vernünfftige Gedancken von der Menschen Thun und Lassen, zu Beförderung ihrer Glückseeligkeit, den Liebhabern der Wahrheit mitgetheilet*, 4th edn; now also reprinted in Wolff, 1962–, 1st ser., 4 (1976).

Wolff, C. (1740), 'Von einer Erwegung der Staatsgeschäffte', in *Gesammlete kleine philosophische Schrifften, Sechster und letzter Theil, darinnen besonders die zur Staatsklugheit und der damit verbundenen Rechtsgelehrsamkeit gehörige Stücke enthalten* (tr. and ed. G.F. Hagen); now also reprinted in Wolff (1962–), 1st ser., 21(6), 1981, 323–528.

Wolff, C. (1754), *Grundsätze des Natur- und Völckerrechts, worin alle Verbindlichkeiten und alle Rechte aus der Natur des Menschen in einem beständigen Zusammenhange hergeleitet werden*, now also reprinted in Wolff, 1962–, 1st ser., 19 (1980).

Wolff, C. (1934), *Jus gentium methodo scientifica pertractatum* 2: *The Translation*, tr. J.H.

Drake, Oxford: Clarendon and London: Milford. (Translation of the 2nd edn of 1764; 1st edn, 1749.)
Wolff, C. (1962–), *Gesammelte Werke*, ed. J. École et al., Hildesheim, Zurich and New York: Olms.
Wolff, C. (1985), *Oratio de Sinarum philosophia practica / Rede über die praktische Philosophie der Chinesen* (ed.), M.Albrecht, Hamburg: Meiner.

Index

accidents 87, 94, 96–7, 352–3
  cost of 89
acidification policy in Netherlands 371
acquis communautaire 261–2, 266
Acta Borussica (Schmoller) 663
action, Weber on 734–5
activity level incentives 94, 95, 96
administration, science of 692–3
administrative costs 87
administrative law and economics 239–45
administrative structures 691
adoption laws 105, 108, 114
adversarial proceedings 580, 581, 582
adverse possession 254–5
agents, independent 80
agricultural indebtedness 577
agricultural production 592
agricultural reform 594
agriculture
  age of 677
  Roscher on 648–9
air pollution 365
Alcoa v. Essex 151, 155, 158
alimony 113
allocation 68, 89
  of accidents costs
  of resources 18, 20
  of rights 19
  systems 68–9
  theory 629
allocative effects 19–21
alternative dispute resolution 399
alternative liability rules 15–16
alternative precaution 89, 90, 92
American Economic Association 532
Amsterdam Treaty of European Union 279
anarchy 47
Angelini, Elizabetta Croci 261–79
Anglo-American legal analysis 61
Anglo-Saxon system 582
annulment of marriage 104, 107
Anson (*Law of Contract*) 574
Anti-Corn Law League 444, 561
antitrust law 494–7, 512
Arab and Israeli occupation 565
arbitration 399
artisan firms 646, 649
asceticism, religious and worldly 738–9
assent 454
asset change (*abusus*) 40
asset disposal 119
asset proprietors 42
asset specificity 9, 146
assignment of property rights 21, 249
assignment of rights and liability 19
assurance cover in US 360
atrocity, crimes of 479–80
attenuation of property rights 42
Austrian Civil Code (ABGA) 577
Austrian Civil Procedural Code (ZPO) 577
Austrian Criminal Code 577
Austrian school 560, 627, 628, 632
authority 153, 154
authors' rights 130

Backhaus analysis 236
Backhaus, Jürgen 280–88, 297–309, 329–44, 381–9
balance of payments, Stein on 512, 695–6
bank, central 650
banking 303, 554
  free 555, 631
Bank of England 165
bankruptcy law 55
barbarianism, Veblen on 729–30
bargaining 13, 26, 87, 97, 190, 256
  bargaining power 154, 194
  Coasean 15–18
  costless 41
barter 571, 573, 574, 630
Bass v. Gregory 54–5
Beccaria, Cesare (1738–94) 475–88
benefits and costs 68, 204, 263
benevolent government 200

Bentham, Jeremy (1748–1832) 58, 66–9, 129, 588
Berle-Means hypothesis 686
best practice approach 330–36
bilateral monopoly situations 669
bilateral precaution 89
bilateral relationship 156
bills of exchange 686
biology, child rearing 111
Bismarck's social reforms 693
Böhm, Franz (1895–1977) 489–98
bonding of investments 301–304, 306
Boomer v. Atlantic Cement Co. 250, 258, 345–6
borrowing, domestic 695–7
boundary errors 255
bounded rationality 9, 10, 31, 290, 453, 455–6
bourgeoisie 622
Brandeis brief 444–5, 448
Brandeis, Louis Dembitz 444–5, 448
breach of contract 529, 574
breach of promise action 107
brewery case 54–5
Bright, John 561
Brinig, Margaret 103–17
British Act of Navigation 602, 606
Brown v. Board of Education of Topeka 448
Buchanan, James M. 188–91, 208, 227, 515
budget 199, 695, 697
Buechtemann, Christoph 311–28
burden-sharing rule 696
bureaucracy, inefficiency of 594–5
business administration 585
business and economic law 58
business contracts 456
business cycle theory 554, 556, 643
businesses as entities 686
business transactions, legal forms 685

Calabresi, Guido 59, 61, 62, 97
Calvinism 673
capital 645, 666, 679, 694
capital and labour 681
capital and needs 655
capital markets 56, 206, 305, 307
capital punishment 481–2 *see also* death penalty
capitalism 193, 506, 622, 669, 688, 737–9
capitalist development 646
capitalist system 685
capitalists and workers 662–3
carbon taxes 385
care versus activity level 91–5
Carnival Cruise Lines, Inc. v. Schute (1991) 455, 457
carpal tunnel syndrome 438
Carroll Towing 90, 91
cartel agreements 231, 492, 5120, 668, 681
Cartel Law 577
causation 100
cause of external harm 247
cement factory pollution 345–6
central banks 165–81
centralization 200, 217, 654
change and power structures 601
Charter of Fundamental Rights 266
Chicago school 63–4, 187
child-bearing out of wedlock 106
child labour 446
child rearing 107–10, 123
children
  and divorce 113, 114
  in a family 103–105
  state provision for 107
Christianity 516, 654
Civil Procedure Code, Austrian 576–83
civilization 596–7, 653–4
class system 534, 623, 693,
climate policy 365
CO^2 emissions trading 365, 366, 367, 371
Coasean bargaining 15–18, 256
Coasean view on pollution 382–4
Coase, Ronald H. 7–39, 59, 63, 67
Coase theorem 7–39, 41, 46, 87, 247–9
Cobden, Richard 561
code law systems 448
Code Napoléon 430
codes of conduct 53
codetermination legislation 297–306, 494–7
codified customs 430
coercion 536, 537
coercive government regulation 252
cognitive science 453–57

collective bargaining 297, 300, 303
collective choice theory 63, 65
collective inheritance 119, 120–25
collectivism 652–3, 656–8
command-and-control instruments 356–8
command structure 8
commercial banking 166
Commercial Code 520
commercial contracts 289
commercial law 55, 56, 186, 524
commodities and wealth 666
common good 190, 200
common land 622, 624
common law 63, 195, 244, 448, 502, 549, 550–52
common policies 217–18, 273, 276–8
common property, problem 42–8
common resource, overuse of 49, 74
commons and anticommons 74–82
Commons, John R. 9, 499–507
communication 593, 605
communitarian legislation 214
community of property 647
company board representation 297–309
company law 577
comparative causation 94–5, 98
comparative negligence 93, 98
compensation 21, 41, 45, 68, 292, 358–60
  employees 45
  liability 93
  for pain and suffering 99
  right 252
  schemes 315–16
competition 8, 13, 17, 188, 514, 605
competition law 54, 58
competition market 205
complementarity 186, 212, 213
complete contract 158–9
compulsion 635, 537
computers 46, 436, 437, 438, 442
confiscations of goods and assets 482
conflict 168, 624
consensus test 227
constitutional economics 184–219, 223–37, 244–5, 285, 546
constitutional law 533
constitutional reform 553–4
constitution formation and maintenance 229–32
constraint 224, 573
consumer sovereignty 678
consumer valuation 629
consumption goods 655, 657
contractarianism 227–8, 475–6, 486–7
contract law 453–7, 526–10, 570–71
contracts 59, 197, 572, 573, 670
contractual agreement 22, 147, 428, 431
contributory negligence 93
cooperation 266–7, 276, 277, 624, 667
coordination 550
copyright-authors' rights 131–2
copyright law 11, 127
Corn Laws (1815) 623
corporate city tax, Germany 154–5
corruption of governments 201
cosmopolitanism 673
cost-benefit analysis 48–9, 62
costs 77, 385–7, 628
cost valuation 657
courts 53, 469, 578–9
creativity and knowledge production 130, 139
credit 512, 555, 638
credit basis of banks 555
credit policy 512
crime and criminal behaviour 623–4
  and punishment 71, 476–86, 625, 678
criminal bankruptcy 587
criminal law 56, 475, 524, 616
  in Austria 577
currency 165, 169
custody laws 106, 114
custom and precedent 551
customary law 70, 424–6, 427–31
customs unions 604
cyberspace 436

damages, punitive 638–9
day-care system 205
death duties 526–7, 528
death penalty 476, 481–2 *see also* capital punishment
debt management 523, 694–7
debtor-creditor law 55
decentralization 204, 201, 216
decision sharing 42
decoupled liability 96
default rule 54

delegation of decision-making powers 413–14
demand and supply 23, 643
democracy 56, 124–5, 553, 597
democratic accountability (EU) 271, 272
'democratic federalism' 197
'democratization' 188, 594
Demsetz, Harold 16–18, 20, 21–3
dependency 107
Depoorter, Ben 74–85
deposit protection 180
deterrence 483, 486
developers and the environment 78, 348, 562
devolution 219
Dewey, John 532
dialectics 601
diligence 95
dismissal protection regimes 311, 318, 319, 320, 321–5
dispute resolution 56, 248, 250, 393–401
distributional effects 14. 18. 20
divorce 106, 107, 110–14, 456, 665, 679
domain partitioning of land 42
domestic production 170, 605
dowry system in marriage 525
due care 90, 91, 92, 94, 98
duopolisitc anticommons 77–8
Dutch economic theory 586
dynamic adjustments 16

ecological tax credit 388, 389, 390
economic analysis 30, 465–71, 561, 701
  and the legal system 565
economic constitution 492–3
economic democracy (EU) 272
economic development 254, 591, 596, 693
  in the Roman *campagna* 683
economic efficiency 209, 592
economic growth 55, 169
economic integration and political institutions 212–14
economic liberalism 192
economic order 491, 492–3, 511
economic studies, Loria on 607–608
'economic system', Sombart on 684–5
economic thought, history of 683, 700
economists and law 32
economy, public and private 653
education 564, 591, 605, 658, 666, 669
efficiency 227–8, 333, 337, 346, 604–605, 637
electoral rules 194, 198–9, 523
eminent domain, government power of 251
emissions and tradable rights 352, 364–78
empathy 184
empathy in humans 676 *see also* sympathy
employment contracts 47, 48, 293, 496–7
employment protection 311–25
employment/unemployment 170, 290–91, 516, 616, 680
enforcement 53, 280, 317
Engels, Friedrich 618–26
England and English jurisdiction 562, 563
English Rule, legal costs 396–7
Enlightenment, French 673
Enlightenment, Scottish 676
enterprise, legal forms 685
entrepreneurship 563–4, 601
entropy, laws of 77–82
environment 267
environmental law and economics 345–49
Environmental Protection Agency, US (EPA) 365
equal opportunities 120–24, 194, 272
'equal treatment' 191
ethical and moral values 438
Eucken, Walter 508–18
European Central Bank 196, 264
European Commission 262, 263, 371
  emissions scheme 365–6, 371–5, 377
European constitution 273, 280
European Court of Justice (ECJ) 262, 263
European Economic Community (EEC) 261–79
European Monetary System (EMS) 224, 263, 264
European Union, institutional design 261–79
evaluation of goods 70, 656

evaluative analysis 465
evidence, Beccaria on 477
evolution of institutions 630, 631
exchange process 8, 252–2, 666
exchange rate, control of 170, 264
exchange, rules of 679
exchange value 599, 604, 629
exclusivity clauses 151
expenditure 227, 657
externalities 75, 76, 246, 247, 250
extortion problem 19–20

factories 646
Factory Acts 623
family 103–14, 123, 612, 613, 622
  and friends 282–4
  firm 109–10
  law 103, 104, 525
Farina, Francesco 184–222
farming, tenant 683
fathers, duties and obligations 106
fault 88, 89, 90
  in divorce law 107
  in liability 720, 721, 725
federalism 70, 200, 209–12, 236
Federal Cartel Office 494–5
Federal Rules of Civil Procedure 397
feminism 107
feudal states 692
Feuerbach, Ludwig 619
field cultivation 621
finance, public 450, 690
fiscal policies 216, 264, 270
Fisher Body v. General Motors 150–51, 152, 155, 158
fishing industry and power plant 382–4, 388
fixed-rent contract 47
force majeure 717–18, 724
foreign exchange control 171
formal rules 424
fortuitous events, liability 716–9
foster families 109
franchise model 105
Franklin, Benjamin 444
'free banking' theory 165, 166
freedom and constraint 597, 685
free enterprise, Jhering on 572
free inheritance 119, 120–25
free market approach 187, 637
free trade 263, 444, 586, 590–603, 597, 663
Freiburg School 489–91, 508, 516
French Revolution 242–3
fringe benefits for employees 46
functional schools in law and economics 58–71

general norms and customs 424–31
genetic heritage 108
German Civil Code 520–21
German corporate law 298–308 *see also* codetermination
German Economic Association 520
German economy 490–97
German Historical School 532, 560, 683
German *Länder* border controls 692
German reunification 695
Gierke, Otto von 519–21
globalization 212, 219, 670
goods and labour 655
goods and welfare 678
goods, public and non-public 45, 656–7
governance 10, 49
government 194, 197–9, 536–8, 658, 678
  and central bank 173, 175
  expenditure 45, 650
  and the judiciary 412
  and public enterprises 335–7
  regulation and laws 250, 598, 599, 703
Graziani, Augusto 522–30
greenhouse gas emissions 365
Guizot, François, Pierre Guillaume 444

Hale, Robert Lee 531–44
Hamilton, Alexander 444
harm caused 30, 47, 354, 355
harmonization 268–71, 374
Hayek, Friedrich Auguste von 187–8, 545–58
health and safety issues 269
Hegel, Georg 601, 619, 691
Henricus Stephanus 640
hereditary succession 526, 612, 614–15
Herzl, Theodor (1860–1904) 559–67
Hobbesian state of nature 192
Hobbes, Thomas (1588–1679) 43, 184
Hohfeld, Wesley Newcomb 533

hold-out problem 148–9, 251, 252
Holmes, Oliver Wendell Jr. 533–4
'homogeneity' in jurisdictions 207
homogenization 240
horizontal competition 191, 209
horizontal separation of powers 194–6
households 109, 110, 282–3, 746–7, 748
human behaviour 61, 123–5, 348–9, 442, 568
human capital 294, 496, 497
humanism 619
human performance 639
human rights 565, 566, 569
Hume, David (1711–76) 103, 184
hunter-gatherer economy 621, 677
hyperinflation 521
hypotheses, testable 61

ideas market 134, 135, 139, 140
ignorance 234–6, 244
immigration 563
immunity to accomplices in crime 480
impartiality of the judiciary 416
imperialist model 665
implementation failures 319–20
implicit contracts 154–5, 158, 294, 295, 296
imports and exports 512
incentive-based pay 292
incentive mechanisms 10, 11, 239, 242–4
incentives 95, 97, 150, 153, 191, 658
incentive system 171, 195, 203, 564, 600–601, 643
incomplete contract 145–60, 155–8, 195
independence of central bank 171–4, 177, 180, 196
independence of judiciary 413–18
index number formula, Laspeyres 585–6
individual action 18–22, 61, 68, 70, 226
individual and the state 662, 746, 748
individualism 64, 66, 230, 625, 657–8, 685
individualistic law systems 653–4, 656
individual rights 190–204, 246
individuals and goods 40
induction versus deduction 665
industrial emissions, social cost 27–9
industrial employment 649
industrialization 562, 594, 596, 730
industry, using machines 623, 645–6
inequality 193
infant mortality rate 110
inflation, high 170, 171, 172, 196
information 138, 257, 702
inheritance 119–25, 524, 526–8, 622, 649
  and children 123
injurer and victim 8, 89–90, 92–4, 98, 247–8
inquisitorial proceedings 579–82
institutional arrangements for judges 408–409
institutional complementarity 124–5, 156–8
institutional economics model, new 31
institutional law and economics 540, 541, 543
institutional theory of Schmoller 666
institution-building 216–18
institutions 10, 53, 145–60, 185, 683, 728–9, 731
insurance 87, 97, 219, 696
integration process in European Community 263–71
intellectual property 11, 127–44, 443, 448
interdependence costs 229
interest costs 695, 697
interest groups 217, 230, 231
interest rates 583, 623
international law and commerce 593, 617
international politics, criticism 599
international protection 650
international trade 424
interpersonal effects 68
interventionism, mercantilist 681
inventions 129, 130, 131, 257, 601
investment 298, 514, 655, 657, 679, 697
involuntary transfers 249, 254–5
irrationalism, legal 728–9
Islamic law 439, 448
Israel 559, 565, 566
Italian city states 686
Italian Civil Code 710–21, 725–6

Jefferson, Thomas 444
Jewish Company 562
Jewish law 439, 448, 563, 565, 566

Jewish state 564
Jhering, Rudolf von (1919–92) 568–75
joint-stock banks 165
joint-stock companies 562, 650, 686
Josselin, Jean-Michel 239–44, 424–32
judges 87, 401, 581–3
judicial accountability 415–20
judicial independence 232–3, 407–20
judicial opinions 533
judicial rewards 408
judiciary 173, 194
jurisdiction, political 204
jurisprudence 504, 625, 680, 681, 767–8
jury 449, 592–3, 600
justice 434 61, 64, 70, 434, 676, 677

Kaldor-Hicks test of potential compensation 64, 67–8
Kant, Immanuel (1724–1804) 184, 675
Keynes, John Maynard (1883–1946) 515, 545
Klein, Franz (1854–1926) 576–84
knowledge 435–9, 440, 442, 443, 546–9, 629–10
Kyoto Protocol 356, 365, 370, 378

labour 570, 679
labour contracts 289–96
labour force 109, 110, 112, 290, 645
labour legislation 495, 577
labour market institutions 264, 269, 314–16
labour relations 645
labour representatives 298, 300–302
labour unions and markets 499, 502
laissez faire 511, 515, 535, 541–2, 603
land 48, 594, 649, 680, 683
landed interests, criticism 599
Landes-Posner analysis 411–12
landlord-tenant law 257
landowners and polluters 24
language 66
language and meaning of law 504, 505
Laspeyres, Etienne (1834–1913) 585–9
lateral subsidiarity 284
law 54, 55, 58, 70, 524, 525, 577, 669, 741
  Schmoller 665–70
  see also administrative law; antitrust law; commercial law; common law; contract law; criminal law; environmental law; family law; Jewish law; patent law
lawsuit avoiding 354, 398–9, 583
lease contracts, Loria on 615
legacies 122, 123, 124
legal agreements, international 604
legal change in economic analysis 53–6, 347–8, 459, 540
legal framework 550–51, 631
legal institutions 56, 612, 688
legal problem, economic analysis 465–71
legal protection of property rights 249–52
legal realism 533, 639, 727–8
legal rules 11, 19, 27, 59, 60
legal sanctions 609, 610
legal structure 66, 691
legal studies movement 539
legal systems 561, 592, 593, 600, 691, 702
legal theory 465, 546, 547
legislation 631
legislative coalitions 232
legislative contracts 412
legislative power and the central bank 173
legislature 439
leisure class 534
lenience in punishing 481
Leviathan 200, 210, 429
liability 90, 91, 95, 454, 709–12, 719
liability of firm 351–2, 353
liability regimes 361
liability rules 22, 88, 95, 249
liability system 354, 355
liberalism 124–5, 192, 553, 649, 685
liberty 192, 600, 681
List, Friedrich (1789–1846) 590–606
litigation 100, 241, 417–19
local communities 203, 301
local government 201, 208, 216, 236
Lochner v. New York 533–4
Locke, John (1632–1704) 103, 128, 184, 194, 483
London School of Economics 10–12, 32
Loria, Achille (1857–1943) 607–17
loss-sharing 94–5

Lucas v. South Carolina Coastal Council (1992) 258

Maastricht Treaty 269, 280–81, 284
macroeconomics 643
majority rule 198, 658
manufacturing 592, 597
Marciano, Alain 239–45, 424–32
marginal utility 68, 629
Mark, German 276
market 548, 549, 552
market dynamics 34, 155–6, 186
market economy 212, 512, 563, 676
market exchange 13, 97, 192, 550
market failure 22–8, 64, 71, 246, 311–25
market institutions 66, 75
market interrelationships 604–605
market prices 61, 629, 668
market process 667–8
market rate of interest 555
market regulation 59, 347, 669, 681
marriage 103, 104, 106–108 *see also* divorce
Martineau. Harriet 103
Marx, Karl (1818–83) 601, 618–26, 646–7
Marxism 662, 737
materialism 620
mathematical techniques 66, 701
matriarchy 612
Mattiacci, Giuseppe Dari 87–102
mediation 399
Menger, Anton 561, 578
Menger, Carl 560, 627–34, 644, 659, 737
mental capital, List on 599, 600
mercantilism 681, 685
merchant interests 599, 604
mergers 294, 295, 514
'merit goods' 44
metal reserves 165–6
methodenstreit 631–3
methodologies 14–15, 62, 226, 665
Miceli, Thomas J. 246–60, 393–406
microeconomics 13, 60, 62, 700, 701, 703, 706
migration of workers 264
Mill, John Stuart (1806–73) 629, 646–7, 706
monetarism, List on 598
monetary and banking regimes 554
monetary policy 167, 171, 196, 224
money 254, 555, 628, 630, 679
monism 240, 241, 242
monitoring 291, 292, 293, 304
Monopolies Office 494–5
moral behaviour 565, 566, 609, 610, 665, 676
moral hazard 195, 207, 211, 252
moral imperative, Bentham 67, 68
morality 51, 184, 575, 576, 596
moral perfection, duty of man 605
moral rights 192
morals and culture 565, 566, 665
moral sentiments and jurisprudence 674–8
motivation for human action 67
Mugler v. Kansas 258
multinational companies 212
multiple equilibria 185
multiple exclusion rights 75, 76, 82
multiple rights 82
mutual assistance 582
mutual recognition 268–71

Nash, John 68–9
national debt 691, 696, 697
national defence 677
national economy 169, 510
national legal arrangements 593
natural law 281, 677
nature 46, 128
needs, private and public 656–7, 659–60
'negative freedom' 191, 193, 194
negligence 89, 90–95, 353, 454
negotiation and transfer of right 14
neoclassical theory 53, 629, 633
neoliberalism, Germany 489–90
Netherlands
  emissions trading 370–71
  economic power 586, 587
Nicita, Antonio 145–63
no liability 89–90
no-fault divorce 112–13
Nomoi (Plato's Laws) 636–8
non-market institutions, influence 66
non-observable precaution 91
non-pecuniary loss 99
normative analysis 63, 64, 65, 465–6

normative schools in law and economics 58–71
note-issuing banks 166, 168
novelty of invention 131
nuisance 54–5, 250–51

Oates Decentralisation theorem 203–204
oaths, Beccaria on 478
objectivization 686
observable precaution 91
oligopolistic anticommons 80–82, 702
opportunism 31, 146–51, 159, 211
opportunistic behaviour 195, 294, 295, 315, 316
opportunity costs 374–6
'optimal activity level' 95, 96
option fixing 669
oral transmission of rules 428
order, spontaneous, of the market 547–9, 550, 553, 630
ORDO, journal of German neoliberals 490, 495–6
Ordoliberals 508–509, 511, 513–6
ownership 13, 253–7, 346, 525–6, 621, 702

Pagano, Ugo 145–65
'pain' motive 67
pardons, Beccaria on 483–4
parenting 105, 108–109, 110–112
Pareto efficiency 13, 63–9, 187, 190, 202
Pareto Rule 348
Pareto, Vilfredo 67
Parisi, Francesco 7–39, 58–85, 87–100
parliament 243, 244
partial property rights 78
patent law 127, 131, 137, 139, 257, 600
path dependence 186, 306–308
patriarchy 103, 612, 613, 622
pecuniary rights 141
penal servitude 481–2
Pennsylvania Coal Co. v. Mahon (1922) 258
pension rule 696
perception 546
performance analysis 331, 336, 338, 341–3
permit trading in emission rights 366–70, 377
personal liability 686
Pigouvian view 12, 46, 247
plaintiff win rate 400
Plant, Sir Arnold 11
Plato (c.427–349 BC) 635–41
policy design failures 318–19, 320–21
policy goals 280–81
Politeia (Plato) 636, 640
political and economic history 415, 690–91
political decision makers 409, 411
political freedom 591, 596, 647
political science and law 532–3
politicians' interest in the judiciary 413–14
polluter-pays principle 374–8
pollution 352, 357–8, 382, 639
Popper, Karl R. 60–61, 640
population growth 48, 74
Porrini, Donatella 350–63
positive analysis 465, 509–10
positive externalities 42, 82
positive schools in law and economics 58–71
Posner, Richard 61, 63, 77, 103
possession 253–4, 526, 655
post-contractual opportunism 230, 291
Powell, Thomas Reed 532–3
power 195, 536
pragmatists 665
precaution 87, 88, 89
precedent 438, 448
press, freedom of 600
prevention 91
pricing 42, 643, 658, 667
price theory 512, 628–9
principal-agent theory 231, 241–2
prisoner's dilemma games 413
private appropriation 128, 129
private contracting 317
private corporations 339–41
private goods 26, 45, 189, 210
private incentive to sue 397–8
private interest 619
private pension funds 269
private property 40–41, 123, 255–8, 516, 523, 655
privatization 46, 47, 48, 334–6
privilege of use 74
procedural rules 232, 583

production 612, 620–21, 644, 648–51, 654
product liability 99
product warnings 454
profit 572, 679, 680, 685
profit-maximizing prices 80
proletariat 622, 625
property 246–58, 526, 574, 678
property law 55, 570, 613–18
property rights 27, 61, 75–82, 128, 236, 257, 563, 691
  partitioning of 40–52
property status on death 119
property values 539
proportional rule of government 198
protectionism 586, 587, 590, 603–605
Prussia, Schmoller on 663, 664
public and private goods 336–41, 648, 659
public choice theory 62–3, 217
public enterprise 329–43, 658
public expenditure 195
public finance 512, 587, 695
public institutions 330–41, 657–9
public law 179, 691
public policy 316–18, 450–52
public transportation system 205, 206
public utility regulation 538
punishments and crimes 476–86, 676

quantitative methods 62

racial discrimination 448
railroad construction 436, 446, 590, 596
Ramello, Giovanni B. 127–43
rationalism, Western 737, 738, 739
rationality of law 63, 741–2
Rawls, John 66–9
real estate 613
reality, conception 601
recording systems 254
regional government 236, 643
regulation 46, 593, 681, 685
regulation systems 355–6
regulation versus liability 352–6
regulatory instruments 603, 604
regulatory taxation 381–90
relationships, compulsive 609, 610
religion, Weber's study of 737, 738, 740
  freedom 591, 647
  laws 53, 439
remedies, private and public 28
rent seeking 201, 230–31, 236–7, 239_40, 244–5
representative government 523
reproductivity rule 654, 696
reputation 154–5, 294, 448
residence jurisdictions 28, 206
residual control rights 153
resource allocation 628, 629
resource, overuse of 74
resource use (*usus*) 40
responsibilities of private bodies 284
retirement age 588
retirement policies, mandatory 295
revenue, public 334, 656, 694
reward and punishment 40–41, 568–9, 573
right exhaustion in EU jurisprudence 129
rights and power 541
rights in policy decisions 495, 496
rights, non-optimal allocation 15
rights of exclusion 76
rights of individuals 655
rights of private bodies 284
rights ownership 137
right to judge the state 241
right to returns (*usus fructus*) 40, 41
'right to all' 43
right to sell or lease (*alienation*) 40
rights transfer and allocation 15, 19, 23, 26–8, 249
risk and profit 97, 216, 563
risk and uncertainty 702
risk control 354
risk-creating activities 88, 96
risk estimation 453, 454, 455
risk liability, strict 709–14, 718, 719–24
risk sharing 218, 316, 322
risks and policy failure 318–21
road tolls 46
Roman Catholic Church 435–6
Roman law 430, 568, 613–15, 619, 654, 673, 686
Roscher, Wilhelm Georg Friedrich 560, 642–51
rule enforcement 280
rule of Law, Hayek's doctrine 552–3

sales in official markets 637
same-sex marriages 107
Sax, Emil (1845–1927) 652–61
scarce resources 16–17, 44
Schengen Agreement 267
Schleswig-Holstein 689–90
Schmid, A. Allan 345–49
Schmoller, Gustav von 444, 520, 561, 632, 662–70
science and technology 131, 139, 330
science in law 433–7, 592
seat belt law 454
secession 218–19
securities law 55–6
sedition 595
self-curing failures 15–18
self-enforcement 295, 296
self-interest 9, 31, 569, 609, 610
self-realization 193
Seligman, E.R. A. 532
Senn, Peter R.
sensitivity analysis 468
separation of powers, horizontal and vertical 194–7
separation process costs 218
services, provision 655
settlement-trial decision 393–5
Šević, Željko 165–83
Sherman Act 54
shirking 291
simple contract 154, 158
simple negligence 92–3
single currency 263
Single European Act 268
Single Market 263, 270, 274
slavery, List on 600, 614–15, 621, 644–6
Smith, Adam (1723–90) 129, 601–603, 672–81
Smyth v. Ames (169 U.S. 466, 1898) 539
social and economic conditions 451
Social Charter 1989 (EU) 269
social constructivism 501, 503–504
social contracts 189, 191, 244, 428
social cost 10–12, 14, 20, 22, 25, 28
  of litigation 398
social institutions 549, 564, 621
social learning 546–8
social norms 70, 424
social policies 267, 663, 664, 665
social problems 446, 450–51, 561, 669
social relationships 687, 735
social responsibility 178–9
social rights 191–4
social science 62, 437, 443, 459, 546
  as source of law 442–52
social security provision 552, 693
social welfare 66–9, 187–8, 488, 657, 701–702
  Beccaria on 475–6
socialism
  Roscher on 646–8
  Sombart on 684
Society of Jews 562
sociological theory 461
sociology, Weber on 733–4, 735, 740
Sombart, Werner 683–88
sovereign nation model 105
Soviet Union 224
space law 437
special interest projects 231–2
species, prolongation of 111
spontaneous order 187, 188
Spur Industries v. Del E. Webb Development Co. (1972) 248, 258
Stammler, Rudolf 611–12
starvation 68
state 239–40, 622, 654, 658
statistics 564, 585, 586
statutes 53
Stein, Lorenz von (1815–90) 689–98
Steinbach, Erich 578
Stigler, George Joseph ((1911–92) 700–707
strategic behaviours 20–21, 240–41, 277, 278
strategic stalls in Coasean bargaining 21–2
strict liability 89–94, 96
strikes 300, 303
structure analysis, public enterprises 341–2
subsidiarity, principle of 209, 266–7, 280–86, 649
subsidization by government 44
substitution nexus 212
succession model 665
supervision 647
supply and demand 667, 668
supporting aged parents 111
supranationalism 264, 265

surrogacy contracts 108–109
Swedish Rijksbank 165
sympathy in humans 425, 426, 675, 676

tariffs 596, 603–604
taxation 247, 270–71, 350, 587–8
  see also regulatory taxation
tax credits, environmental 388 , 389, 390
taxes 28–9, 357, 381, 389
  see also inheritance
tax incentives, environmental 386–7
tax liabililty 119–20
tax revenues 336
tax system 267, 691–7
taxonomy of liability rules 88–95
technology and law 48, 138, 436, 564
technology and natural science 339–41
television in court 438
temporal division of ownership 257
tenants 615–16
territory for new Jewish state 564
third party 78
Tiebout model 204–12, 236
Tietzel, Manfred 40–56
time lapse 629, 630, 633
title assurance 252–5
Torrens registration system 254
tort law 87–100, 352. 353, 453
torture, Beccaria on 478–9
tradable emission rights 364–78
trade 272, 587, 604, 630–31, 650
Trade Association 596
trade cycle, Austrian theory of 554
trade in Germany 595, 663
trade secrets 127, 132–3
trade unions 171, 302
trademark 127, 130, 133–4, 136, 137
Trade-Related Aspects of Intellectual Property Rights (TRIPS) 130
traffic congestion 44, 46
training of workers 295, 295
transaction cost economics 7–39,146–6
transaction costs 187, 248, 251, 491–2, 702
transferability of individual rights 14
transfer of property 252–5
transnational alliances 217–18
transport and communication 650
transport systems 44, 591
Treaty of Rome 53–4, 267–8
trespass 250–51, 254
Treuhand legacy 329
trials 479
Trimarchi, Pietro (1934–) 709–26
two-party accidents 89
Tyson & Bro v. Banton 534, 543

unanimity test 227
uncertainty 98, 188
underdeveloped nations 591
understanding, Weber on 734, 735
unemployment, insurance against 669
unification of Germany 692
unilateral precaution 89, 92
unions 292, 300 *see also* trade unions
unitary state 199–200
United States
  Constitution 194–5
  Federal Reserve 196
USA v. Microsoft Corporation 142
USA v. United Shoe Machinery 137
use and exclusion right 76
utilitarianism 58, 67–8, 482, 484, 486–7
utility function 200
utility maximization 61, 70
utopia, Plato's 636, 638, 640

value and motivation 67, 667
value-enhancing transformation 78
value judgements 465–6
Van den Hauwe, Ludwig 223–37
Veblen, Thorsten 534, 535, 727–32
vertical composition 209
vertical equity 202–203
vertical integration 152–3, 158, 294
veto power 267
vicarious liability 98
victims 89–90, 92, 93–4
Virginia University 11–12
voluntary and involuntary exchange 254
voluntary social interaction 71
voluntary transfers 249
voting system 194, 197–8, 231

wage differentials 292
wage policy 512
wages and incentives 614–15
wage stability 290
Wagner, Adolf Heinrich Gotthilf 561

Wagner, Richard E. 119–25
Walrasian orientation 629, 630, 633
war economy 521, 601
warlord economies 56
wealth accumulation 282–3, 527–8, 614
wealth and property 686
wealth and utility 69
wealth creation, global 590
wealth distribution and redistribution 60, 574
wealth maximization 61, 70
Wealth of Nations, The 672–81
wealth transfers 14, 15, 17, 18
Weber, Max (1864–1920) 733–44
Weimar Republic 665
welfare 269–70, 565
welfare economics 190, 200, 469
welfare-maximizing prices 80
welfare services and the state 693
welfare state 506, 578, 743, 746–7
welfare state provision 45, 264, 281–2
well-being of members of society 69
Wicksell, Knut 226–9
wilderness preserve 75
will of human beings 612
wine growing industry and power plant 382–4, 385, 388
witness credibility 477
Woerdman, Edwin 364–80
Wolff, Christian (1679–1754) 745–9
workers on works councils 297
working conditions 525
working hours 534, 669
working rules 505
working-class 506
work organization 686
works councils in Germany 300, 301, 302
World Wide Web 436, 437

yardstick competition 209

zero cost mobility 208, 210
Zionism 560, 562, 567